MONOGRÁFIA

Korszerű módszerek az empirikus társadalomtudományi kutatásban

Kutatói kézikönyv

Turós Mátyás, Soltész-Várhelyi Klára

L'Harmattan

ESZTERHÁZY KÖNYVTÁR

This work is licensed under Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International. https://creativecommons.org/licenses/by-nc-sa/4.0/

Korszerű módszerek az empirikus társadalomtudományi kutatásban

Modern Methods in Empirical Social Science Research

Turós Mátyás, Soltész-Várhelyi Klára

https://orcid.org/0000-0002-4710-550X
https://orcid.org/0000-0003-4369-6936

Interdiszciplináris társadalomtudományok / Interdisciplinary social sciences (12965), Ökonometria, statisztikai módszerek / Econometrics, statistical methods (12778), Ipari és társadalomtudományi alkalmazott matematika / Industrial and social sciences applied mathematics (10025)

társadalomtudományi kutatásmódszertan, kvalitatív szövegelemzés, strukturális egyenletmodellezés, lineáris kevert modellek, természetesnyelv-feldolgozás
social science research methodology, qualitative text analysis, structural equation modeling, linear mixed models, natural language processing

ISBN 978-615-7145-03-0 (pdf)

DOI: https://doi.org/10.56037/978-615-7145-03-0

L'Harmattan Open Access

https://openaccess.pub

Turós Mátyás, Soltész-Várhelyi Klára

Korszerű módszerek az empirikus társadalomtudományi kutatásban

Kutatói kézikönyv

Eszterházy Könyvtár – Bibliotheca Eszterhazyana

MONOGRÁFIA ALSOROZAT

A sorozatot szerkeszti a szerkesztőbizottság:
Prof. Dr. Yoo Jinil, Seoul (KR)
Prof. Dr. Bednanics Gábor, Eger (HU)
Prof. Dr. Rainer M. János, Eger (HU)
Prof. Dr. Sárkány Péter, Eger (HU)
Prof. Dr. Schwendtner Tibor, Eger (HU)
Prof. Dr. Szűts Zoltán, Eger (HU)
Dr. Boros Lili, Eger (HU)
Dinnyés Patrik, Eger (HU)

Turós Mátyás, Soltész-Várhelyi Klára

Korszerű módszerek az empirikus társadalomtudományi kutatásban

Kutatói kézikönyv

Eszterházy Károly Katolikus Egyetem – Líceum – L'Harmattan

Eger – Budapest, 2026

Lektorok:
dr. habil Letenyei László
dr. habil Racsko Réka
dr. habil Vida Gergő

© Szerzők, 2026
© Eszterházy Károly Katolikus Egyetem, 2026
© Líceum Kiadó, 2026
© L'Harmattan Kiadó, 2026
© L'Harmattan Open Access, 2026

ISSN 3058-1931
ISBN 978-963-646-605-3
ISBN 978-615-7145-03-0 (pdf)
DOI: https://doi.org/10.56037/978-615-7145-03-0

L'Harmattan France
5–7 rue de l'Ecole Polytechnique
75005 Paris
T.: (33) 1 40 46 79 20
E-mail: diffusion.harmattan@wanadoo.fr

L'Harmattan Italia
SRLVia Degli Artisti 15
10124 Torino
T.: (39) 011 817 13 88 / (39) 348 39 89 198
E-mail: harmattan.italia@agora.it

A kiadó kötetei megrendelhetők,
illetve kedvezménnyel megvásárolhatók:
L'Harmattan Könyvesbolt
1053 Budapest, Kossuth L. u. 14–16.
Telefon: (36) 70 554 3177
harmattan.hu

Nyelvi lektorálás és tördelés: Líceum Kiadó
Borítóterv: Kára László
Nyomdai kivitelezés: Prime Rate Zrt.
Felelős vezető: Tomcsányi Péter

Tartalomjegyzék

ELŐSZÓ 13

I. A MEGISMERÉS ÉS TORZÍTÓ TÉNYEZŐI A TÁRSADALOMTUDOMÁNYI KUTATÁSBAN 19

1. Bevezetés 21
2. A hétköznapi megismerés torzításai és a tudományos módszer 25
3. Tudományos és nem tudományos szabványok 32
4. Torzítások a tudományos publikálásban 38
5. A káros „kvantitatív vagy kvalitatív" dichotomizálás 43
6. Zavaró tényezők (Confounding factors) – a harmadik változó problémája 46
7. A zavaró tényezők kezelése a kutatási folyamatban 54
8. Miért volt fontos a zavaró tényezőkkel ennyit foglalkozni? 60
9. Összefoglalás 61
10. Önellenőrző kérdések 62

II. INTERPRETATÍV FENOMENOLÓGIAI ELEMZÉS (IPA) 67

1. Bevezetés 69
2. Elméleti alapok 69
3. A kutatás folyamata 71
   3.1. Adatgyűjtés 71
   3.2. Adatelemzés 76
4. A minőség biztosítása 82
5. Mikor használjuk és mikor ne használjuk a módszert? 84
   5.1. A kutatási kérdés 84
   5.2. A minta 85
   5.3. A kutatás jellege 86
6. A módszer változatai és haladó alkalmazásai 87
7. Összefoglalás 89
8. Önellenőrző kérdések 90

III. REFLEXÍV TEMATIKUS ELEMZÉS (RTA) 93

1. Bevezetés 95
2. Elméleti alapok 99
3. A kutatás folyamata 101
   3.1. A kutatás tervezése és az elméleti pozicionálás 102
   3.2. Az adatelemzés 104
4. A minőség biztosítása 107
5. Mikor használjuk és mikor ne használjuk a módszert? 109
   5.1. A kutatási kérdés 109
   5.2. A minta 109
   5.3. A kutatás jellege 110
6. A módszer változatai és haladó alkalmazásai 111
7. Összefoglalás 112
8. Önellenőrző kérdések 113

IV. SZEMANTIKUS TARTALOMELEMZÉS (SCA) 117

1. Bevezetés 119
2. Elméleti alapok 120
3. A kutatás folyamata 123
   3.1. A kutatási kérdés megfogalmazása 123
   3.2. A kezdeti munka a szöveggel: olvasás, jegyzetelés és eset-összefoglalók készítése 124
   3.3. A kategóriarendszer fejlesztése és a kódolás folyamata 125
   3.4. Az elemzés 126
   3.5. Az eredmények bemutatása és a folyamat dokumentálása 128
4. A minőség biztosítása 130
5. Mikor használjuk és mikor ne használjuk a módszert? 132
   5.1. A kutatási kérdés 132
   5.2. A minta 132
   5.3. A kutatás jellege 133
6. A módszer változatai és haladó alkalmazásai 134
7. Összefoglalás 135
8. Önellenőrző kérdések 135

V. KVALITATÍV ÖSSZEHASONLÍTÓ ELEMZÉS (QCA) 139

1. Bevezetés 141
2. Elméleti alapok 143
3. A kutatás folyamata 146
   3.1. A kutatási kérdésnek megfelelő esetek és feltételek kiválasztása 146
   3.2. Adatgyűjtés és kalibráció 147
   3.3. A szükséges feltételek elemzése 151
   3.4. Az igazságtábla létrehozása és elemzése 155
   3.5. Logikai minimalizálás 159
   3.6. Az eredmények értelmezése és validálása 160
4. A minőség biztosítása 165
5. Mikor használjuk és mikor ne használjuk a módszert? 166
   5.1. A kutatási kérdés 166
   5.2. A minta 167
   5.3. A kutatás jellege 168
6. A módszer változatai és haladó alkalmazásai 168
7. Összefoglalás 170
8. Önellenőrző kérdések 171

VI. EPISZTEMIKUS HÁLÓZATELEMZÉS (ENA) 175

1. Bevezetés 177
2. Elméleti alapok 178
3. A kutatás folyamata 179
   3.1. Adatgyűjtés és előkészítés 179
   3.2. Kódolás 179
   3.3. Sztanzaelemzés 180
   3.4. Értelmezés 183
   3.5. Az eredmények értelmezése 184
4. A minőség biztosítása 186
5. Mikor használjuk és mikor ne használjuk a módszert? 187
   5.1. A kutatási kérdés 187
   5.2. A minta 188
   5.3. A kutatás jellege 188
6. A módszer változatai és haladó alkalmazásai 191
7. Összefoglalás 194
8. Önellenőrző kérdések 195

VII. A TERMÉSZETESNYELV-FELDOLGOZÁS NÉHÁNY MÓDSZERE (NLP) 199

1. Bevezetés 201
2. Elméleti alapok 203
3. A kutatás folyamata 210
   3.1. Névelem-felismerés (Named Entity Recognition – NER) 212
   3.2. Szentimentelemzés (Sentiment Analysis) 215
   3.3. Témaelemzés (Topic Modeling) 219
4. A minőség biztosítása 225
   4.1. Felügyelt módszerek értékelése 225
   4.2. Felügyeletlen módszerek értékelése 227
   4.3. Általános minőségi követelmények minden NLP-kutatásban 228
   4.4. Modellek összehasonlítása és statisztikai szignifikancia 229
5. Mikor használjuk és mikor ne használjuk a módszert? 230
   5.1. A kutatási kérdés 230
   5.2. A minta 231
   5.3. A kutatás jellege 231
6. A módszer változatai és haladó alkalmazásai 233
7. Összefoglalás 236
8. Önellenőrző kérdések 237

VIII. MEGERŐSÍTŐ FAKTORELEMZÉS (CFA) 241

1. Bevezetés 243
2. A CFA koncepciója 243
   2.1. A modell specifikációja (model specification) 244
   2.2. A modell azonosíthatósága (model identification) 248
   2.3. Paraméterbecslés (model estimation) 249
   2.4. A modell értékelése (model evaluation) 252
   2.5. A modell újraspecifikálása (model respecification) 254
   2.6. A modell értelmezése (model interpretation) 257
   2.7. További elemzések 257
3. Kutatási folyamat 262
   3.1. Mintaválasztás 263
   3.2. Adatgyűjtés, az adatok előkészítése 263
   3.3. Az elemzés beállításai 264
   3.4. Elemzés értelmezése 268
4. Mikor használjuk és mikor ne használjuk a módszert? 270
5. A minőség biztosítása 271

6. A módszer változatai és haladó alkalmazásai 272
7. Összefoglalás 273
8. Önellenőrző kérdések 274

IX. ÚTVONALELEMZÉS (PA) 277

1. Bevezetés 279
2. Elméleti alapok 280
2.1. A modell specifikációja (model specification) 280
2.2. A modell azonosíthatósága (model identification) 291
2.3. Paraméterbecslés (parameter estimation) 293
2.4. A modell értékelése (model evaluation) 296
2.5. A modell újraspecifikációja (model respecification) 297
2.6. A modell értelmezése (model interpretation) 301
3. Kutatási folyamat 304
3.1. Mintaválasztás 304
3.2. Adatgyűjtés, az adatok előkészítése 304
3.3. Az elemzés beállításai 306
3.4. Az elemzés értelmezése 308
4. Mikor használjuk és mikor ne használjuk a módszert? 312
5. A minőség biztosítása 313
6. A módszer változatai és haladó alkalmazásai 315
7. Összefoglalás 315
8. Önellenőrző kérdések 316

X. STRUKTURÁLIS EGYENLETMODELLEZÉS (SEM) 319

1. Bevezetés 321
2. A kovarianciaalapú modellezés (CB-SEM) 322
2.1. A CB-SEM elmélete 322
2.2. CB-SEM kutatási folyamata 325
2.2.1. Mintaválasztás 325
2.2.2. Adatgyűjtés, az adatok előkészítése 325
2.2.3. Az elemzés beállításai 327
2.2.4. Az elemzés értelmezése 329
2.3. A minőség biztosítása 330
3. A parciális legkisebb négyzetek módszerén alapuló SEM (PLS-SEM) 332
3.1. A PLS-SEM elmélete 332
3.2. PLS-SEM kutatási folyamata 339

3.2.1. Mintaválasztás 339
3.2.2. Adatgyűjtés, az adatok előkészítése 339
3.2.3. Az elemzés beállításai 340
3.2.4. Az elemzés értelmezése 341
3.3. A minőség biztosítása 342
4. Mikor használjunk CB-, és mikor PLS-SEM-et? 344
5. A módszer változatai és haladó alkalmazásai 345
6. Összefoglalás 347
7. Önellenőrző kérdések 348

XI. BEÁGYAZOTT ANOVA (NESTED ANOVA) 351

1. Bevezetés 353
2. Elméleti háttér 354
2.1. A függetlenség feltétele 354
2.2. Keresztezett és beágyazott kísérleti elrendezések 356
2.3. A modell kiegyensúlyozottsága 361
2.4. Fix és random hatások 363
3. Kutatási folyamat 367
3.1. Mintaválasztás 367
3.2. Adatgyűjtés, az adatok előkészítése 368
3.3. Az elemzés beállításai 369
3.4. Az elemzés értelmezése 371
4. Mikor használjuk és mikor ne használjuk a módszert? 372
5. A minőség biztosítása 375
6. A módszer változatai és haladó alkalmazásai 377
7. Összefoglalás 377
8. Önellenőrző kérdések 378

XII. LINEÁRIS KEVERT MODELLEK (LMM) 381

1. Elméleti háttér 383
1.1. A hagyományos kísérleti elrendezések 383
1.2. Fix és random hatások 387
1.3. Az adatok strukturálása 389
1.4. A varianciák modellezése 391
1.5. Döntés a fix és random hatásokról 394
1.6. A fix hatások modellezése 396
1.7. A random hatások modellezése 397

1.8. Első szint, a random intercept 401
1.9. Ismételt mérések modellezése 418
1.10. Második szint, a random slope 432
1.11. Összegzés 441
1.12. Modellilleszkedés 443
2. Kutatási folyamat 444
2.1. Mintaválasztás 444
2.2. Adatgyűjtés, az adatok előkészítése 445
2.3. Az elemzés beállításai 446
2.4. Az elemzés értelmezése 450
3. Mikor használjuk és mikor ne használjuk a módszert? 451
4. A minőség biztosítása 452
5. A módszer változatai és haladó alkalmazásai 453
6. Összefoglalás 454
7. Önellenőrző kérdések 454

XIII. AZ EREDMÉNYEK ÁLTALÁNOSÍTHATÓSÁGA ÉS PREDIKTÍV EREJE: SZIGNIFIKANCIA, BOOTSTRAP ÉS KERESZTVALIDÁCIÓ 457
1. Bevezetés 459
2. A szignifikancia 463
3. Bootstrap 469
4. Keresztvalidáció 474
5. Összefoglaló 481
6. Ellenőrző kérdések 482

IRODALOM 485

ANGOL–MAGYAR TERMINOLÓGIAI EKVIVALENCIAJEGYZÉK 499

GLOSSZÁRIUM 505

Előszó

A magyar nyelvű tudományos módszertani irodalom – kutatói kézikönyv vagy felsőoktatási tankönyv formájában – a terjedelem nagyobb részében a társadalomtudományi kutatás elméleti kereteit és alapjait tárgyalja, kisebb részében pedig a 20. század elejétől kialakított olyan módszereket ismerteti elméleti síkon, mint például a megfigyelés, a kísérlet vagy a kikérdezés. Kutatói kézikönyvünk nem a már jól ismert információkat ismétli vagy mutatja meg újabb oldalról – ha ez egyáltalán lehetséges lenne –, hanem az a célja, hogy a 20. század utolsó évtizedeitől széleskörűen elterjedt, valamint az utóbbi két évtizedben egyre inkább alkalmazott eljárásokat az elméleti keretek rövid megadása mellett alapvetően a gyakorlati alkalmazást elősegítve ismertesse.

A könyv nem vállalkozik a statisztikai és tudományfilozófiai alapok (például a módszerek alapvető matematikája, a szignifikanciatesztelés elemi logikája vagy a klasszikus episztemológiai keretek) bemutatására, mivel ezeket a hazai szakirodalom már kiterjedten tárgyalja (Csíkos, 2020; Hunyadi & Vita, 2019; Ivanović, 2016; Kaposi & Szőke-Milinte, 2022; Kontra, 2011; Vargha, 2015, 2019). Hasonlóképpen, könyvünk nem foglalkozik részletesen a kvalitatív kutatásmódszertannal sem, mivel ezt a területet a hazai szakirodalom már átfogóan feldolgozta (Rácz et al., 2023; Sántha, 2009, 2015a, 2015b, 2022). Könyvünk elsősorban azoknak szól, akik már ismerik a társadalomtudományi kutatásmódszertan elméleti alapjait, fogalmait és alapvető módszereit, vagyis tisztában vannak az empirikus kutatás folyamatával, például a kutatási kérdések lehetséges forrásaival, a hipotézisek megfogalmazásával és tesztelésével, az alapsokaság és minta kapcsolatával, a mintavételi eljárásokkal, az adatbázis létrehozásának összes kérdésével, vagy például könnyen értelmeznek egy korrelációs mátrixot. Ez utóbbival összefügg, hogy kötetünk nem statisztikai tárgyú (lásd például Hajtman & Hajtman, 2018; Vargha, 2015, 2019, 2020), de azoknál a módszereknél, ahol lényeges vagy elengedhetetlen, bemutatjuk a statisztikai eljárások alapjait. Könyvünkben jól vagy kevésbé elterjedt szoftverek használatát sem ismertetjük, az egyes módszerekhez szándékosan nem mellékelünk kódokat vagy a grafikus felületű szoftverek menüpontjairól képernyőképeket, mert szoftverekből sok van, néhány eltűnik, a többi folyamatosan alakul, újak jelennek meg. Szoftverképek és leírások megadása helyett azokat a fogalmakat és azoknak a mutatóknak a neveit közöljük, amelyek egy elemzési eljáráshoz és az elemzés eredményeinek

leírásához szükségesek. Ezeket a szoftverek ugyan különböző módon és formában, de biztosítani fogják.

A kötet első fejezete röviden a megismerés kérdéseivel foglalkozik kutatásmódszertani szempontból. E fejezet célja az is, hogy az olvasót a témába bevezesse. A további fejezetekben mutatjuk be az egyes módszereket. Ezek sorrendjét, vagyis a kötet központi fejezeteinek felépítését nem a kvalitatív és kvantitatív paradigma, hanem a nemzetközi irodalomból a kvalitatív és kvantitatív kifejezésekhez hasonlóan jól ismert strukturálatlan és strukturált adatok szerinti megközelítés adja. Ezt azért tesszük, mert a strukturálatlanság és strukturáltság nem megismerési paradigmákra, hanem a kutatási adatok szervezettségére és feldolgozhatóságára utal a következők szerint. A strukturált adatok előre meghatározott alakúak, könnyen kereshetők, általában adatbázisokban vagy táblázatokban találhatók, felépítésük egyértelmű sémát követ (például Excel-táblázatok, SQL-adatbázisok). Ezzel szemben a strukturálatlan adatok nem követnek előre meghatározott formát vagy sémát, változatos formában jelenhetnek meg (például hangfelvételek, videók, képek), nehezebben kereshetők és elemezhetők. A különbség gyakorlati jelentősége abban áll, hogy a strukturált adatok közvetlenül alkalmasak matematikai feldolgozásra, a strukturálatlan adatok pedig a) különböző nem matematikai módszerekkel elemezhetők (értelmezhetők), vagy b) először előkészíthetők (strukturálhatók) a matematikai elemzéshez, és ezt követően a matematikai elemzésből származó eredmények értelmezhetők. Vagyis szövegeket (például szakértőkkel készített interjúkat), képeket (például fotókat), műalkotásokat (például egy médiaóra diákmunkáit) lehet tisztán hermeneutikai (nem matematikai) és matematikai módszerekkel is elemezni. A kvantitatív és kvalitatív elkülönítés kerülését indokolja az is, hogy a strukturált adatokon nyugvó matematikai elemzés sem mellőzi az interpretációt – gondoljunk faktorok vagy klaszterek elnevezésére – vagy a megértést – gondoljunk az eredmények értelmezésére. Akkor milyen fogalmakat és hogyan használunk? A kvalitatívnak nevezett kutatási hagyomány intellektuális gyökereit a hermeneutikában találjuk meg. Ennek megfelelően a kvalitatív jelző helyett gyakran fogjuk a pontosabb és mélyebb tartalmat hordozó hermeneutikai kifejezést használni. Ugyanakkor nem hagyjuk teljesen figyelmen kívül a kvalitatív és kvantitatív kifejezéseket. Egyrészt a tudományos diskurzusban annyira elterjedtek, hogy használatuk megkönnyíti a tájékozódást, másrészt pedig egyes módszerek, mint például a kvalitatív összehasonlító elemzés, elnevezésükben is hordozzák ezt a megjelölést. A továbbiakban ezeket a szavakat majd azzal a kritikai háttértudással használjuk, amelyet az első fejezet egy része a kvalitatív–kvantitatív dichotómia problémáiról bemutat.

Kötetünkben tizenegy korszerű empirikus kutatási eljárást tárgyalunk: szemantikus tartalomelemzés (SCA), reflexív tematikus elemzés (RTA), interpretatív fenomenológiai analízis (IPA), a természetesnyelv-feldolgozás (NLP) néhány módszere, kvalitatív összehasonlító elemzés (QCA), episztemikus hálózatelemzés (ENA), megerősítő faktorelemzés (CFA), útvonalelemzés (PA), strukturális egyenletmodellezés (SEM), beágyazott ANOVA (Nested ANOVA), lineáris kevert modellek (LMM). Végül egy elméleti fejezetben ismertetjük a szignifikanciatesztelés, a bootstrap és a keresztvalidáció módszertanát. E módszerek választását az indokolta, hogy az elmúlt évtizedekben egyre nagyobb szerepet kaptak a társadalomtudományi kutatásokban, ugyanakkor a magyar nyelvű módszertani irodalomban – különösen a fejlett modellezés területén – még viszonylag kevés gyakorlati útmutatás található alkalmazásukról.

A fejezetek sorrendjét a strukturálatlan adatoktól a strukturáltakig tartó íven a fokozatosság határozza meg. A módszertani fejezeteket a strukturálatlan adatok feldolgozására talán leggyakrabban alkalmazott eljárás, a tematikus elemzés módszercsaládjának bemutatásával kezdjük. A „tematikus elemzés" kifejezést szándékosan használjuk ernyőfogalomként, mivel álláspontunk szerint több, filozófiájában eltérő irányzatot foglal magába. Kötetünk három fő irányzatot egy-egy meghatározó, korszerű módszeren keresztül mutat be: a) a megélt tapasztalatra összpontosító, specializált fenomenológiai megközelítést az interpretatív fenomenológiai elemzés (IPA) (II. fejezet); b) a korszerű, interpretatív vonalat a reflexív tematikus elemzés (RTA) (III. fejezet); c) a hagyományosabb, szabálykövetőbb irányzatot – amelyet gyakran kvalitatív tartalomelemzésnek is neveznek – a szemantikus tartalomelemzés (SCA) képviseli (IV. fejezet). Míg az IPA tisztán, az RTA alapvetően strukturálatlan adatokkal dolgozik, az SCA esetében az elemzés jellege kutatói döntéstől függ.

A tájékozott olvasó számára feltűnhet a Grounded Theory (GT) módszerének hiánya. A szerzői döntés oka, hogy bár a GT fontos és nagy hatású módszertan, egyrészt nem a 20. század utolsó harmadában alakult ki és nem a 20. század vége felé terjedt el, másrészt bőséges magyar és nemzetközi szakirodalom áll róla rendelkezésre. Kötetünk célja a korszerűbb, az elmúlt évtizedekben nagyobb szerepet kapott módszerek gyakorlati bemutatása, így a már jól ismert eljárások részletes ismertetésétől eltekintettünk.

A QCA és az ENA (V. és VI. fejezetek) alkalmazható strukturált adatok elemzésére közvetlenül, de strukturálatlan adatok strukturálása (jellemzően kódolása) után is. Az NLP (VII. fejezet) a strukturálatlan szöveges adatokat először strukturált, numerikus formátummá alakítja, és a tényleges, jelentős számítási kapacitást igénylő elemzést már ezen a strukturált adatbázison végzi.

Ezt követően a kötet az eleve strukturált adatokkal dolgozó modellezési eljárásokra összpontosít. A sort a megerősítő faktorelemzés (CFA) (VIII. fejezet) nyitja, amely a mérési modellek alapjait fekteti le. Ezt követi az útvonalelemzés (PA) (IX. fejezet), amely a megfigyelt változók közötti oksági utakat modellezi, és amelynek keretében a mediációs modelleket is tárgyaljuk. Ez a két fejezet készíti elő a strukturális egyenletmodellezés (SEM) (X. fejezet) bemutatását, amely a mérési és strukturális szinteket egyesíti. A következő logikai egységet a függő, egymásba ágyazott adatstruktúrák kezelésére szolgáló módszerek alkotják: a beágyazott ANOVA (XI. fejezet) bemutatása készíti elő a terepet a korszerűbb és rugalmasabb lineáris kevert modellek (LMM) (XII. fejezet) megértéséhez. A kötet utolsó fejezete elméleti jellegű (XIII. fejezet), és az eredmények általánosíthatóságának és megbízhatóságának mérésére szolgáló három nagy módszertant, a hagyományos szignifikanciatesztelést, a bootstrapet és a keresztvalidációt hasonlítja össze.

Az egyes fejezetek belső rendjét a gyors tájékozódás, vagyis a kötet kézikönyvként való használatának segítése határozza meg, ezért minden módszer tárgyalása hasonló felépítést követ. A fejezetek rövid bevezetéssel kezdődnek, amit az elméleti alapoknak, a kutatás folyamatának, a minőségbiztosításnak, a módszer változatainak vagy haladó alkalmazásainak leírása követ. A módszereknél ismertetjük a legfontosabb gyakorlati döntési pontok dilemmáit. Ez a szerkezet lehetővé teszi az olvasó számára, hogy gyorsan felmérje, egy adott eljárás illeszkedik-e a saját kutatási kérdéséhez, mintájához és céljaihoz, mielőtt elmélyülne a részletekben.

A kötetben található példák, ábrák, táblázatok, az elemzési lépések sorrendje és a magyarázatok saját szerzői koncepciók. A könyv kialakításakor a gyakorlati alkalmazhatóságot és az érthetőséget helyeztük a középpontba. E didaktikai cél érdekében a szöveg törzse mellőzi a klasszikus elméleti monográfiákra jellemző, az olvasást gyakran megakasztó sűrű szövegközi hivatkozásokat. Célunk, hogy az olvasó figyelmét a módszertani lépések logikai láncolata kösse le. A tudományos visszakövethetőség érdekében azonban az egyes fejezetek végén ajánlottirodalom-jegyzéket helyeztünk el, valamint a könyv végén az összes felhasznált irodalom megtalálható. A módszertani fejezetekhez tartozó frissebb monográfiák és tanulmányok egyrészt bemutatják az adott módszer elméleti alapjait és történeti fejlődését, másrészt ismertetik a jelenlegi kutatási standardoknak megfelelő legújabb technikákat és szoftveres alkalmazásokat.

Végül, az olvasóban joggal merülhet fel a kérdés, hogy egy kutatási módszerekre gyakorlati szempontból összpontosító kötet miért egy látszólag elméleti, közismert problémákat is felvető, viszonylag szétágazó témákat tárgyaló fejezettel, a társadalomtudományi kutatás torzító tényezőinek tárgyalásával kezdődik. Ennek oka, hogy a kötet nem kizárólag kutatói kézikönyvként, hanem doktoranduszoknak

szóló tankönyvként is használható. Továbbá a könyvben bemutatott eljárások hatékony és kifinomult eszközök. Azonban a hatékony és kifinomult eszközök önmagukban még nem garantálják a megbízható és érvényes tudás létrehozását. Egy rosszul feltett kutatási kérdés, egy fel nem ismert zavaró tényező vagy a kutató saját, reflektálatlan kognitív torzításai még a legkorszerűbb elemzési technikával is szükségszerűen téves, félrevezető következtetésekhez vezetnek. Az első fejezet az eljárások helyes kontextusba helyezésében segít. E fejezet célja nem a kutatásmódszertan filozófiai és gyakorlati hátterének teljes körű bemutatása; általa csak megpróbáljuk néhány kutatással kapcsolatos buktatóra felhívni a figyelmet.

Turós Mátyás

Ajánlott irodalom

Csíkos, C. (2020). *A neveléstudomány kutatásmódszertanának alapjai*. ELTE.

Hajtman, E., & Hajtman, B. (2018). *Gyakorló példák a pszichológus hallgatók statisztikai tantárgyainak anyagához*. Akadémiai Kiadó. https://doi.org/10.1556/9789634542513

Hunyadi, L., & Vita, L. (2019). *Statisztika I*. Akadémiai Kiadó. https://doi.org/10.1556/9789634542216

Ivanović, J. (2016). *Bevezetés a pedagógiai kutatás módszertanába*. Újvidéki Egyetem.

Kaposi, J., & Szőke-Milinte, E. (Eds.). (2022). *Kutatási módszerek pedagógusjelölteknek*. Pázmány Péter Katolikus Egyetem.

Kontra, J. (2011). *A pedagógiai kutatások módszertana: Egyetemi jegyzet*. Kaposvári Egyetem.

Rácz, J., Karsai, S., & Tóth, V. (2023). *Kvalitatív pszichológia*. ELTE Eötvös Kiadó.

Sántha, K. (2009). *Bevezetés a kvalitatív pedagógiai kutatás módszertanába*. Eötvös József Kiadó.

Sántha, K. (2015a). *Multikódolt adatok kvalitatív elemzése*. Eötvös József Kiadó.

Sántha, K. (2015b). *Trianguláció a pedagógiai kutatásban*. Eötvös József Kiadó.

Sántha, K. (2022). *Kvalitatív tartalomelemzés*. Eötvös József Kiadó.

Vargha, A. (2015). *Matematikai statisztika – Pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. Pólya.

Vargha, A. (2019). *Többváltozós statisztika dióhéjban*. Pólya.

Vargha, A. (2020). *Normális vagy? És ha nem? – Statisztikai módszerek nem normális eloszlású változókkal pszichológiai kutatásokban*. Pólya.

I. A MEGISMERÉS ÉS TORZÍTÓ TÉNYEZŐI A TÁRSADALOMTUDOMÁNYI KUTATÁSBAN

TURÓS MÁTYÁS

A fejezet a következő kérdéseket válaszolja meg:

a) Miben különbözik a tudományos megismerés a hétköznapi gondolkodástól, és miért van szükség módszertanra a megbízható tudás létrehozásához?
b) Milyen alapvető filozófiai paradigmák (például pozitivizmus, hermeneutika) határozzák meg a társadalomtudományi kutatás céljait és eszköztárát?
c) Miért vezethet félre a kutatási módszerek egyszerűsítő kvalitatív kontra kvantitatív felosztása, és milyen gyakorlatiasabb megközelítés létezik helyette?
d) Hogyan ismerhetjük fel és kezelhetjük a kutatásainkban a rejtett, zavaró tényezőket (confounding factors), amelyek az összefüggések erősségének torz észleletéhez vezethetnek?
e) Milyen megbízható ismérvek alapján tudunk különbséget tenni a tudományos kutatás és az áltudományosság között?

1. Bevezetés

Mindenki alkot elméleteket a világról. Az ember jellemzően személyes tapasztalatai és a környezetében hallott vélemények alapján – különböző írásos forrásokat is felhasználva – próbálja megérteni és megmagyarázni a társas és társadalmi jelenségeket is. Szinte közhely, hogy e törekvésnek egyik kihívása az egyéni tapasztalat és megismerés természete, amely korlátozott és gyakran torzított. Például az egyén hajlamos azokat az eseményeket észrevenni és felidézni, amelyek megerősítik a már meglévő meggyőződéseit, miközben a cáfoló esetek felett átsiklik, vagy a legkönnyebben felidézhető, legélénkebb vagy legutóbbi események jelentőségét hajlamos túlbecsülni. Aki a saját munkahelyén tapasztalja, hogy a fiatal munkatársak kevésbé elkötelezettek, könnyen általánosít egy egész nemzedékre, miközben tapasztalata néhány tucat emberrel való találkozásra korlátozódik. A sikeres vállalkozók tanácsait olvasva hajlamos lehet valaki elfelejteni azt a több ezer embert, akik ugyanazt a stratégiát követve buktak meg, de az ő történetüket senki sem írta meg könyvben. A saját tapasztalat korlátozott térben és időben, szűkös mennyiségben és gyakran kétes minőségben.

A társadalomtudományos módszertan arra szolgál, hogy az egyéni tapasztalatból és a zárt, önigazoló rendszerekből fakadó torzításokat és korlátokat kiküszöbölje, és egy olyan keretrendszert biztosítson, amelyen belül az ember tárgyilagos, megbízható és érvényes tudást hozhat létre. A társadalomtudomány nemcsak egyéni kutatói, hanem társadalmi vállalkozás is. Ha a valóság megismerésének igénye tudományos módszerrel kapcsolódik össze, az jobb eséllyel és hatásfokkal szolgál egyetemes értékeket, mint az esetleg önérdeken nyugvó, de mindenképpen egyéni észlelésen alapuló kinyilatkoztatás. Ennek ellenére hiba lenne az egyéni megismerés értékét alábecsülni. Az egyén egyrészt fejlesztheti megismerését, másrészt sok tudományos vizsgálat egyéni megfigyelésből, megérzésből vagy személyesen megtapasztalt gyakorlati problémából fakad. A tudományos módszertan tehát nem az egyéni tapasztalat és észlelés ellensége, hanem annak kritikus és fegyelmezett társa.

Azonban a társadalomtudományi kutatások minőségét több tényező kikezdheti. A kutatásokat a politika is megérintheti. A különböző ideológiákat vagy érdekeket képviselő csoportok gyakran igényeket fogalmaznak meg, vagy olyan kutatási eredményeket várnak el és támogatnak, amelyek a saját álláspontjukat vagy eszméiket támasztják alá, ezért a szakmában vagy a közbeszédben téves

vagy egyoldalú állítások is tudományos eredményként jelennek meg. Sokan azt gondolják, hogy a tudományos közösség gyorsan és egyértelműen feloldhatja az ellentmondásokat, azonban ez nem így van. A gyors konszenzus kialakulását a szervezeti, ideológiai és anyagi érdekek éppúgy hátráltatják, mint a tudományos rendszer működésének módszertani és publikációs sajátosságai:

a) Egy téves állítás oka nemcsak a szándékos torzítás lehet akár szűk csoport-, akár egyéni érdekek mentén, hanem előfordulhat nem megfelelő módszertan vagy az eredmények téves értelmezése miatt tiszta szándék mellett is. Utóbbi rejtve maradhat a szerző, a terjesztő és a befogadó előtt is.
b) A kutatás folyamata során a kezdeti, kategorikus állítások gyakran finomodnak, és kiderül, hogy egy egyetemesnek hitt megoldás csupán bizonyos esetekben, egyedi célokra alkalmas.
c) A tudományos közösség sem mentes a társadalmi dinamikától. Egy tanulmány hivatkozásainak számát (ami összefügg presztízsével, hírértékével) torzíthatja a szerzők ismertsége, a kölcsönös hivatkozások gyakorlata (szűkebb kutatói csoportok egymásra hivatkozása), valamint az, hogy a kutatók gyakran nem a módszertanilag jobb, hanem a saját nézeteiket alátámasztó tanulmányokat idézik (mikroklímák kialakulása).
d) Habár a folyóiratok és a szerkesztők hangsúlyozzák, a módszertani kurzusok pedig tanítják a null eredmények közlésének fontosságát, a pozitív, feltűnő eredmények publikálási esélyei továbbra is nagyobbak. A szignifikáns hatásokat mutató kutatások gyakrabban jelennek meg, mint azok, amelyek cáfolják a várt összefüggést – a negatív eredmények sokkal ritkábban látnak napvilágot (bár létezik a registered riport intézménye, elenyésző az ilyen formában megjelenő tanulmányok száma). Ennek következtében a szakirodalom torz képet adhat. Úgy tűnik, mintha egy hatás erősebb vagy gyakoribb lenne, mint valójában. Ez félrevezeti a metaanalíziseket, a szakpolitikai döntéseket és az újabb kutatásokat egyaránt. Ráadásul kutatók fölöslegesen ismételnek meg olyan vizsgálatokat, amelyek negatív eredményét mások már korábban megkapták, de nem publikálták.
e) A „publish or perish” kultúra arra készteti a kutatókat, hogy triviális vagy kis jelentőségű, gyengén megalapozott eredményeket közöljenek, amelyeket azonban nagy léptékű állításokként, esetleg áttörésként mutatnak be.

f) Az ipari finanszírozás nyilvánvaló problémákhoz vezet. Például a dohányipar évtizedeken át finanszírozott olyan kutatásokat, amelyek kétségbe vonták a dohányzás egészségkárosító hatását.
g) Bizonyos területeken az ideológiai megfelelés és a bonyolult szakzsargon képes elfedni a tudományosnak álcázott értelmetlenséget.[1]

Ezek a torzítások valósak, mégis az elméletek elfogadása vagy elutasítása – az például a közlés fórumától is függ – hosszú távon nem publikációs torzításokon vagy szociális beágyazottságon múlik. A paradigmaváltó elméletek és áttörő eredmények nem azért váltak meghatározóvá, mert támogatottabbak vagy hangosabbak voltak, hanem mert jobban magyarázták a megfigyelhető jelenségeket, pontosabb előrejelzéseket tettek, és kiállták a módszeres empirikus tesztelést. Wegener kontinensvándorlás-elméletét végül nem a tudományos közhangulat változása tette elfogadottá, hanem a tengerfenék-tágulás felfedezése révén előkerült empirikus bizonyítékok.

Tehát amikor egy kérdésről vallott nézetek ütköznek, és a megszokott vitarendezési módszerek – mint a saját álláspont melletti érvelés vagy a tekintélyre való hivatkozás – nem vezetnek eredményre, és különösen, amikor a szakmai közösség is megosztott, a tudományos módszertan nyújthat kiutat. A tudományos módszer szisztematikus, kritikai keretrendszere és empirikus tesztelése hosszú távon képes túllépni az egyéni vagy csoportérdekeken, és közelebb visz a megbízható tudáshoz.

A tudományos módszer általános értelemben tehát megbízható. De nem egységes. Minden módszer kapcsolható olyan filozófiai kerethez, amely meghatározza, hogyan lehet érvényesen gondolkodni és állításokat megfogalmazni a természet és társadalom természetéről (ontológia), megismerhetőségéről (episztemológia) és vizsgálatának módszereiről (metodológia). A mai napig a társadalomtudományokban három olyan filozófiai előfeltevés-rendszert, három fő hagyományt szokás elkülöníteni és módszertani kötetek elején valamilyen módon ismertetni, amelyek mélyen gyökereznek abban az alapvető emberi tapasztalatban, hogy az ember egyszerre éli meg magát a világot szemlélő, értelmező tudatként (szubjektum) és a világ egy megfigyelhető részeként (objektum):

[1] A „Grievance Studies" botrány ezt bizonyította: egy kutatócsoport szándékosan abszurd, áltudományos cikkeket nyújtott be rangos társadalomtudományi folyóiratokba. A szerkesztők és szakmai bírálók elfogadták és közölték például azt a tanulmányt, amely a kutyák parkbeli viselkedését a „nemi erőszak kultúrájának" bizonyítékaként értelmezte, vagy azt, amely Hitler Mein Kampfjának fejezeteit írta át feminista nyelvezetre.

a) Az externalista (külső) hagyomány, amelynek legismertebb képviselője a pozitivizmus. Ez a természettudományok mintájára képzeli el a társadalomtudományos kutatást. A kutató feladata az ok-okozati összefüggések feltárása. E megközelítés jellemzően matematikai módszereket alkalmaz, és a jelenségek általánosítható magyarázatára törekszik.
b) Az internalista (belső) hagyomány, amelybe a hermeneutikai és fenomenológiai megközelítések tartoznak, azt hangsúlyozza, hogy a társadalmi valóság alapvetően különbözik a természeti világtól. A társadalmi jelenségek ezért emberi értelmezések és jelentések által létrehozott konstrukciók. A cél a jelenségek egyedi megértése, jellemzően hermeneutikai módszerekkel.
c) Az interakcionista vagy pragmatista megközelítés a tudást nem leképződésnek, hanem (főleg problémamegoldó) cselekedetek által létrejött jelenségnek tekinti. E megközelítés nem a külső és belső világ különbségtételéből indul ki, hanem a kettő közötti erőteljes kölcsönhatást, a gyakorlati következményeket helyezi a középpontba.

E tetszetős és jogos felosztás ellenére ezek a paradigmák azonban – gyakorlati kutatói szempontból – inkább analitikus kategóriák. A kutatók valójában ritkán választanak módszertani irányzatot úgy, hogy előbb évtizedeket töltenek több filozófiai iskola tanulmányozásával, egynek igazáról meggyőződnek, majd ennek alapján 60 évesen kutatni kezdenek. A PhD-fokozatot szerzett, az egyetemi oktató, a kutató, a tudós és a filozófus tudása és tevékenysége eltér, és ritkán adódik egyetlen személyben. A kutatókat a gyakorlatban kompetenciáik (jellemzően, hogy idegen-e számukra a statisztika, a programozás), érdeklődésük, sőt ízlésük és szimpátiáik, kutatási kérdésük természete és a rendelkezésre álló erőforrások vezetik el egy-egy megközelítéshez. Ezért a társadalomtudományi kutatást ebben a kötetben *mesterségként* értelmezem, amelynek középpontjában a módszertan ismerete, valamint a módszertani csapdák elkerülésének gyakorlati tudása áll.

Ez a megközelítés nem jelenti azt, hogy a tudományos módszer merev szabályrendszer vagy üres technicitás lenne. A tudományfilozófia radikális kritikája, melyet legélesebben talán Paul Feyerabend és Thomas Kuhn fogalmazott meg, arra is figyelmeztet, hogy a legnagyobb tudományos áttörések gyakran éppen a kor elfogadott módszertani szabályainak megszegésével születtek.[2] Ezt az

[2] A kutatásmódszertan az ellenőrizhetőségre, átláthatóságra, logikusságra, átfogóságra és reprodukálhatóságra törekszik, különösen a természettudományos és kvantitatív paradigmában. Célja a megalapozott és igazolható tudás előállítása (nem feltétlenül bináris igaz/hamis ítéletek formájában). A legtöbb kutatás valószínűségi vagy valószínűsített állításokat produkál (például a statisztikai

ellentmondást itt azzal oldjuk fel, hogy különbséget teszünk a kutatásmódszertan – amely a megbízható tudás létrehozásának jelenlegi legjobb gyakorlatát írja le – és a tudományfilozófia között – amely a tudományos megismerés természetét és végső korlátait vizsgálja. A könyv a továbbiakban ezért a szükséges fogalmi keretek tisztázása után azokat a konkrét módszertani eszközöket mutatja be, amelyekkel a kutató a ma ismert módon megbízható tudást hozhat létre.

2. A hétköznapi megismerés torzításai és a tudományos módszer

Eddig az olvasó kiindulópontot kapott a hétköznapi és a tudományos megismerés közötti különbségről, valamint bemutattam néhány torzítást a tudományon belül. Most vizsgáljunk meg néhány torzítást a hétköznapi megismerésen belül. A hibák forrásai itt jellemzően logikai hibák, kognitív torzítások, félrevezető heurisztikák, valamint társas és motivációs hatások.

Például valaki azt tapasztalja, hogy felgyorsult az algásodás az akváriumában, ezért egy új algamentesítő szert önt a vízbe. Másnapra az aranyhal elpusztul. A hétköznapi következtetés azonnali és magától értetődő lehet: „Az algamentesítő szer ölte meg a halat, soha többé nem használom." Ez a gondolatmenet a post hoc, ergo propter hoc (ezután, tehát emiatt történt) logikai hibán alapul. Figyelmen kívül hagy számtalan egyéb lehetséges magyarázatot, például a hal korát, esetleges rejtett betegségét, a víz jellemzőinek változását.

Hasonlóképpen, ha A egy távoli országról szóló elrettentő élménybeszámolója alapján B eldönti, hogy soha nem utazik oda, egyetlen, szubjektív tapasztalatot általánosít egy egész országra. Ez az elérhetőségi és reprezentativitási heurisztika kombinációja:[3] az élénk, érzelmileg telített történet aránytalanul

szignifikancia ≠ igaz), sőt a hermeneutikai kutatás hitelességről, mélységről, transzferálhatóságról beszél. A popperi tudomány is inkább cáfolható állításokat termel, nem „igaz" állításokat. Bármely paradigmában mozog egy kutatás, a kutatásmódszertan hivatott arra, hogy elválassza a tudományos tudást a hiedelmektől, és eszközöket biztosítson a kutató számára az adatok értelmező, elméleti keretbe helyezett feldolgozásához és kommunikálásához – megkülönböztetve azt a puszta adatközléstől vagy leírás nélküli adatgyűjtéstől.

[3] Szakirodalmi értelemben itt az elérhetőségi heurisztika és a reprezentativitási heurisztika összjátékáról van szó. Az egyén hajlamos a könnyen felidézhető, élénk és érzelmileg telített információkat

nagy súlyt kap az unalmasabb statisztikákhoz képest, és egyetlen esetet a teljes populációra nézve reprezentatívnak tekint.

Hasonlóan, ha egy politikai elemző egy grafikont mutat be arról, hogy egy gazdasági mutató hogyan változott egy kormány hivatalba lépése óta, és ebből azt a következtetést vonja le, hogy a változás a kormány teljesítményének köszönhető, akkor figyelmen kívül hagyja a globális gazdasági trendeket, a korábbi kormány örökségének késleltetett hatásait (a gazdasági intézkedések hatása gyakran csak évekkel később jelentkezik) vagy más egyidejű társadalmi változásokat (lag time effect). Ez az egyszerűsítő ok-okozatiság torzítása, amikor összetett jelenségeket egyetlen tényezővel próbál valaki megmagyarázni. Az egyszerűsítő ok-okozatiság szemléletes ellenpéldája Geoffrey West és munkatársainak kutatása a városi skálatörvényekről. A városméret megkétszerezése minden vizsgált országban egyidejűleg eredményezett ~15%-kal több jövedelmet és ~15%-kal több bűnözést (kitevő: 1,12–1,27). A jelenség matematikai törvényszerűség, nem politikai döntés következménye – a hétköznapi megismerés mégis hajlamos egyetlen tényezőnek (például egy kormánynak) tulajdonítani az ilyen összetett összefüggéseket.

Ha a hírek tele vannak gyilkosságokkal, azt gondolhatja valaki, hogy az a gyakoribb bűncselekmény, még ha a statisztikák szerint a betörések száma jóval magasabb is. Ez az elérhetőségi heurisztika, amely miatt hajlamos lehet az egyén felülbecsülni azoknak az eseményeknek a jelentőségét, amelyekről könnyen lehet példát felidézni.

A hétköznapi megismerés azonban nem mindig megbízhatatlan, és sok esetben a tudományos módszerre való csere nem gazdaságos vagy nem is észszerű. Például egy tapasztalt erdész ránézésre megmondhatja, hogy egy adott fa egészséges-e, vagy érdemes-e kivágni, figyelembe véve a helyi mikroklímát, talajviszonyokat és az erdő ökoszisztémájának egyedi jellemzőit. Egy általános tudományos protokoll ezeket az egyedi kontextuális tényezőket csak nehézkesen vagy jelentős erőforrás-ráfordítással tudná kezelni. A helyzet természetesen gyökeresen megváltozik, ha nem egy fáról, hanem százról, több ezerről vagy több millióról kellene dönteni. Ezekben az esetekben a hétköznapi megismerés skálázhatósága korlátokba ütközne, és a feladat léptékét is figyelembe kellene venni a módszer megválasztásakor.

Fontos azonban, hogy amikor a hétköznapi megismerés hatékonyságát vagy megbízhatóságát állítjuk, ezt magát is érdemes ellenőrizni. Például ha valaki azt állítja, hogy egy tapasztalt ápoló megérzi a beteg állapotromlását, mielőtt

(mint egy ismerős elrettentő beszámolója) fontosabbnak vagy gyakoribbnak ítélni, mint a statisztikai adatokat, és egyetlen esetet a teljes populációra nézve reprezentatívnak tekinteni.

a műszerek jeleznék, a tudomány metaszinten kompetens marad: eldöntheti, hogy az ápoló megérzései szignifikánsan jobbak-e a véletlennél. Ha igen, a hétköznapi megismerés működik, de ezt is a tudomány állapítja meg. Ha nem, megerősítéses torzításról van szó. A tapasztalt megfigyelő intuíciójának korlátait jelzi Barabási és munkatársainak mobiltelefon-alapú mozgásvizsgálata is. Ötvenezer felhasználó hat hónapos adatai alapján a következő tartózkodási hely 93%-os pontossággal megjósolható volt, miközben az érintettek szabadnak érezték döntéseiket. Ez arra utal, hogy amit a hétköznapi megfigyelő megérzésnek vagy tapasztalatnak tulajdonít, az statisztikailag leírható.

A hétköznapi megismerést torzíthatja továbbá:

a) a motivált gondolkodás, vágyteljesítő gondolkodás (azt hiszem el, amit hinni akarok);
b) a társas hatások: konformitás, tekintélytisztelet, csoportgondolkodás;
c) az érzelmi állapotok hatása a következtetésekre;
d) a kulturális előfeltevések, amelyeket észre sem veszünk;
e) a korlátozott információhoz való hozzáférés (szűrőbuborékok, médiaszelekció).

A hétköznapi megismerés ugyanakkor természetes és indokolt, mert ez teszi lehetővé a hatékony mindennapi cselekvést. A gondolkodási rövidítések gyorsak és általában hasznosak, mivel a mindennapi életben nincs idő és mentális erőforrás arra, hogy minden állításunkat a tudományos kutatás alaposságával indokoljuk. A mindennapi kommunikációban természetes és többnyire értelmes is, ha gyorsan, intuitívan és röviden közöljük álláspontunkat, véleményünket. A hatékonyságnak azonban ára van. A gondolkodási rövidítések gyakran szisztematikus hibákhoz vezetnek.

Ezzel szemben a tudományos megközelítés a pontosságra, a körülhatároltságra és a kritikai alapállásra épül. Kerüli a pontatlan és általános állításokat. Például „A digitalizáció a megoldás!" jelszava helyett olyan pontos kutatási kérdésekre épülő vizsgálati eredmények alapján fogalmaz meg állításokat, mint: „Milyen összefüggés mutatható ki a Redmenta platformon végzett, heti rendszerességű, játékosított matematikai gyakorlás és a 7. osztályos diákok órai aktivitása, valamint a matematika tantárgyhoz fűződő attitűdjének változása között egy féléves vizsgálati időszak alatt?" Egy ilyen megfogalmazás már hordozza a tudományos gondolkodás ismérveit. Pontos fogalomhasználatot, körülhatárolt érvényességi kört és ellenőrizhető állításokat feltételez.

1. doboz: A kutatási folyamat logikai lépései egy elképzelt digitális oktatási vizsgálatban

„A digitalizáció a megoldás" jelszavával kapcsolatban a hétköznapi megközelítés talán így hangzana: „A digitalizáció megoldja az oktatás problémáit. A gyerekek szeretik a kütyüket, biztosan jobban tanulnának tabletről." Ezzel a kérdéssel és a rá adott rövid válasszal szemben a tudományos tevékenység keretein belül körülbelül, a példában elnagyolva, ezeket a lépéseket tennénk meg:

a) Elméleti keretet keresnénk, vagyis figyelembe vennénk az összes eddig létrehozott tudást a témában.
b) Megkeresnénk a kutatási rést, vagyis azt, ami még feltáratlan a témában, amit mások nem vizsgáltak, vagy amiben ellentmondás mutatkozik az eddigi eredmények között.
c) Pontosan meghatároznánk, hogy mit értünk „digitalizáció" és „oktatási problémák" alatt.
d) Meghatároznánk az alapsokaságot és a mintát, a mintavétel módszerét, a példa esetében kontrollcsoportos kísérletet végezhetnénk.
e) Mérnénk a tanulási eredményeket – nem csak a szubjektív tetszést –, figyelembe vennénk a zavaró tényezőket (például újdonság varázsa, tanári attitűd).
f) Elfogadott, biztos módszerekkel dolgoznánk, eredményeket hoznánk létre.
g) Eredményeinket összevetnénk mások eredményeivel.
h) Deklarálnánk, hogy hol mik a munkánk korlátai (például a módszerből, a minta méretéből, az adatfelvétel torzításaiból, nem vizsgált tényezőkből stb. fakadóan).
i) Következtetéseket fogalmaznánk meg.

A tudományos megismerés tehát egy gondolkodási fegyelmezettségre, a szubjektivitás megszüntetésére, de legalábbis korlátok közé szorítására vagy átláthatóvá tételére törekvő tevékenység. Nemcsak az előbbi hibák kiküszöbölésében, hanem abban is különbözik a hétköznapi megismeréstől, hogy a) alapvetően kollektív természetű (a jól ismert szentencia szerint minden kutatás mások munkájának folytatása és egyben előzménye); b) a szisztematikusság elvét követi (például nem véletlenszerű eseteket, hanem reprezentatív vagy célzott mintát vizsgál); c) nem az igazolást, hanem a cáfolat lehetőségét keresi (nem részrehajló).

Vagyis nem azt kérdezi: „Az én ismerőseim mit gondolnak?", hanem: „Egy megfelelően kiválasztott minta mit mutat?" Továbbá nem azt kérdezi: „Hogyan bizonyíthatnám, hogy igazam van?", hanem: „Mi cáfolhatná az eredményeimet?" Végül nem azt mondja: „Higgyétek el nekem!", hanem: „Így csináltam,

ellenőrizzétek!" Ez a nyitottság és ellenőrizhetőség az, ami megkülönbözteti a tudományt nemcsak az egyéni intuíciótól, hanem szervezett, zárt, önigazoló rendszerektől is.

Ugyanis az esetleges, egyéni, hétköznapi gondolkodás korlátain kívül létezik még csapda a megismerés számára: a szervezett, de nem tudományos megismerési rendszer. Például a 19. században virágzó frenológia tana szerint a koponya formájából, a dudorokból és mélyedésekből kiolvasható az ember jelleme, tehetsége és hajlamai. Ez a rendszer nem csupán egyének ötletelése volt, hanem szervezett mozgalomként működött, saját szakirodalommal, képzésekkel és praktizáló szakértőkkel. Intuitíven vonzónak tűnt – hiszen az agy a koponyában van, így logikusnak látszott, hogy annak formája leképezi a belső tulajdonságokat –, és a frenológusok számtalan esettanulmánnyal támasztották alá állításaikat, gondosan feljegyezve azokat a példákat, amelyek megerősítették elméletüket.

Az elméletet végül nem egy másik, hasonlóan anekdotikus rendszer buktatta meg, hanem a tudományos módszertan alkalmazása. Az agysérülések következményeinek szisztematikus vizsgálata, majd később a pszichológiai mérések (például intelligencia, személyiségjegyek) és a koponyaalak közötti korreláció hiánya egyértelműen kimutatta, hogy a frenológia állításai alaptalanok. A tudomány itt nem egy egyéni vélekedéssel, hanem egy kollektív, intézményesült tévhittel szemben bizonyult hatékonyabbnak.

Míg a frenológia ma már legfeljebb tudománytörténeti érdekesség, a mögötte húzódó gondolkodási mintázatok és kognitív torzítások, heurisztikák és szisztematikus hibák egyáltalán nem tűntek el. Például a frenológiához hasonlóan az asztrológia is azt ígéri, hogy egy külső, objektívnek tűnő tényezőből (a születés pillanatában látható csillagállásból) képes levezetni belső, személyes tulajdonságokat és életeseményeket. A rendelkezésre álló bizonyítékok elsöprő többsége szerint azonban nem létezik statisztikailag szignifikáns és megismételhető korreláció egy személy csillagjegye és a viselkedése, személyiségjegyei vagy életeseményei között. Az egyik leghíresebb vizsgálat az 1985-ben a Nature folyóiratban publikált Carlson-tanulmány. Egy kettős vak kísérletben neves asztrológusoknak kellett személyiségprofilokat párosítaniuk a résztvevők születési horoszkópjával. Az asztrológusok nem teljesítettek jobban a puszta véletlennél. A találati arányuk körülbelül 1/3 volt, ami az adott kísérleti elrendezésben pontosan a véletlennek, a tippelésnek felelt meg. Geoffrey Dean és Ivan Kelly 2003-as metaanalízise még tovább ment, több száz, az asztrológiával foglalkozó tudományos vizsgálat eredményeit elemezték. Az eredmények szerint a csillagjegyek és az emberi tulajdonságok közötti korreláció hatásnagysága (a kapcsolat erőssége) gyakorlatilag nulla. Természetesen statisztikai értelemben a nulla közeli korreláció nem minden esetben jelenti a hatás teljes

hiányát. Léteznek nemlineáris összefüggések (például egy fordított U alakú kapcsolat, ahol a szélső értékek kioltják egymást a képletben) vagy olyan rejtett interakciók, amelyek csak bizonyos alcsoportokban érvényesülnek. A lineáris (Pearson-féle) korreláción túl több eszköz is létezik nemlineáris összefüggések kimutatására. A Spearman-ρ és Kendall-τ monoton, de nem feltétlenül egyenes kapcsolatokat azonosít; a távolsági korreláció bármilyen függőséget jelez (ahol a 0 valódi függetlenséget jelent); a maximális információs együttható (MIC) pedig akár ciklikus vagy többágú mintázatokat is felismer. A modern vizsgálatok ezeket az eszközöket szoftveresen kombinálják, sőt gépi tanulási algoritmusokat (Random Forest, neurális hálók) is alkalmaznak – amelyek a legrejtettebb nemlineáris összefüggéseket is képesek lennének detektálni. Azonban az asztrológia és a frenológia esetében a modern, többváltozós és nemlineáris modellekkel végzett vizsgálatok sem tudtak a véletlennél (zajnál) erősebb szisztematikus mintázatot kimutatni. Továbbá számtalan kutatás vetette össze a tudományosan elfogadott személyiségtesztek (mint például a Big Five-modell) eredményeit az emberek csillagjegyeivel. Az eredmény következetesen negatív volt, tehát az extrovertáltak, lelkiismeretesek vagy neurotikusok aránya nem magasabb egyetlen csillagjegyben sem.

Akkor miért érzi mégis oly sok ember, hogy a saját horoszkópja „működik" és ráillik? Elsősorban azért, mert az ember hajlamos az általános, homályos és szinte mindenkire igaz személyiségleírásokat nagyon pontosnak és egyedül rá jellemzőnek elfogadni (ez az úgynevezett Barnum-effektus[4]). Egy tipikus Barnum-állítás: *„Bár külsőleg fegyelmezettnek tűnsz, belül hajlamos vagy a bizonytalanságra."* Ez az állítás a népesség nagy többségére igaz, sokan mégis, illetve talán pont ezért személyesnek és találónak érzik. A horoszkópok tele vannak ilyen állításokkal. Továbbá a megerősítési torzítás arra késztet, hogy azokat az információkat keressük, vegyük észre és idézzük fel, amelyek megerősítik a hiedelmeinket. Ha Péter elhiszi, hogy a Bikák makacsok, és Péter Ádám nevű, Bika jegyű barátja egyszer makacskodik, Péter azt fogja gondolni: *„Látod, Ádám tipikus Bika!"* Amikor viszont Ádám alkalmazkodó, meggyőzhető, azt Péter egyszerűen nem fogja Ádám csillagjegyéhez kötni. Végül, a horoszkópok nyelvezete szándékosan általános és többértelmű (*„egy kihívás vár a munkahelyeden"*), ami szinte bárkivel megtörténhet. Emellett ha valaki elhiszi, hogy ő egy „szenvedélyes

[4] Fontos különbséget tenni a Barnum-állítás és a Barnum-effektus (vagy Forer-effektus) között. Az állítás maga egy olyan általános leírás, amely szinte bárkire igaz. Az effektus pedig az a pszichológiai jelenség, amelynek során a befogadó az ilyen homályos leírásokat kifejezetten önmagára nézve egyedinek és pontosnak fogadja el. A jelenség nevét P. T. Barnum showmanről kapta, aki szerint a jó cirkuszban mindenki talál valamit a maga számára – a Barnum-állítások pedig pontosan így működnek a személyiségleírások-ban.

Kos", az ösztönözheti arra, hogy valóban így viselkedjen, ezzel megerősítve önmagában a hiedelmet.

A két példa – a frenológia és az asztrológia – rávilágít, hogy a tudományos módszertan egyedisége nem csupán a szervezettségében rejlik, hanem abban a specifikus cél- és eszközrendszerben, amelyet a megbízható és általánosítható tudás létrehozására fejlesztett ki, és amely képes kiszűrni az egyéni és a kollektív emberi gondolkodás rendszerszintű hibáit is.

A fenti gondolatmenet további tanulságokkal is szolgál. A példák alapján nem gondolhatjuk, hogy minden, amit a tudomány egy adott korban nem képes kimutatni, szükségszerűen hamis. Elképzelhető, hogy léteznek olyan valós jelenségek, amelyek méréséhez a kor tudományos eszköztára még nem elég fejlett. Ahogy a baktériumokat sem lehetett közvetlenül megfigyelni a mikroszkóp feltalálása előtt, úgy lehetséges, hogy léteznek olyan finom összefüggések, amelyek kimutatásához a mai módszerek nem megfelelők. A tudományos hozzáállás azonban itt sem a hiten alapuló elfogadás, hanem a módszerek folyamatos fejlesztése iránti elköteleződés. Ha egy hatás valós, akkor az egyre kifinomultabb és erősebb mérőeszközöknek előbb-utóbb képesnek kell lenniük kimutatni azt. Ha azonban egy jelenséget évszázadokon át, egyre fejlettebb módszerekkel sem sikerül igazolni – mint az asztrológia esetében –, akkor a racionális következtetés az, hogy a jelenség nem létezik.

A tudomány egyedisége tehát nem abban áll, hogy sosem téved, hanem abban a képességében, hogy hosszú távon képes felismerni és javítani saját tévedéseit és megkülönböztetni a termékeny hipotéziseket a bizonyítékokkal alá nem támasztott állításoktól. Ezzel szemben az áltudományos elméletek évtizedek óta ugyanazokat a cáfolt, alátámaszthatatlan vagy alá nem támasztott állításokat ismétlik, képtelenek prediktív erejű állításokat megfogalmazni, és amikor az empirikus tesztek cáfolják őket, összeesküvés-elméletekbe menekülnek. A tudományos szkepticizmus tehát nem önkényes kapuőrködés, hanem a megbízható tudás létrehozásának feltétele. A replikációs válság nyílt elismerése a pszichológiában és más területeken, az open science mozgalom térnyerése, az előregisztráció gyakorlatának terjedése, a megismétlési kísérletek támogatása mind azt mutatják, hogy a tudományos közösség – ha néha lassan is, de – képes az önkorrekcióra.

3. Tudományos és nem tudományos szabványok

Az empirikus tudományok állításai tényeken alapulnak, nem spekuláción vagy tekintélyi érvelésen. Ez természetesen az empirikus kutatásokban és azokon kívül sem jelenti az elméletalkotás kizárását, de az elméleteknek kapcsolódniuk kell megfigyelhető jelenségekhez. Ezzel kapcsolatban a következőkben felsorolunk néhány – de nem minden tudományfilozófiai iskola által egyöntetűen elfogadott – kutatási szabványt:

a) Empirikus tesztelhetőség és cáfolhatóság (falszifikálhatóság). Egy tudományos állításnak olyannak kell lennie, hogy elvileg meg lehessen cáfolni tapasztalati bizonyítékokkal. Olyan kísérletet vagy megfigyelést kell tudnunk tervezni, amelynek elképzelhető olyan kimenetele, amely ellentmond az állításnak. A közgazdaságtan klasszikus modelljének azon tételét, miszerint a „racionális döntéshozó" mindig a profitmaximalizálásra törekszik, cáfolhatjuk olyan kísérleti helyzetekkel, amelyekben az emberek viselkedése szisztematikusan eltér ettől a modelltől (például az altruista vagy méltányos viselkedést választják a nagyobb anyagi haszonnal szemben). Az állítás, hogy „a rövid távú memóriában egyszerre legfeljebb hét (± kettő) információegységet tudunk tárolni" (Miller mágikus száma), cáfolható olyan kísérletekkel, amelyekben a résztvevők szisztematikusan kevesebb (például komplex vizuális minták esetén) vagy több (például „tömbösítéssel") egységet képesek felidézni. Vagy az állítás, hogy „a demokráciák nem háborúznak egymással" (demokratikus béke elmélete), cáfolható akár egyetlen, egyértelműen demokratikus államok között kitörő háború történelmi példájával. A kutatói vita így nem a tényről, hanem a fogalmak (például „mi számít demokráciának?", „mi számít háborúnak?") definíciójáról folyik. E példák a falszifikálhatóság elvét illusztrálták.

b) Reprodukálhatóság és nyilvánosság. A kutatást úgy kell dokumentálni, hogy más, független kutatók is képesek legyenek megismételni a kísérletet vagy elemzést. Az eredményeknek interszubjektíven ellenőrizhetőnek kell lenniük, és a tudományos közösség számára (például lektorált publikációkban) elérhetővé kell tenni őket. Egy új gyógyszer

hatékonyságát vizsgáló tanulmánynak részletesen le kell írnia a résztvevők kiválasztását, az adagolást, a mérés módját, hogy más laboratóriumok is megismételhessék a vizsgálatot.

c) Objektivitásra való törekvés. A tudományos módszerek célja a kutatói szubjektivitás, az előítéletek és a személyes meggyőződések torzító hatásának minimalizálása. Ezt szolgálja a szakmai lektorálás (peer review) rendszere, ahol a tudományos közösség alapos, elfogulatlan és kritikai értékelésnek veti alá az eredményeket. Egy kutató, aki egy általa kifejlesztett oktatási módszer hatékonyságát vizsgálja, vak vagy kettős vak elrendezést használ, hogy a saját elvárásai ne befolyásolják az eredmények értékelését.

d) Parszimónia (Ockham borotvája). Ha két vagy több elmélet is képes ugyanazt a jelenséget megmagyarázni, a tudomány azt részesíti előnyben, amelyik a kevesebb feltételezéssel él, vagyis az egyszerűbb. Ha egy beteg meggyógyul, az egyszerűbb magyarázat az, hogy az immunrendszere vagy a kapott, bizonyított hatású gyógyszer győzte le a kórokozót, nem pedig az, hogy egy másik galaxisból érkező kozmikus energia segített neki. Az utóbbi feleslegesen sok új, bizonyítatlan feltételezést vezet be.

e) Konzisztencia és fejlődőképesség. Egy új tudományos felismerésnek összhangban kell lennie a tudomány más, jól megalapozott területeivel, és idővel képesnek kell lennie újabb bizonyítékok beépítésére, finomodásra. A tudományos tudás folyamatosan fejlődik, nem megváltoztathatatlan. A kvantummechanika nem tette érvénytelenné Newton törvényeit, csupán pontosította azok érvényességi körét (a hétköznapi méretekben továbbra is tökéletesen működnek), és kiterjesztette a tudást az atomi szintekre.

A tudományos kutatás eredményének tehát megbízhatónak, ellenőrizhetőnek és nyilvánosnak kell lennie. Amikor ezek a szabványok sérülnek, vagy csupán látszólagosan teljesülnek, áltudományosságról beszélünk. Az áltudományosság nélkülözi a tudomány módszertani szigorát és kritikai szellemét. Jellemzői a következők:

a) Cáfolhatatlan hipotézisek. Olyan komplex és rugalmas elméletek, amelyek bármilyen megfigyelést képesek utólag megmagyarázni, így lehetetlenné teszik az empirikus cáfolatukat. Például Karl Popper szerint ilyen a klasszikus pszichoanalitikus elmélet, ahol egy páciens viselkedése

– legyen az a terapeutával való egyetértés vagy azzal való szembeszegülés („ellenállás") – ugyanannak az elfojtott tudattalan konfliktusnak a bizonyítékaként értelmezhető. Az elmélet minden lehetséges kimenetelt képes bekebelezni, így nem tesztelhető.

b) Immunitás a kritikára. Az ellentmondó bizonyítékokat nem az elmélet felülvizsgálatára ösztönző anomáliaként, hanem a kísérleti módszer hibájaként, a kutató rosszindulatának jeleként vagy a jelenség „túl törékeny" voltával magyarázzák. Egy vitatott terápiás módszer (például a múltban a „facilitált kommunikáció") hatástalanságát kimutató, szigorúan kontrollált kísérletekre a módszer hívei azzal reagáltak, hogy a „laboratóriumi körülmények mesterkéltsége" és a „szkeptikus kutatók jelenléte" tette tönkre a finom terápiás kapcsolatot, ezért nem működött a módszer.

c) A bizonyítási teher áthárítása. A módszer hívei nem mutatnak be pozitív bizonyítékokat saját állításaik alátámasztására, hanem elvárják a kritikusoktól, hogy ők bizonyítsák be minden kétséget kizáróan, hogy az állítás hamis. Egy új energiagyógyászati technika támogatója bizonyítékok helyett azt mondhatná: „Több százan számoltak be javulásról. Amíg a szkeptikusok nem tudják bebizonyítani, hogy ez az energia nem létezik, addig nincs okunk kételkedni a hatásosságában."

d) Mítoszokra és tekintélyre hivatkozás. Empirikus adatok helyett gyakran hivatkoznak ősi szövegekre, karizmatikus alapítók megkérdőjelezhetetlen tanításaira vagy homályos bölcsességekre, amelyeket a modern tudomány állítólag elvesztett. Például egy menedzsmenttanácsadó cég egy új vezetési modellt az „ősi szamuráj harcosok titkos kódexére" alapozza ahelyett, hogy modern szervezetpszichológiai kutatásokkal támasztaná alá annak hatékonyságát a mai vállalati környezetben.

e) Szelektív adatkezelés (mazsolázgatás). A támogató bizonyítékokat kiemelik és széles körben terjesztik, míg a negatív vagy ellentmondásos eredményeket figyelmen kívül hagyják, elhallgatják, vagy módszertani hibákra hivatkozva elvetik. Egy „agytréning" szoftvert forgalmazó cég a honlapján kizárólag a saját, belsőleg végzett kutatásai pozitív eredményeit és a lelkes felhasználói visszajelzéseket publikálja, miközben nem tesz említést azokról a független, lektorált tanulmányokról, amelyek nem találtak szignifikáns, a mindennapi életre átvihető hatást.

f) Változatlanság. Az elmélet központi tételei évtizedek, akár évszázadok alatt sem változnak, a felmerülő anomáliák nem vezetnek az elmélet finomításához vagy felülvizsgálatához. Jó példa erre a Myers–Briggs-típusindikátor (MBTI) személyiségmodell, amelynek alapjai évtizedek

óta lényegében változatlanok annak ellenére, hogy a modern pszichometria számos komoly kritikát fogalmazott meg a modell megbízhatóságával és validitásával kapcsolatban. Eközben a tudományos konszenzus által elfogadott Big Five modell folyamatos kutatás és empirikus fejlődés tárgya.

g) Grandiózus, lehengerlő állítások minimális/nulla bizonyítékkal. Az állítások aránytalanul nagyok és általánosak a mögöttük álló, szűkös vagy gyenge minőségű bizonyítékokhoz képest. Gyakran ígérnek forradalmi áttörést vagy egyszerű megoldást összetett problémákra, megfogalmazásuk hatásvadász. Például azok a népszerű menedzsmentkönyvek, amelyek „a siker három egyszerű titkát" ígérik, és ezt néhány sikeres cégvezetővel készített, anekdotikus interjúval támasztják alá, figyelmen kívül hagyva a több ezer sikertelen vállalkozást, amelyek esetleg ugyanezeket az elveket követték.

A tudomány és az áltudomány közötti határvonal azonban nem mindig éles. Gyakran előfordul, hogy valaki a médiában vagy egy népszerű előadásban leegyszerűsítő, hatásvadász kijelentéseket tesz, ez azonban még nem áltudomány. Példa erre a következő, egy interjúban elhangzott kijelentés. „Kérdés: Mit tanácsolna egy mai középiskolásnak? Válasz: Azt, hogy öt dologra lesz szüksége a következő évtizedekben: értő olvasásra, írásra, elemi számolásra (nem a rengeteg marhaságra, amit matematika címén tanítunk ma), számítógéphasználatra és angol nyelvre. Fontos lesz a kritikai érzék és a kreativitás. Amit ma követelnek az iskolában, az sajnos minden kreativitást megöl." Ez az állítás grandiózus, primitivizáló és receptszerű, mégsem teljesen áltudományos, mivel egyes elemei mögött komolyan vehető szakirodalom áll. A leegyszerűsítés itt retorikai eszköz: célja a provokálás, a közgondolkodás formálása. A kontextusából kiragadott üzenet azonban elveszíti a teljes kép árnyalatait, és a leegyszerűsítés itt is együtt jár a bizonyítékok válogatásával. Összetett társadalmi kérdésekben (mint az oktatás vagy a gazdaságpolitika) a tudományos irodalom ugyanis ritkán foglal el egységes álláspontot. Tele van vitákkal, ellentmondó eredményekkel és eltérő értelmezésekkel, néhol szilárd bizonyítékokkal. Amikor valaki ebből csak a saját álláspontjának megfelelő eredményeket emeli ki – figyelmen kívül hagyva az ellenérveket, a limitációkat és a tudományos vita egészét –, akkor nem tudományos ismeretterjesztést, hanem politikai tevékenységet végez. Nem érvek ütköztetésén alapuló, óvatos megállapításokat tesz, hanem kinyilatkoztat, ami a propaganda módszereihez közelít. Az idézet mögött két évtized oktatási diskurzusa áll: előbb a finn modell (amelynek sikerét puszta korrelációból vezették le, s amelynek eredményei azóta jelentősen visszaestek),

majd a 4K-modell (kritikus gondolkodás, kreativitás, kommunikáció, kollaboráció) – mindkettő jól hangzó, de gyengén megalapozott recept. Önellentmondás az is, hogy a kreativitás és a kritikai gondolkodás fontosságát hangsúlyozó üzenetek nemritkán dogmatikus, néhány pontból álló receptként jelennek meg, miközben épp a központosított, uniformizált gondolkodást kritizálják. Az ilyen tevékenység talán még az egyértelmű áltudománynál is veszélyesebb: részigazságokat tartalmaz, a tudományos tekintély látszatával, így hihetőbb, mint egy nyíltan ezoterikus tan. Hatása a tudományos vitában azonban paradox módon csekély: a hergelő, kinyilatkoztató hang és a szelektív érvelés miatt a szakma komolyabb szereplői nem reagálnak rá. Így nem ott fejti ki hatását, ahol a felvetett kérdés eldőlhetne, hanem a közbeszédben, ahol inkább bizonytalanságot szít, mint tájékozottságot.

Fentiek nem jelentik azt, hogy a tudományban nincs helye az értékelvű álláspontoknak. A tudomány szerepe nem merül ki a puszta tényfeltárásban, a kérdések pártatlan megválaszolásában. A tudományos diskurzusban helyet kap a normatív gondolkodás is, vagyis megfogalmazható, hogy a rendelkezésre álló bizonyítékok alapján milyen irányba lenne érdemes haladni. Tehát a tudományon belül is lehet értékelvű álláspontot képviselni, de csak a módszeres szigor és a párbeszédre való nyitottság megőrzésével. Továbbá a kritikus gondolkodás nem áll meg az áltudomány felismerésénél. A tudomány képviselőjének – és ideális esetben befogadójának – azt is meg kell tudnia különböztetni, hogy egy állítás mögött tudományos alapokon nyugvó információ, bizonyíték nélküli kijelentés vagy egy összetett tudományos vita egyoldalú bemutatása áll. Mindez a tudományos írástudás olyan része, amely ismereteket és tapasztalatot egyaránt igényel.

2. doboz: A dihidrogén-monoxid esete

Az áltudományosság megtévesztő erejét kiválóan szemlélteti a híres dihidrogén-monoxid álhír. Egy tudományos versenyen egy diák felhívást indított a „dihidrogén-monoxid" nevű veszélyes vegyület betiltásáért az alábbi, tudományosan helytálló érvekkel: a) Túlzott izzadást és hányást okozhat. b) A savas eső fő komponense. c) Gáz halmazállapotban súlyos égési sérüléseket idézhet elő. d) Hozzájárul a talajerózióhoz. e) Számos fém korrózióját felgyorsítja. f) Rákos daganatokban is kimutatható.

A megkérdezett 50 emberből 43 támogatta a betiltást, és csupán egyetlen ember jött rá, hogy a dihidrogén-monoxid (H_2O) valójában a víz.

Hogyan védekezhetünk az áltudomány és a rossz minőségű információk ellen? Hasznos eszköz lehet egy kritikai kérdéslista, amely segít gyorsan felmérni egy forrás hitelességét. Mielőtt egyáltalán mérlegelnénk egy állítás tartalmát, tegyük fel a következő kérdéseket:

A szerzőről és a forrásról:

a) Ki a szerző? Ismert, elismert szakértő az adott területen?
b) Milyen oldalon vagy folyóiratban jelent meg a cikk?
c) Van impresszum és felelős kiadó?

A hivatkozásokról és adatokról:

a) Hivatkozik a szöveg forrásokra?
b) Ezek a források léteznek, tudományosak, és valóban a tárgyhoz kapcsolódnak?
c) A számadatok ellenőrizhetők?

Az érvelés stílusáról:

a) Az érvelés észszerű és következetes, vagy inkább érzelmekre támaszkodik?
b) Szenzációhajhász a megfogalmazás?

A szöveg szándékáról:

a) A szöveg célja a tárgyilagos tájékoztatás, vagy inkább meg akar győzni valamiről?
b) Terméket akar eladni vagy ideológiát népszerűsíteni?

A tudományos módszertanról:

a) Van tapasztalati bizonyíték az állítások mögött?
b) Megadják a kutatás pontos módszerét?

c) Elismerik a korlátokat és ellentmondásokat?
d) Változhatnak az állítások új bizonyítékok fényében, vagy deklarációval, kizárólagosságokkal állunk szemben?

A tudományos közlést óvatos állítások, a bizonyítékok bemutatása, a módszertan ismertetése és a korlátok elismerése jellemzi. Ezzel szemben a nem tudományos közlést bombasztikus jelenségek, ígéretek, anekdotikus „bizonyítékok", titokzatosság és a kritika elutasítása jellemzi.

A fenti kérdések hatékony szűrőként működnek a mindennapi tájékozódáshoz, laikusok számára is, aki azonban mélyebb forráskritikát kíván végezni, annak a források minőségi különbségeinek megalapozott megítélésén és a tudományos közlésen belüli torzítások felismerésén keresztül érdemes jobban megismernie a tudományos publikálás világát, például a következők szerint.

4. Torzítások a tudományos publikálásban

Amikor valaki haladó forráskritikát gyakorol, tisztában kell lennie a tudományos publikálás nagyságrendjével és hierarchiájával is. A világon több mint 100.000 olyan folyóirat létezik, amelyek tudományos publikációkat jelentetnek meg, de ezek színvonala drámaian eltérő. Pusztán az a tény, hogy valami „tudományos" lapban jelent meg, már önmagában minőségi ugrást jelent a politikai kommunikációhoz vagy a bulvársajtóhoz képest. Azonban a tudományon belül is óriási szakadékok léteznek.

A nemzetközi tudományos életben a minőség egyik biztosítéka az adatbázisokba való bekerülés (indexáltság). A legfontosabb szűrőt a Scopus és a Web of Science adatbázisok jelentik.[5] Az összes folyóiratból a Scopus körülbelül 27.000-28.000, míg a Web of Science Core Collection (SCIE, SSCI, A&HCI együtt) 21.000-22.000 lapot listáz. Az ESCI további 8.000-9.000 lapot tartalmaz. A két adatbázis között jelentős, de nem teljes átfedés van, továbbá mindkét adatbázison

[5] Ez az állítás a legtöbb természet- és társadalomtudományra érvényes, azonban területspecifikus eltérések léteznek. Az orvostudományban a PubMed/MEDLINE, a matematikában a MathSciNet, a műszaki tudományokban az IEEE Xplore ugyanilyen vagy nagyobb jelentőségű lehet.

belül létezik rangsor. A Scopus adatbázison belül az úgynevezett kvartilis besorolás (Q1–Q4), amely az adott folyóirat idézettségi hatását és presztízsét mutatja tudományterületenként:

a) Q1 (felső 25%): a világ élvonalába tartozó, legmagasabb presztízsű folyóiratok (ezen belül a felső 10%: D1);
b) Q2–Q3: a középmezőny, megbízható tudományos sztenderdekkel;
c) Q4: az indexált lapok alsó negyede.

A Web of Science adatbázison belül más hierarchia érvényesül, amely indexek szerint különít el minőségi szinteket:

a) Core Collection indexek (SCIE, SSCI, A&HCI): A legmagasabb presztízsű indexek, szigorú befogadási kritériumokkal. Az itt szereplő folyóiratok általában magas impaktfaktorúak és széles körben idézettek.
b) ESCI (Emerging Sources Citation Index): Újabb vagy regionális fontosságú folyóiratok, alacsonyabb presztízs, gyakran váróterem a Core Collection felé.
c) További indexek: Conference Proceedings Citation Index, Book Citation Index.

Ugyanaz a folyóirat lehet egyszerre Scopus Q1 besorolású és WoS ESCI indexált, míg más esetben lehet Scopus Q2-Q4 és egyidejűleg WoS Core Collection (SCIE/SSCI) indexált, vagyis a két rendszer különböző szempontokat értékel.

A helyzetet tovább bonyolítja, hogy a tudományos világban több száz különböző index és adatbázis létezik, amelyek közül sok nem rendelkezik minőség-ellenőrzéssel. A predátor és alacsony minőségű folyóiratok gyakran kihasználják ezt: honlapjukon hosszú listát tüntetnek fel olyan indexekről és adatbázisokról, amelyekbe való bekerülés semmiféle szakmai bírálatot nem igényel (például: Google Scholar, Research Bible, CiteFactor, Index Copernicus, International Innovative Journal Impact Factor stb.). Továbbá egyes lapok honlapjukon Scopus- vagy Web of Science-indexáltságot tüntetnek fel, miközben nem szerepelnek ezekben az adatbázisokban. Egy átlagos olvasó számára a sok név impozánsnak tűnhet, és azt a látszatot keltheti, hogy a lap rangos és elismert. Ezért minden esetben érdemes ellenőrizni az indexáltsági állításokat a hivatalos forrásoknál:

a) Scopus: ellenőrzés a Scimago Journal & Country Rank oldalon (www.scimagojr.com);

b) Web of Science: ellenőrzés a Master Journal List oldalon (mjl.clarivate.com).

A predátor folyóiratok elkerülése érdekében több forrás is rendelkezésre áll. Jeffrey Beall eredeti listája 2017-ben megszűnt, de annak frissített, közösségi változatai továbbra is elérhetők (például beallslist.net). Emellett a predatoryjournals.org oldal is gyűjti a gyanús kiadókat és folyóiratokat. Magyarországon a Magyar Tudományos Művek Tára (www.mtmt.hu) honlapján listázza a predátor folyóiratokat. A lapok száma folyamatosan változik, de jellemzően 8.000 ilyen folyóirat működik világszerte.[6]

Bár a Scopus-indexáltság és a Q-besorolás jó irányú, fontos tudni, hogy ez sem jelent abszolút garanciát a minőségre. Egyes nagy volumenű Open Access kiadóknál (különösen Frontiers- vagy MDPI-folyóiratok esetében) minőségi ingadozásokat figyeltek meg. Mivel a szerzők magánszemélyként, egyetemeiken vagy támogató szervezeteiken keresztül mindenképpen fizetnek a megjelenésért (open access APC-modell), a kiadó anyagi értelemben ösztönzött a több cikk befogadására, ami a szakmai bírálati rendszer felhígulásához vezethet. Előfordul, hogy Q1-es vagy Q2-es besorolású lapokban is megjelennek módszertani vagy egyéb hibákat tartalmazó tanulmányok, ahol a szerkesztői döntés felülírja a bírálók elutasító véleményét (vagy a bírálók nem vettek észre hibákat). Ezzel szemben a nagy múltú és presztízsű kiadók (Elsevier, Springer Nature, Wiley, Taylor & Francis, Sage) és egyetemi kiadók (Oxford, Cambridge University Press) általában jó hírnevük megőrzése érdekében szigorúbb minőség-ellenőrzést folytatnak – bár ez sem jelent abszolút garanciát.

Ugyanakkor nem minden jó minőségű folyóirat szerepel a Scopus vagy WoS adatbázisokban. Az indexálás nem automatikus, hanem a lap szerkesztőségének kérelmeznie kell azt, ami jelentős adminisztratív terhet és költséget is jelent. Nem túl gyakran előfordul, hogy egy szakmai folyóirat minőségi bíráltatási rendszert működtet, de nem törekszik indexálásra. Ugyanakkor egy komoly tudományos folyóirat esetében az indexálásért való aktív erőfeszítés indokolt és

[6] A predátor folyóiratok azonosítására szolgáló feketelisták (beallslist.net, predatoryjournals.org) használata szakmai körökben vitatott. E listák fő kritikái: (1) a predátorság megállapítására vonatkozó kritériumrendszerük nem kellően átlátható és szubjektív, (2) felmerült, hogy egyes esetekben személyes indokok is szerepet játszhatnak a listázásban, (3) hiányzik a független szakmai ellenőrzés. Ezen problémák miatt számos intézmény nem ajánlja használatukat. Alternatív, megbízhatóbb megközelítést jelent a fehérlista elvű minősítés, amelynek nemzetközi sztenderdje a norvég oktatási hivatal által fenntartott publikációs-csatorna-regiszter (https://kanalregister.hkdir.no/publiseringskanaler/). Ez pozitív listaként működik. Az itt szereplő folyóiratok elfogadott, minőség-ellenőrzött publikációs fórumok. A magyar Magyar Tudományos Művek Tára (MTMT) minősítési rendszere is jelentős mértékben erre a norvég listára épül, amely ezáltal a hazai tudományos értékelésben is referenciát jelent.

várható, mivel ez objektív, külső minőségi validációt jelent a nemzetközi tudományos közösség számára. Ezért bár egy nem indexált lapban megjelent tanulmány nem automatikusan gyenge minőségű, az indexáltság hiánya – különösen univerzális tudományterületeken – mindenképpen további ellenőrzést igényel.

Sőt, a nemzetközi indexáltság különösen fontos a szakértői kompetencia értékelésekor. Ugyanis a tudományos fokozatok és pozíciók mögött nem mindig áll nemzetközileg is mérhető teljesítmény. Előfordulhat, hogy jelentős titulusokkal rendelkező szakértők publikációs listája (www.mtmt.hu, szerzői összefoglaló táblázat) kizárólag alacsonyabb presztízsű, helyi lapokban megjelent tanulmányokból, könyvfejezetekből, magánúton vagy a munkáltató egyetemen keresztül megjelentetett monográfiákból és belföldi vagy alacsonyabb presztízsű nemzetközi konferencia-előadásokból áll. Ezek tudományos értéke töredéke egy D1, Q1 vagy Q2 besorolású folyóiratcikknek. Ideális esetben egy magasabb szervezeti pozíció mögött elvárható lenne – az adott tudományterület sajátosságaitól függően – legalább 3-4 olyan tanulmány, ahol a szakértő egyedüli vagy első szerzőként szerepel magasan rangsorolt (D1, Q1–Q2) nemzetközi folyóiratokban, független hivatkozásokat gyűjtve, nem predátor folyóiratoknál.

Természetesen e mérce nem lehet mechanikus. A nemzeti tudományok (például magyar irodalomtörténet, magyar jogtudomány) esetében a monográfia és a hazai vezető folyóiratok szerepe fajsúlyosabb lehet, és az idősebb nemzedékek pályája is más elvárások mentén épült. Tehát a nemzeti nyelvű folyóiratokban való publikálás nagyon értékes is lehet. Magyarországon az MTA által „A" kategóriásként besorolt folyóiratok magas presztízsűek, és minőségi szakmai bírálatot biztosítanak. Ám azokban a diszciplínákban, ahol a tudomány nyelve univerzális (természettudományok, pszichológia, szociológia, neveléstudomány, gyógypedagógia stb.), a nemzetközi megmérettetés hiánya ma már nehezen magyarázható. Ennek hiányában az illető szakértői státusza – bármilyen pozíciót is töltsön be – nemzetközi tudományos mércével mérve akár súlytalan is lehet.

Azonban a nemzetközi publikációs lista megléte sem jelent garanciát a tényleges szakmai kompetenciára. A formális kvalifikáció és a valós tudás között szakadék tátonghat. Magas publikációs mutatókkal (akár 20–25-ös H-indexszel) rendelkező empirikus kutatók is követhetnek el alapvető módszertani hibákat. A Peter-elv szerint az emberek gyakran olyan magasra emelkednek a hierarchiában, ahol már inkompetensek – vagyis túllépik kompetenciájuk határát. Így előfordul, hogy valaki korai szakmai eredményeinek és adminisztratív vagy politikai képességeinek kombinációja révén jut tudományos vezető pozícióba, miközben kutatói kompetenciája idővel nem felel meg a pozíció elvárásainak. Természetesen a szervezői, adminisztratív munka is értékes és elengedhetetlen a tudományos kutatás működéséhez. És bár léteznek eszközök – mint például

a CreDIT authorship statement (Contributor Roles Taxonomy) –, amelyek a tanulmányok végén átláthatóan rögzítik, hogy ki mihez járult hozzá egy kutatásban (koncepció, adatgyűjtés, elemzés, kézirat elkészítése stb.), ezek hiányában vagy következetlen alkalmazása esetén a formális szerzőség félrevezető képet ad a tényleges szakmai teljesítményről. Így nem ritkán előfordul, hogy valaki publikációinak 90%-ában másodlagos szerzőként (utolsó, középső) szerepel olyan munkákon, amelyeket valójában mások készítettek el. Vagyis kizárólag rangja, pozíciója alapján – akadémiai kifejezéssel: tiszteletből – került a szerzők közé. Sőt, a tudományos közösségben jól ismert az indokolatlan társszerzőség még súlyosabb, bár nehezen kimutatható problémája is: amikor valaki politikai nyomás, intézményi hierarchia vagy pénzügyi ellenszolgáltatás révén kerül egy publikáció szerzői közé anélkül, hogy bármilyen érdemleges munkával hozzájárult volna ahhoz. Ez jelentősen torzítja a tudományos rendszer működését, mivel az így megszerzett publikációk alapján az illető a tudományos hierarchiában is előreléphet, ami alapján tudományos pályázatokat, folyóiratcikkeket, doktori jelentkezéseket és disszertációkat bírál el, vagy fiatal kutatókat vezet.

Mindez azt mutatja, hogy sem az egyéni szakértők formális mutatói (pozíció, publikációs lista), sem egyetlen tanulmány önmagában nem nyújt megbízható alapot a tudományos forrásértékeléshez. Még ha egy tanulmány Q1 besorolású folyóiratban jelent is meg, az nem jelent végső igazságot, csupán hozzájárul *a bizonyítékok összességéhez*. Ezért egy kérdést illetően a tudományos összképet érdemes keresni:

a) Több független kutató vagy kutatócsoport jutott hasonló eredményre megfelelő minőségű folyóiratokban?
b) Sikerült replikálni a megállapításokat?
c) Léteznek metaanalízisek vagy szisztematikus irodalmi áttekintések a témában?
d) Milyen a szakmai konszenzus a kérdésről?

Az eddigi torzítások a tudományos publikálás technikai és intézményi aspektusait érintették. Azonban létezik egy ennél is alapvetőbb torzítás, amely magát a módszertani gondolkodást formálja. Ez a társadalomtudományos kutatási módszerek kvantitatív és kvalitatív kategóriákra való felosztása. Ez a felosztás különösen a bölcsészettudományi közegben vált hobbilóvá, illetve kapott ideológiai töltetet. A következő fejezet azt mutatja meg, hogy ez a dichotómia nemcsak elavult, hanem káros is.

5. A káros „kvantitatív vagy kvalitatív” dichotomizálás

A társadalomtudományi módszertani irodalomban évtizedek óta uralkodó megközelítés a kutatási módszerek kvalitatív és kvantitatív kategóriákba sorolása. Ez a felosztás olyan mélyen beépült a tudományos gondolkodásba, hogy szinte természetesnek tűnik – holott kialakulása hosszú tudományfilozófiai viták eredménye. A látszólagos természetességgel szemben a dichotomizálás számos problémát vet fel, és gyakran inkább akadályozza, mintsem segíti a megfelelő módszer kiválasztását. A hagyományos felosztás szerint a kvantitatív kutatás ugyanis:

a) számszerűsített adatokkal dolgozik,
b) nagy mintán, statisztikai módszerekkel elemez,
c) egyedi eseteken túlmutató, vagyis általánosítható eredményekre törekszik,
d) az ok-okozati összefüggések feltárását célozza,
e) benne a kutató távolságtartó, semleges megfigyelő;

amivel szemben a kvalitatív kutatás:

a) szöveges, képi vagy egyéb nem számszerű adatokat elemez,
b) kis mintán, részletes feltárással vizsgálódik,
c) szubjektív jelentéseket, egyedi eseteket vizsgál,
d) a megértést és értelmezést helyezi előtérbe,
e) benne a kutató tevékeny, értékekre hangolt résztvevő.

E tíz jellemző valóban különböző lehetőségeket mutat be az adatok fajtájára, a minta méretére, az elemzés módszereire és a kutatói szerepelvárásra, működésmódra vonatkozóan. A hozzájuk kapcsolódó kvantitatív-kvalitatív felosztás és különösen az értékalapú szembeállítás azonban számos problémát rejt magában. Talán a legfontosabb, hogy egy ilyen felosztás áldilemmát teremt, és morálisan sugalmazó, mert választást ösztönöz az ugyan „kemény”, de felszínes számok és a „valódi”, mély értelmezések között. Ez a szembeállítás akkor válik kártékonnyá, amikor fetisizálja a hermeneutikai módszerekkel feltárt „mélységet”, miközben felszínesnek bélyegzi a számszerűsítést: a romantikus vélekedés szerint egy 10

oldalas interjú elemzése érvényesebb képet ad egy emberről, mint a tőle származó 10 szám egy kérdőíven. De vizsgáljuk meg ezt kritikusan, például a szóbeli kikérdezés példáján. Az interjúhelyzet egy többszörösen torzított folyamat: 1) a válaszadó reflektál önmagára, ami önmagában is szelektív; 2) gondolatait, reflexióit szavakba önti, ami a nyelvi képességeitől és a szituációról alkotott képétől függ; 3) a kutató ezt a szöveget redukálja és értelmezi kisebb-nagyobb mértékben a saját előfeltevései, tévedései és valamilyen külső szempontrendszer alapján. Miért gondolná bárki, hogy ez a többlépcsős, szubjektív interpretációs lánc eredendően megbízhatóbb, érvényesebb, mint egy strukturáltabb adatfelvételt követő matematikai elemzés? Ráadásul ez a nézet gyakran a „választartálymodell" (mintha az emberek fejében kész, konzisztens és raktáron lévő válaszok lennének, amelyeket csak elő kell venni) naiv feltételezésén nyugszik. A valóság az, hogy az ember a válaszait kisebb-nagyobb mértékben helyben alkotja meg, figyelembe véve a környezetet, a kérdező személyét, a vélt és valós elvárásokat.

Ezzel szemben egy jól megszerkesztett, több kérdésből álló Likert-skála esetében az állítás és az állítás minősítésének fokozatai is egyértelműek. Továbbá bár itt sem lehet a válaszadó fejébe látni, a több, hasonló irányba mutató kérdés együttesen egy látens változót mér, ami csökkenti a torzítást, vagyis növeli annak a valószínűségét, hogy azt a mögöttes attitűdöt vagy vélekedést mérjük, amelyre kíváncsiak vagyunk.

A kérdés tehát nem az, hogy egy hosszú szöveg „mélyebb-e" egy számnál, hanem az, hogy melyik eszköz milyen torzítások mellett képes a leghatékonyabban megragadni a vizsgált jelenséget. A hermeneutikai elemzés mélysége és érvényessége nem magától értetődő, a számszerűsítés nem eleve felszínes. A „kvalitatív" megnevezésű vizsgálatokat nem lehet a „művészetekkel" összefüggésbe hozni, és nem szerencsés a kvalitatív szó „minőségi" áthallása sem. A tudomány történetében a dichotómiák által keretezett konfliktusokat ritkán vagy sosem oldották fel. A klasszikus természet kontra nevelés (nature vs. nurture), nevelésművészet kontra neveléstudomány vagy objektív kontra szubjektív vitákban nem lehet győztest hirdetni, mert már maga a problémafelvetés kategorizáláson, aggregáláson, primitivizáláson alapul (lásd még: hamis dilemma).

A hagyományos kvantitatív és kvalitatív felosztás is ilyen durva kategorizáláson, aggregáláson, túlzó egyszerűsítésen alapul. Strukturálatlan adatokat lehet hermeneutikai és matematikai módszerekkel is elemezni, metrizált skálán megválaszolható kérdéseket statisztikával lehet feldolgozni, de az eredmények értelmezése már hermeneutikai feladat. Például a faktorelemzés matematikai eljárással azonosít rejtett dimenziókat, de a faktorok elnevezése és értelmezése a kutató tevékenységének már tisztán értelmező mozzanata. Elképzelhető, hogy egy kutatásban 50 fővel készítenek strukturált mélyinterjút, illetve

léteznek 30 fővel lebonyolított olyan értelmes kérdőíves vizsgálatok is, melyben a kérdéseket a válaszadók Likert-skálán válaszolták meg.

Az interpretáció, a mély értelmezés igénye minden kutatásban általános, és nem csak a „kvalitatív" kutatások jellemzője. A faktoranalízisnél említett interpretatív aktus elmondható a regresszióról (melyik változót vonom be?), a kísérletről (mit mérünk egyáltalán?), a tartalomelemzésről (hogyan kategorizálunk?) és bármely egyéb módszerről is. A „kvantitatív" kutató is értelmez, csak a statisztikai formalizmus laikus szemmel feltűnőbb, mint az interpretációs réteg. Persze a valóságban a két pólus ritkán jelenik meg tisztán, ilyen a kutatás természete. A hermeneutikai elemzés sem teljesen strukturálatlan, kódolási sémák, protokollok, elméleti keretek részben strukturálják az adatot már a gyűjtés előtt. A matematikai módszer sem teljesen értékmentes algoritmus, a strukturált adat létrehozása (operacionalizálás, mérőeszköz-alkotás) már interpretatív aktus. Vagyis a két tiszta forma inkább ideáltípus (Weber értelmében), és a valódi kutatási gyakorlat mindig valahol a kettő között helyezkedik el. Egyszerűen megfogalmazva: a legtöbb kutatás a vegyes zónában él. Vagyis nem létezik kvantitatív és kvalitatív, csak tudományos és nem tudományos állítás. A tudományosság kritériuma nem a paradigmához, hanem az episztemológiai fegyelemhez kötődik, a kutató mindig egy többdimenziós összefüggés-komplexummal áll szemben. A valóság összetett, bármilyen módszer alkalmazása már önmagában reduktív aktus: kiválasztjuk, mit nézünk, mit mérünk, mit kérdezünk.

A paradigmaharc ezért részben kategóriahibán alapul. Akármit is tegyünk, akármelyik paradigmával dolgozzunk, az adatvesztést csak elméletileg lehetséges elkerülni! A cél ezért nem a helyes paradigma megtalálása, hanem az adatvesztés minimalizálása, az elvesztett adat átlátható kezelése és az új információ létrehozása. Ez a paradigmaharc végső, elvi lebontását jelenti. Összefoglalva, sem a kiinduló, nyers adatok típusa, sem az adatfelvétel módja vagy a minta mérete alapján nem lehet döntést hozni arról, hogy a kutatás melyik paradigmába tartozik.

Ennek ellenére, ha elkerülhetetlen lenne két kategóriát két végpontban meghatározni, lehetnek olyan társadalomtudományos kutatások, amelyek alapvető elemzési eljárása

a) a matematikán, vagyis a strukturált adatokon végrehajtott algoritmuson vagy
b) a hermeneutikán, vagyis a strukturálatlan adatokon végzett jelentésképzésen nyugszik.

Ez a megkülönböztetés gyakorlatias és nem ideologikus, mert a) nem zárja ki egymást a két megközelítés; b) a módszerválasztást az adatok természete, nem pedig értékalapú megfontolások határozzák meg; c) a fogalmak szintjén is elfogadja a létező átjárhatóságot, mert strukturálatlan adatokat strukturálhatunk (például tartalomelemzéssel), strukturált adatokat értelmezhetünk (például faktorok elnevezésével).

Az a – nem Winston Churchill által megalkotott – közhely, miszerint „csak abban a statisztikában hiszek, amit én magam hamisítottam", laikus élcelődés, amivel könnyen szembeállítható Frederick Mosteller állítása: „Statisztikával hazudni könnyű, de statisztika nélkül még könnyebb." De a társadalomtudományokban a tisztán hermeneutikai megközelítésű kutatás ugyanúgy szisztematikus, módszeres és tudományos igényű tevékenység lehet, mint az elemzését tekintve matematikai megközelítésű. A minőség nem az adattípustól (strukturált vagy nem strukturált) függ, hanem számos egyéb, a kutatás minőségét jellemző tényezőtől, például a zavaró tényezők kezelésétől. A továbbiakban ezt a kérdést érintjük.

6. Zavaró tényezők (Confounding factors) – a harmadik változó problémája

A társadalomtudományi kutatás egyik legnagyobb kihívása az, hogy a vizsgált jelenségek ritkán léteznek elszigetelten. Az emberi viselkedés és a társadalmi folyamatok összetett, korrelációs vagy ok-okozati hálózatokban zajlanak, ahol számos tényező hat egyszerre. Míg a hétköznapi gondolkodás hajlamos két együtt járó esemény között azonnal oksági kapcsolatot feltételezni, a tudományos módszertan óvatosságra int. A „korreláció nem jelent kauzalitást" közhely igazságtartalma mögött gyakran egy fel nem ismert, harmadik (századik) változó rejtőzik, amely alapvetően megváltoztatja a viszonyokról alkotott képet. Ezért a kutatások érvényességének biztosításához elengedhetetlen, hogy különbséget tudjunk tenni e harmadik, zavaró tényező különböző típusai között. Mielőtt vizsgálnánk a zavaró tényezők kezelésének módszereit, meg kell jegyezni, hogy ez a probléma elsősorban a strukturált adatokon alapuló, prediktív kapcsolatokat vizsgáló kutatásokban merül fel az itt bemutatott formában.

Amikor a cél egy hatás (például egy oktatási beavatkozás hatása a teljesítményre) méretének pontos becslése és általánosítása, ezen tényezők statisztikai kontrollja elkerülhetetlen. Hiszen gyakran nemcsak az lehet érdekes, hogy egy jelenség (A) okoz-e egy másikat (B),

1. ábra: Az oksági kapcsolat alapmodellje

hanem az is, hogy hogyan teszi azt. A mediátorváltozó erre a „hogyan"-ra ad választ, mivel feltárja a hatás mögöttes mechanizmusát. A mediátor a magyarázó (független változó) és a magyarázott (függő változó) közötti logikai láncban helyezkedik el, ahol a független változó a mediátoron keresztül fejti ki hatását.

2. ábra: A mediátor szerepe az oksági láncban

Egy közismert társadalmi összefüggés szerint a magasabb képzettség jellemzően magasabb jövedelemmel jár együtt. De miért? A képzettség önmagában nem eredményez jövedelmet. A mediátor ebben a kapcsolatban a megszerzett munkahelyi pozíció lehet. A magasabb képzettség jobb pozíció elérését teszi lehetővé, és a jobb pozíció magasabb fizetéssel jár.

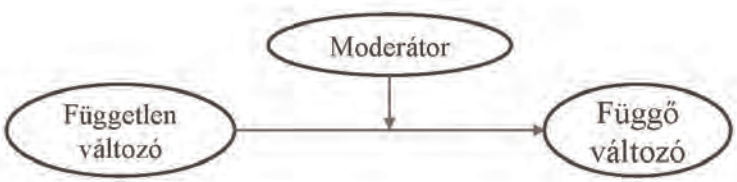
A mediáció tesztelésére eredetileg 1986-ban kidolgozott módszer logikája szerint a mediátor akkor tölti be szerepét, ha annak statisztikai kontrollálásával az eredeti (a példában a képzettség-jövedelem) kapcsolat szignifikánsan gyengül vagy teljesen el is tűnik. A módszer konkrét lépései a következők:

a) Kimutatni, hogy a független változó (képzettség) szignifikánsan hat a függő változóra (jövedelem). Ha nincs kimutatható kapcsolat, nincs mit mediálni.
b) Kimutatni, hogy a független változó (képzettség) szignifikánsan hat a feltételezett mediátorra (munkahelyi pozíció).
c) Kimutatni, hogy a mediátor (pozíció) szignifikánsan hat a függő változóra (jövedelem), amikor a független változót is figyelembe vesszük.

e) Amikor a mediátort statisztikailag kontrolláljuk, az eredeti kapcsolat (képzettség-jövedelem) szignifikánsan gyengül (részleges mediáció) vagy teljesen eltűnik (teljes mediáció).

A mediáció jelenségét ma már összetettebben értelmezik. Az eredeti, lépésenkénti módszer ugyan megalapozta a mediáció vizsgálatát, de ma már meghaladottnak tekinthető, mivel nem képes azonosítani minden mediációs folyamatot. A korszerű álláspont szerint mediáció akkor is létezhet, ha nincs kimutatható teljes hatás a független és a függő változó között, mivel előfordulhat, hogy az ellentétes irányú közvetlen (X → Y) és közvetett (X → M → Y) hatások statisztikailag kioltják egymást. Éppen ezért a korszerű elemzések már közvetlenül a közvetett hatás szignifikanciáját tesztelik, jellemzően bootstrap eljárással.

Míg a mediátor a hatás útvonalát írja le, a moderátor a független és a függő változó közötti kapcsolat erősségét vagy irányát befolyásolja. Arra a kérdésre, hogy „Van hatás?", a moderációs vizsgálat azt válaszolja: „Attól függ".



3. ábra: A moderátor hatása az oksági kapcsolatra

Ha a kutatás a társas támogatásnak a stressz egészségre gyakorolt negatív hatásait enyhítő szerepét vizsgálja, akkor a moderátor a társas támogatás lehet. Ebben az esetben az várható (statisztikai modellel tesztelhető), hogy a stressz és a megbetegedések közötti kapcsolat csak azoknál erős, akik alacsony társas támogatással rendelkeznek, míg a támogató szociális hálóval rendelkező egyéneknél a stressznek alig van kimutatható egészségkárosító hatása (ez a moderációs vagy interakciós hatás, ami jellemzően regressziós vagy ANOVA-modellekben tesztelhető, a független változó és a moderátor szorzatának vizsgálatával).

A harmadik változók leginkább problematikus típusa a zavaró tényező (confounder, amiről ez az alfejezet szól), amely nem egy valódi kapcsolat értelmezését árnyalja, hanem egy teljesen hamis, látszólagos összefüggést hoz létre, vagy egy létező összefüggést fed el. Ez természetesen nagyon jelentős probléma a tudományos tudás érvényességének szempontjából. A zavaró tényező egyszerre van kapcsolatban a magyarázó és a megmagyarázott változóval is,

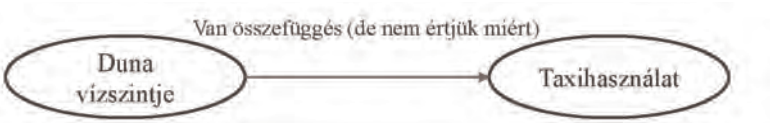
megváltoztatva ezáltal a kettő közötti látszólagos összefüggést, például azt az illúziót keltve, hogy összefüggés van közöttük.

4. ábra: A zavaró tényező (confounder) által keltett látszólagos kapcsolat

Egy gondolatkísérlet szerint számos városban megfigyelhető, hogy azokban a hónapokban, amikor több fagylaltot adnak el, a bűncselekmények, különösen az erőszakos bűncselekmények száma megnő. Ebből levonhatnánk azt a téves következtetést, hogy a fagylaltfogyasztás, mondjuk a fagylaltban lévő valamilyen anyag agressziót vált ki, ami bűnözéshez vezet. A valóságban azonban a meleg időjárás a zavaró tényező:

a) Meleg időben az emberek több fagylaltot esznek (kapcsolat a feltételezett okkal).
b) Meleg időben az emberek több időt töltenek a szabadban, gyakrabban kerülnek társas kölcsönhatásokba, ami növeli a konfliktusok és bűncselekmények esélyét (kapcsolat a feltételezett okozattal).

A fagylaltfogyasztás és a bűnözés közötti kapcsolat tehát csupán látszólagos, és valójában mindkettő mögött a meleg időjárás áll mint közös ok. Ha a kutatás során nem vennénk figyelembe az évszak hatását, teljesen téves következtetésre jutnánk. Egy talán még szemléletesebb példa a Duna vízszintje és a budapesti taxihasználat közötti korreláció. Mire gondolhatunk, ha bizonyíték van arra, hogy a Duna vízszintjének 10 centiméteres emelkedése a budapesti taxihasználat 45%-os növekedését eredményezi?



5. ábra: Látszólagos korreláció a Duna vízszintje és a taxihasználat között

A háttérben itt is egy közös ok, az eső áll, amely egyszerre emeli meg a folyó vízszintjét és növeli az emberek hajlandóságát a taxizásra.

6. ábra: Az eső mint közös ok (zavaró tényező) a Duna vízszintje és a taxihasználat kapcsolatában

A zavaró tényező mindkét eddigi példánkban illuzórikus kapcsolatot hozott létre két másik között, ahogy az eső összekötötte a Duna vízszintjét és a taxihasználatot. A harmadik változó hatása azonban nem csupán ilyen látszólagos kapcsolatokat hozhat létre, hanem épp ellenkezőleg, egy valós összefüggést el is fedhet előlünk. Például a munkatapasztalat és egy tűzoltó stresszhelyzetben nyújtott teljesítménye között meglepő módon nem találunk összefüggést. Ebben az esetben az életkor úgynevezett szuppresszor (elnyomó) változóként[7] van jelen. A szuppresszor egy olyan sajátos tényező, amely – ellentétben a korábban bemutatott confounderrel – nem látszólagos kapcsolatot hoz létre, hanem éppen ellenkezőleg, elrejti vagy gyengíti a változók közötti valós összefüggést. A tűzoltók példájánál az életkor két ellentétes mechanizmust hoz működésbe. Miközben az évek száma növeli a tapasztalatot, ami javítja a teljesítményt, ugyanakkor rontja a fizikai állóképességet, ami rontja a teljesítményt. Ez a két ellentétes irányú hatás statisztikailag kioltja egymást, így a tapasztalat valódi, pozitív ereje rejtve marad. Ha azonban az elemzés során kontrolláljuk (kiszűrjük) az életkor hatását, a tapasztalat és a teljesítmény közötti valós, pozitív kapcsolat azonnal láthatóvá válik.

Bár a fagylaltfogyasztás és a Duna vízszintjének példái tréfásnak vagy túlzóan leegyszerűsítettnek tűnhetnek, komolyan kell venni azt a logikai csapdát, amelyet szemléltetnek. Ugyanis a zavaró tényezők nemcsak a hétköznapi gondolkodást tévesztik meg, hanem gyakran az empirikus tudományok eredményeinek érvényességét is aláássák vagy akár teljesen meg is szüntetik. A tudománytörténet tele van olyan esetekkel, amikor egy-egy fel nem ismert zavaró tényező teljesen téves következtetésekhez és szakmai vitákhoz vezetett, nemcsak a laikusok, hanem a kutatók körében is. Például a korai megfigyeléses vizsgálatok erős kapcsolatot találtak a kávéfogyasztás és a szívbetegségek megnövekedett kockázata között. Később derült ki, hogy a dohányzás volt a

[7] A szuppresszió jelenségét akkor azonosítjuk, ha egy kontrollváltozó modellbe emelése nem csökkenti, hanem növeli a független változó magyarázó erejét, vagy megváltoztatja annak irányát. A szuppresszorváltozó tehát a kimeneti változóhoz (Y) nem kapcsolódó, de a bemeneti változóval (X) korreláló varianciát (a zajt) szűri ki, ezáltal tisztítva meg az X és Y közötti kapcsolatot.

zavaró tényező. Azok az emberek, akik a hetvenes-nyolcvanas években sok kávét ittak, gyakrabban is dohányoztak, és valójában a dohányzás, nem pedig a kávé okozta a megnövekedett kockázatot. Vagy például évtizedekig úgy tűnt, hogy a hormonpótló terápia csökkenti a menopauzán átesett nőknél a szív- és érrendszeri megbetegedések esélyét. A zavaró tényező itt az egészségtudatosság és a társadalmi-gazdasági státusz volt. Azok a nők, akik a terápiát választották, eleve egészségesebben éltek, többet mozogtak, és jobb anyagi körülmények között éltek. A látszólagos pozitív hatást nagyrészt ez a háttér, és nem maga a gyógyszer okozta, sőt a későbbi, randomizált kísérletek ennek épp az ellenkezőjét mutatták ki. Gyakori megfigyelés az is, hogy a magániskolákba járó diákok a többségi intézményekbe járó diákokhoz hasonló vagy akár jobb tanulmányi eredményeket érnek el. Mielőtt ezt az iskola hatékonyságának tulajdonítanánk, figyelembe kell vennünk a családi hátteret (a szülők iskolai végzettsége, jövedelme, motivációja) mint erőteljes zavaró tényezőt. A magániskolát választó és megfizetni képes családok gyakran más módokon is támogatják gyermeküket, ami önmagában is hozzájárul a jobb eredményekhez.

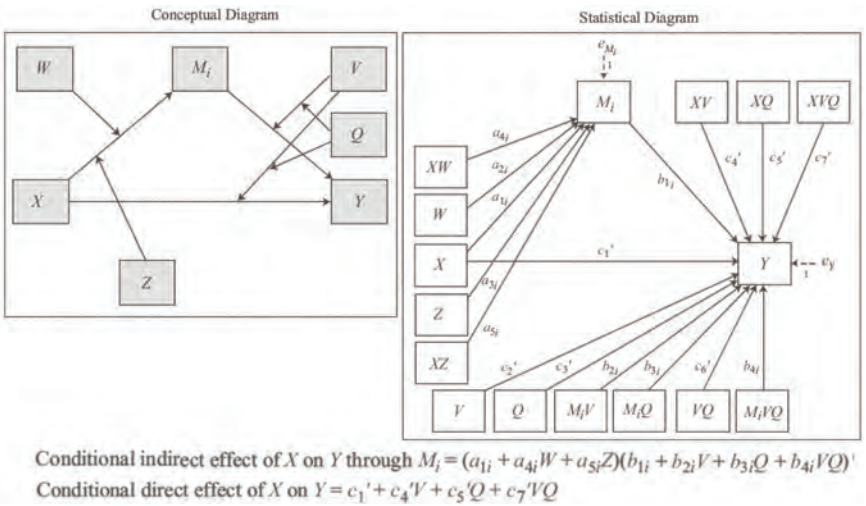
Ezek a példák rávilágítanak arra, hogy a harmadik változó problémájának figyelmen kívül hagyása nemcsak elméleti hiba, hanem a valóság félreértelmezéséhez, téves szakpolitikai döntésekhez vagy akár káros egészségügyi gyakorlatokhoz is vezethet. Tekintsük át táblázatos formában a három változótípus főbb jellemzőit:

Változó típusa	Szerepe	Helye a kauzális láncban	Példa
Mediátor	A hatás mechanizmusa	A → M → B	Képzettség → Pozíció → Jövedelem
Moderátor	A hatás feltétele	Befolyásolja A → B erősségét	Társas támogatás módosítja a stressz-betegség kapcsolatot
Confounder	Hamis kapcsolat forrása	C → A és C → B	Eső → Duna-vízszint és eső → taxihasználat

1. táblázat: A harmadik változók típusai és szerepük a kutatásban

De sajnos a problémákat ezzel az áttekintéssel még nem oldottuk meg. A világ nagyon bonyolult. A valóságban nem egyetlen, hanem számos tényező befolyásolja a kimenetet, és a tényezők nemcsak a kimenettel, hanem egymással is bonyolult kapcsolatban állnak. Az eddig bemutatott egyszerűsített, háromváltozós modellek csupán a probléma alapvető logikájának megértését szolgálták. A valóságban, a társadalomtudományos kutatásban olyan jóval összetettebb, de

a valós bonyolultságot még korántsem tökéletesen leíró összefüggésrendszerekkel kell számolnunk, mint például:



7. ábra: Komplex mediációs és moderációs modell (Hayes PROCESS 54)

E bonyolultság felvetheti a kérdést, hogy van-e egyáltalán értelme egyszerűsített modelleket alkotni a valóságról. Erre a kérdésre egy lehetséges válasz: *„All models are wrong, but some are useful."* (George E. P. Box) Vagyis a társadalomtudományi modellek törekvése a valóság magyarázatára nem reménytelen. Ellenkezőleg, a tudományos modellezés során keressük meg a legfontosabb, legerősebb összefüggéseket, vagy azonosítjuk ezeket először, majd vizsgáljuk meg a finomabb, gyengébb, de még lényeges hatásokat. Vannak olyan elméleti modellek, amelyek néhány változó segítségével meglepően jól képesek magyarázni bonyolult jelenségeket. Például a technológia elfogadásának modellje (Technology Acceptance Model – TAM) egy rendkívül sikeres modell, amely alapvetően azt mondja ki, hogy az emberek akkor fognak használni egy új technológiát, ha azt hasznosnak és könnyen kezelhetőnek tartják. Ezt az egyszerű modellt számtalan kutatás igazolta a legkülönbözőbb kulturális környezetekben.

Természetesen a modellek magyarázó ereje témánként és tudományterületenként erősen eltérő. Egy használt autó eladási árát néhány prediktorral (márka, évjárat, futott kilométer, állapot, motor típusa) akár 90%-os pontossággal is be lehet jósolni. A társas, humán tér természete ennél jóval összetettebb, és az emberi viselkedést, attitűdöket vagy társadalmi kimeneteket 90%-ban magyarázó tényezők száma nem öt (mint az előbbi példában), hanem több száz vagy talán végtelen, és a tényezők egymással is összefüggenek, továbbá hatásuk

gyakran kontextusfüggő. Éppen ezért a társadalomtudományokban a jó modell mércéje más, mint a gazdasági vagy műszaki területeken. Ez utóbbi esetekben elvárás lehet a magas magyarázó- és bejóslóerő, azonban a szociológiában, a pszichológiában vagy a neveléstudományban már egy jóval kisebb megmagyarázott variancia (a valóság változatossága, aminek egy részét a modell magyarázza) is tudományos eredménynek számít. Egy társadalomtudományi modell, amely a vizsgált jelenség varianciájának 20%-át képes megmagyarázni, már értékesnek tekinthető. Különösen erős, jól kidolgozott modellekkel esetenként a változatosság akár 50-60%-a is magyarázható, ami többször megerősített esetben (több független kutatás ugyanarra vagy nagyon hasonló eredményre jut) már önmagában szakpolitikai beavatkozást indokolhat.

Bár a tudományos módszertan továbbra is a legmegbízhatóbb eszköz a világ megismerésére, ez nem jelenti azt, hogy a tudomány tévedhetetlen, vagy hogy egyenes úton halad az igazság felé. Sőt, a tudomány története tele van tévedésekkel, zsákutcákkal és később megcáfolt elméletekkel. A tudomány ereje nem a tévedhetetlenségében, hanem a beépített önkorrekciós mechanizmusaiban rejlik. A tudományos módszertan tehát a leghasznosabb és az eddig legsikeresebb eszköz arra, hogy a bonyolult valóságot megbízhatóan és érvényesen magyarázzuk. A cél sokszor az, hogy a kutatás megtalálja azokat a modelleket, amelyek – bár szükségszerűen egyszerűsítenek – a puszta vélekedésnél megalapozottabb következtetésekre és a jelenségek mélyebb megértésére adnak lehetőséget. Ezért a társadalomtudomány eredményei nem feltétlenül bizonytalanok. A szigorú módszertannal feltárt, bizonyítékokon alapuló felismerések alapjaiban formálják a szakpolitikát, és javítanak emberek millióinak életén. A pszichológia és a közgazdaságtan határterületén végzett, Nobel-díjjal elismert kutatások például feltárták, hogy az emberek döntéseit apró beavatkozásokkal, ún. „ösztönzésekkel" jelentősen javítani lehet. Ennek egyik leglátványosabb eredménye a nyugdíj-megtakarítási rendszerek átalakítása, és azzal, hogy a munkavállalókat automatikusan beíratják a programokba, jelentősen megnő a megtakarítók aránya, ami milliók időskori anyagi biztonságát alapozza meg. Hasonlóképpen, a kriminológia területén az adatalapú kutatások bizonyították, hogy a bűnözés nem véletlenszerűen oszlik el egy városon belül, hanem kis, „forró pontokra" összpontosul. Erre a tudományos eredményre építve a rendőrségek világszerte a járőrözést ezekre a területekre összpontosítják, ami – amint azt számos szisztematikus vizsgálat igazolta – hatékonyan csökkenti a bűncselekmények számát. Vagy jól ismertek azok a több évtizedes, randomizált kísérleteken alapuló fejlődéspszichológiai kutatások, amelyek a hátrányos helyzetű gyerekeknek nyújtott minőségi kora gyermekkori nevelés jelentős társadalmi megtérülését igazolták: a programokban részt vevő gyerekek felnőttként jobb

egészségnek örvendtek, magasabb jövedelmet értek el, és kisebb eséllyel kerültek összetűzésbe a törvénnyel. Az ilyen típusú tudományos eredmények nyilvánvalóan megalapozottak, és komolyan befolyásolják a szakpolitikát, a törvényeket és általában az emberi élet minőségét.

7. A zavaró tényezők kezelése a kutatási folyamatban

A zavaró tényezők azonosítása és hatásuk semlegesítése a kutatás érvényességének és a minőség biztosításának tehát alapfeltétele. Ahogyan láttuk, enélkül hamis ok-okozati következtetésekre juthatunk, ami téves gyakorlati beavatkozásokhoz vezethet, például a fagylaltárusítás betiltásához a bűncselekmények visszaszorítása érdekében. A tudományos hitelesség megköveteli a zavaró tényezők szisztematikus feltárását és kezelését. Erre a kutatás több szakaszában is lehetőség van.

A kutatástervezés szakaszában a leghatékonyabb módszer a randomizálás, például a résztvevők véletlenszerű besorolása kísérleti és kontrollcsoportba. Ez az eljárás – elméletileg, de ezt ellenőrizni szokás az elemzés megkezdése előtt is – minden ismert és ismeretlen zavaró tényező hatását egyenletesen osztja el a csoportok között. Amikor azonban a randomizálás nem etikus vagy nem kivitelezhető, például nem sorolhatunk embereket véletlenszerűen szegény és gazdag csoportokba, olyan eszközökhöz kell nyúlni, mint például a párosítás (matching), ahol minden kezelt személyhez hasonló jellemzőkkel rendelkező kontrollszemélyt keresünk.

Az adatgyűjtés szakaszában is tehetünk a zavaró tényezők hatásának csökkentése érdekében. Például előre azonosítsunk és mérjünk minden olyan lehetséges zavaró tényezőt (kor, nem, iskolai végzettség, jövedelem stb.), amely a tárgyhoz tartozó lehet. Ezeket az információkat később az elemzés során felhasználhatjuk a hatásuk kontrollálására.

Az elemzés szakaszában számos statisztikai eljárás áll rendelkezésre. A legegyszerűbb a rétegzett elemzés, ahol a mintát a zavaró változó értékei szerint alcsoportokra (rétegekre) bontjuk, és az elemzést ezekben a csoportokban külön-külön végezzük el. Így például a képzettség és a jövedelem kapcsolatát külön vizsgálhatjuk férfiaknál és nőknél. Ennél bonyolultabb, de hatékonyabb

módszer a többszörös regresszióanalízis, amely lehetővé teszi több zavaró tényező hatásának egyidejű statisztikai kontrollálását. Például az osztályméret hatását vizsgáljuk a tanulmányi eredményre. A kisebb osztálylétszám jobb tanulmányi eredményekhez vezet? Itt a lehetséges zavaró tényezők a következők.

a) Iskolamodell. A magániskolákban jellemzően kisebb osztályok vannak és jobb eredmények, de ez nem feltétlenül az osztályméret miatt van.
b) Szülői háttér. A tehetősebb szülők gyakran olyan iskolát választanak, ahol kisebb osztályok vannak.
c) Településtípus. Kistelepüléseken demográfiai okokból kisebbek az osztályok, de kevesebb erőforrás is jut oktatásra.
d) Tanári kvalitás. A jobb iskolák nemcsak kisebb osztályokat, hanem képzettebb vagy jóval elkötelezettebb tanárokat is biztosíthatnak.

Fenti példában a zavaró tényezők kezelése a következő lehet:

a) Tervezés. Csak állami iskolákat veszünk be a mintába (így kizárjuk az iskolamodell hatását).
b) Adatgyűjtés. Részletesen mérjük a szülői végzettséget, jövedelmet, a település típusát, a tanárok képzettségét.
c) Elemzés. Többszintű modellezést alkalmazunk, amely figyelembe veszi mind az iskola-, mind az osztályszintű hatásokat, és statisztikailag kontrollálja a mért zavaró tényezőket.

A fejezetben tárgyalt zavaró tényezők mellett más okai is lehetnek az elemzések hamis eredményeinek. Tökéletesen szemlélteti ezt az előző példa, melyben a confounder változók mellett az egy osztályba tartozó diákok között kialakuló hasonlóság torzító erejével is meg kell küzdenünk, azaz kezelnünk kell, hogy a vizsgálatban a méréseink nem függetlenek. A problémára a megoldást a XI. és XII. fejezetben tárgyalt beágyazott ANOVA- és lineáris kevert modellek nyújtják majd.

Természetesen a korszerű kutatásmódszertan olyan fejlett technikákat is kínál, amelyek még összetettebb helyzetekben is képesek a zavaró hatások semlegesítésére. Ezek akkor is használhatók, ha a randomizált kísérlet (RCT) nem alkalmazható, de szeretnénk egy magas bizonyítóerőt elérni (kvázikísérlet):

a) Instrumentális változók módszere. Olyan változót keresünk, amely hat a független változóra, de a függő változóra csak azon keresztül. Például a katonai szolgálat sorsolása mint instrumentális változó használható a jövedelemre gyakorolt hatás vizsgálatában.

b) Difference-in-differences (DiD). Időbeli változásokat hasonlítunk össze kezelt és kontrollcsoportok között. Például egy oktatási reform hatását úgy mérjük, hogy összehasonlítjuk a reformot bevezető és nem bevezető iskolák eredményeinek változását az idő múlásával.
c) Propensity score matching. Összetett statisztikai eljárás, amely egy pontszámba sűríti össze sok zavaró változó hatását. Minden kezelt személyhez vagy egységhez hasonló „hajlandósági pontszámú" kontrollszemélyt vagy egységet keresünk.
d) Regressziós diszkontinuitás (RDD). Amikor a kezelés egy küszöbérték alapján történik (például ösztöndíjat kap, aki 4,5 feletti átlagot ér el), a küszöb közelében lévők összehasonlítása jó közelítést ad a tényleges hatásról.

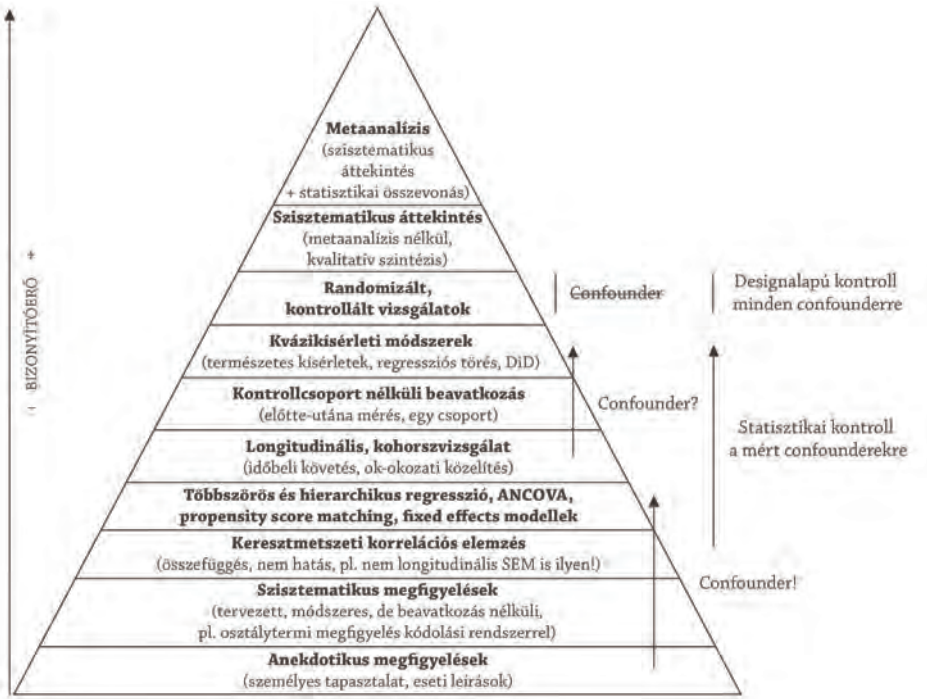
3. doboz: Az instrumentális változók alkalmazása

Az instrumentális változó olyan külső tényező, amely csak a független változón keresztül hat a függő változóra, közvetlenül nem. Például Joshua Angrist klasszikus kutatásában a kutatási kérdés az volt, hogy a katonai szolgálat egy személy jövőbeli jövedelmét növeli vagy csökkenti. Itt a független változó az volt, hogy az egyén volt-e katona, a függő változó pedig a jövedelem. A probléma (zavaró tényező) az, hogy nem hasonlíthatjuk össze egyszerűen a veteránok és a nem veteránok jövedelmét, mert az emberek nem véletlenszerűen lesznek katonák. Ezt a rejtett önkiválasztást (self-selection bias) mint zavaró hatást ki kell küszöbölnünk. Ezért a vietnámi háború idején az USA-ban sorsolással (lottóval) döntötték el, hogy a hadköteles korú férfiak közül kinek kell bevonulnia. Ez a sorsolás véletlenszerű volt. Logikája a következő.

1. Az instrumentum hat a független változóra. A kisorsolt alacsony sorszám drasztikusan megnövelte az esélyét annak, hogy valaki katona legyen.
2. Az instrumentum közvetlenül nem hat a függő változóra. A sorsoláson kapott szám önmagában nincs hatással a jövedelemre. Az egyetlen út, ahogy a sorszám a jövedelmet befolyásolhatja, az a katonai szolgálaton keresztül vezet.

A módszer azoknak a jövedelmét hasonlítja össze, akiket a sorsolás bevonulásra kényszerített, azokéval, akiket megmentett tőle. Mivel a két csoport között a véletlenszerűség miatt egyetlen lényegi különbség van (a katonai szolgálat valószínűsége), a jövedelmük közötti eltérés már megbízhatóan a katonai szolgálat hatásának tulajdonítható. Röviden az instrumentális változó természetes kísérletet talál a világban.

Az eddig tárgyalt eljárások – a statisztikai kontrolltól kezdve a kvázikísérleteken át a randomizált kontrollált vizsgálatokig (RCT) – mind azt a célt szolgálják, hogy a zavaró tényezők (confounderek) kizárásával növeljék az ok-okozati (kauzális) következtetések érvényességét. Az 1. kép ezt a bizonyítóerő-hierarchiát foglalja össze.



1. kép A bizonyítóerő hierarchiája az ok-okozati (kauzális) összefüggések vizsgálatában

Megjegyzés: A fixed effect modellek paneladatokon valójában közelebb állnak a kvázikísérleti módszerekhez, mint az egyszerű regresszióhoz, mert az összes időben állandó confoundert automatikusan kizárják, nem csak a mérteket.

A zavaró tényezők problémája rávilágít arra, hogy a társadalomtudományokban miért nem elég a puszta megfigyelés. A gondos tervezés, adatgyűjtés és a megfelelő elemzési technikák alkalmazása teszi lehetővé, hogy a valódi hatásokat el tudjuk különíteni a látszólagos összefüggésektől, és ezáltal megbízható tudást hozzunk létre.

Végül meg kell említeni, hogy a hermeneutikai elemzésen alapuló kutatások a fentiektől eltérő minőségbiztosítási logikát követnek. Amikor egy jelenség hermeneutikai elemzése a cél, a kutató gyakran nem a változók közötti hatások statisztikai kiszűrésére, hanem a környezet gazdag leírására, a jelentések feltárására és az összefüggések sokrétű, értelmező bemutatására törekszik.

A zavaró tényezők problémája itt is létezik, de nem statisztikai kontrollal, hanem a kutató reflexivitásával, a körülmények alapos ismeretével és az alternatív magyarázatok mérlegelésével kezelik. Lássunk két példát a hermeneutikai kutatásokban rejlő ilyen típusú csapdákra.

Például egy kutató azt vizsgálja, hogyan fogadják a tanárok egy iskola új digitális oktatási platformját. Az interjúk során azt tapasztalja, hogy a bevont pedagógusok kivétel nélkül rendkívül lelkesek, dicsérik a platform funkcióit, és elismerően nyilatkoznak a használatáról. A kutató arra a következtetésre jut, hogy a digitális platform önmagában rendkívül ösztönző és hatékony eszköz, amely a tanárok körében széles körű tetszést arat. Az értelmezés a platformnak tulajdonítja a sikert. A probléma abból fakad, hogy a kutató nem tudja, vagy nem veszi figyelembe, hogy az iskola vezetése a próbaprogramba szándékosan csak a leginkább úttörő, a technikai eszközöket eleve kedvelő és eleve motivált tanárokat vonta be. Ezek a tanárok valószínűleg bármilyen új kezdeményezést lelkesen fogadtak volna. A valódi ok tehát nem a platform, hanem a minta mesterséges, torzított összetétele (a résztvevők rejtett, már meglévő motivációja). Így a kutatás félrevezető eredményt hoz. Ha az iskola e kutatás alapján döntene a platform teljes körű bevezetéséről, valószínűleg csalódna, mert a többi, kevésbé motivált tanár ellenállásába ütközne. A probléma nem a platformmal, hanem a kutatás téves általánosításával van.

A második példában egy szervezetfejlesztő kutató interjúkat készít egy gyár dolgozóival egy új, a munkavállalói jóllétet és bevonódást célzó programról. Azt tapasztalja, hogy a munkások kivétel nélkül cinikusan, bizalmatlanul és elutasítóan viszonyulnak a kezdeményezéshez. A kutató arra következtet, hogy a programot rosszul tervezték meg, az nem képes megszólítani a munkavállalókat, vagy hogy a fizikai munkások eleve bizalmatlanok a vezetőségi kezdeményezésekkel szemben. Ebben az esetben a probléma például abból is fakadhat, hogy a kutató nem tárja fel a szervezet múltját. Ha feltárná, kiderülne, hogy az elmúlt tíz évben a cég már három hasonló programot indított, amelyeket aztán fél év után mindenféle magyarázat nélkül megszüntetett, sőt az egyik után még leépítések is következtek. A munkások tehát nem erre a konkrét programra reagálnak, hanem a vezetőség korábbi, megszegett ígéreteinek mintázatára. A valódi ok a mélyen gyökerező, múltbéli tapasztalatokon alapuló szervezeti bizalmatlanság. A kutatás tehát teljesen rosszul diagnosztizálja a problémát. A megoldás nem egy jobban megtervezett program lenne, hanem egy hosszú, éveken át tartó bizalomépítési folyamat. A kutatási eredmény alapján hozott döntések nemcsak hatástalanok, de tovább mélyíthetik a cinizmust.

Ezek a példák rávilágítanak arra, hogy a hermeneutikai jellegű kutatásban a zavaró tényező figyelmen kívül hagyása a jelenség mélyebb értelmének és valódi

okainak teljes félreértelmezéséhez vezethet. Ezért ebben az esetben a módszertani logikát saját érvényességi követelményrendszer alapozza meg, amely az objektivista módszertan fogalmai helyett azok funkcionális megfelelőit használja. Ekkor a kutatás tudományos szigorát nem a változók elkülönítésével és az elkülönített változók megfelelő kezelésével, hanem olyan eljárásokkal igyekeznek biztosítani, amelyek feltárják és értelmezik a jelenséget befolyásoló bonyolult kontextuális tényezőket. A legfontosabb érvényességet biztosító stratégiák a következők:

a) Hitelesség (credibility). A belső érvényesség megfelelője. Azt biztosítja, hogy a kutatási eredmények a résztvevők valóságát tükrözzék. Ennek eszköze lehet a trianguláció, amely során több különböző adatforrást, módszert, kutatót vagy elméletet vetnek össze az eredmények megerősítésére. Szintén ide tartozik a résztvevői validálás (member checking), amelynek során a kutató visszajelzést kér a résztvevőktől az értelmezések pontosságáról.
b) Átvihetőség (transferability). A külső érvényesség (általánosíthatóság) helyett alkalmazott követelmény. A cél nem a statisztikai általánosítás, hanem a jelenség olyan mélységű bemutatása, hogy az olvasó maga dönthesse el, az eredmények lényegesek-e más feltételek között. Ennek eszköze a sűrű leírás, amely a viselkedés mellett annak környezetét is részletesen bemutatja, feltárva a mögöttes jelentésrétegeket.
c) Konzisztencia (dependability). A kvantitatív kutatásokban használt megbízhatóság megfelelője, de a logikája alapvetően más. Míg a kvantitatív reliabilitás az eredmények megismételhetőségét jelenti, a hermeneutikai kutatásban ez nem elvárás, hiszen a kutató maga a kutatási eszköz, és az eredmények függenek a környezettől, körülményektől. A konzisztencia ezért nem az eredmény megismételhetőségére, hanem a kutatási folyamat következetességére, észszerűségére és átláthatóságára összpontosít. A cél annak bemutatása, hogy a kutató következtetései és értelmezései nem önkényesek vagy véletlenszerűek, hanem módszeresen és észszerűen levezethetők az adatokból. Ennek legfontosabb eszköze a részletes audit trail (döntési napló) létrehozása és dokumentálása. A kutatónak részletesen fel kell jegyeznie a kutatás lépéseit, hogy a folyamat követhető és ellenőrizhető legyen. Ez a gyakorlatban magában foglalja a kutatási kérdés fejlődését, a mintaválasztás indoklását, a nyers adatokat (például interjúátiratokat), a kutatói feljegyzéseket (memókat, önreflexiókat), a kódolás és a témafejlesztés lépéseit és az elemzés során hozott döntések indoklását (például miért lett egy téma főtéma, vagy miért vontak össze két kódot). Egy ilyen „kutatási nyomvonal” teszi

lehetővé egy külső szemlélő (például bíráló, témavezető) számára, hogy megítélje, a kutatói folyamat észszerű és koherens volt-e, és az eredmények valóban az adatokban gyökereznek – még akkor is, ha ő maga esetleg egyes pontokon máshogy értelmezte volna azokat.

d) Igazolhatóság (confirmability). Az objektivitás helyett használatos fogalom. Azt hangsúlyozza, hogy az eredményeknek egyértelműen az adatokból kell származniuk, nem pedig a kutató elfogultságából. Ennek legfőbb eszköze a kutatói önreflexió: a kutató folyamatosan reflektál saját előfeltevéseire, szerepére és a kutatási folyamatra gyakorolt hatására, ezt pedig átlátható módon teszi.

A fenti keretrendszer érvényesítésének egyik kiváló gyakorlati eszköze, ha a hermeneutikai ihletésű kutatás aktívan kezeli az alternatív magyarázatokat a deviánseset-analízis módszerével is. Ahelyett, hogy a mintázatba nem illő eseteket (adatokat) statisztikai zajként kezelné, a kutató célzottan felkeresi és elemzi ezeket, hogy megértse, miért térnek el, és ezáltal finomítsa vagy elmélyítse a kialakulóban lévő elméletet.

8. Miért volt fontos a zavaró tényezőkkel ennyit foglalkozni?

A helyesen megválasztott és végrehajtott módszertan választja el a tudományos tudást a hiedelmek alapján megfogalmazott tudástól. Validitás hiányában a legkifinomultabb eljárások is félrevezető eredményekhez vezethetnek. A tudományos megismerés a hétköznapi tapasztalatszerzéstől a szisztematikusság, a nyilvánosság és a módszertani szigor (objektivitás, érvényesség, megbízhatóság) révén különül el. A zavaró tényezők kezelésének eltérő logikáit azért tárgyaltuk, mert a módszertani ismeretek önmagukban nem garantálják az érvényes következtetéseket. A tapasztalatok szerint a kutatónak a módszertani felkészültség mellett éles valóságérzékeléssel és kritikai gondolkodással is rendelkeznie kell. A megfelelő módszer kiválasztása és alkalmazása tehát elengedhetetlen, de folyamatos kutatói önreflexió és sokszor egyszerűen józan átgondolás nélkül önmagában nem elegendő a helyes következtetések levonásához.

9. Összefoglalás

Ez a fejezet a társadalomtudományi megismerés alapvető kérdéseit járta körül, a hétköznapi intuícióktól eljutva a tudományos módszertan kihívásaiig. Először szembeállítottuk a hétköznapi és a tudományos megismerést. Láttuk, hogy míg a mindennapi gondolkodásunkat praktikus heurisztikák és kognitív torzítások (mint a megerősítési torzítás vagy a túláltalánosítás) jellemzik, addig a tudományos módszertan ezek kiküszöbölésére törekszik a szisztematikusság, objektivitásra, érvényességre és megbízhatóságra törekvés és a kritikai alapállás révén. Megvizsgáltuk az olyan tudományos szabványokat, mint az empirikus megalapozottság és a cáfolhatóság, amelyek elválasztják a tudományt az áltudományosságtól. Ezt követően vizsgáltuk a hagyományos kvalitatív-kvantitatív dichotómiát. Azt állítottuk, hogy a két fogalom és a hozzájuk rendelt jellemzők merev szembeállítása félrevezető. Javasoltunk egy gyakorlatiasabb megközelítést, amely különbséget a strukturált és strukturálatlan adatok között határoz meg. Végül részletesen foglalkoztunk a tudományos megismerés egyik legnagyobb kihívásával, a zavaró tényezők problémájával. Rámutattunk, hogy bár a fogalom leggyakrabban a matematikai elemzésekben kerül elő, a torzító tényezők azonosítása és kezelése a hermeneutikai kutatásokban is elengedhetetlen. Megkülönböztettük a mediátor, moderátor, confounder és szupresszor változókat, és bemutattuk, hogy míg a strukturált adatoknál statisztikai kontrollal, addig a strukturálatlan adatok elemzésekor mélyreható kontextualizálással és kutatói reflexivitással kerülhetjük el a látszólagos összefüggések csapdáját. Áttekintettük azokat a módszereket – a randomizálástól a regressziós modellekig –, amelyekkel ezek a zavaró hatások a kutatás különböző szakaszaiban kezelhetők. Felismertük, hogy bár a társadalmi valóság rendkívül összetett, az egyszerűsítő, de jól megalkotott modellek hasznosak lehetnek a jelenségek magyarázatában. Mindez elengedhetetlen ahhoz, hogy a következő fejezetekben bemutatott kutatási módszereket megfelelő összefüggésben tudjuk értelmezni és alkalmazni.

10. Önellenőrző kérdések

1. Mi a különbség a hétköznapi és a tudományos megismerés között? Mondjon konkrét példákat mindkettőre!

Válasz: A hétköznapi megismerés személyes tapasztalatokon alapul, szelektív, gyakran anekdotikus és praktikus célokat szolgál (például „a szomszédom szerint a C-vitamin meggyógyította a náthájából"). A tudományos megismerés szisztematikus, objektivitásra törekvő, kritikai és átlátható (például randomizált kontrollált vizsgálat a C-vitamin hatásáról, pontos módszertannal, statisztikai elemzéssel).

2. Magyarázza el egy példán keresztül, mi a különbség a mediátor, moderátor és confounder változók között!

Válasz: Vegyük a „tanulási idő → vizsgaeredmény" kapcsolatot. Mediátor: a megértés szintje (a tanulási idő a megértésen keresztül hat a vizsgaeredményre). Moderátor: a tanulási módszer (a hatékony módszerrel tanulóknál erősebb a tanulási idő hatása). Confounder: az intelligencia (az intelligensebb diákok többet tanulnak ÉS jobban teljesítenek, így ha nem kontrollálunk az intelligenciára, túlbecsüljük a tanulási idő önálló hatását).

3. Miért problematikus a kvalitatív-kvantitatív dichotómia? Mi a könyvben javasolt alternatív megközelítés?

Válasz: A dichotómia áldilemmát teremt, mintha választani kellene számok és jelentések között. Elfedi, hogy sok módszer mindkét elemet tartalmazza. A javasolt alternatíva: strukturált (matematikai algoritmusokkal feldolgozható) és strukturálatlan (hermeneutikai értelmezést igénylő) adatok szerinti felosztás.

4. Soroljon fel legalább öt ismérvet, amely alapján megkülönböztethető a tudományos megközelítés az áltudományosságtól!

Válasz: Tudományos: falszifikálható hipotézisek, empirikus bizonyítékok, reprodukálhatóság, nyitottság a kritikára, fejlődőképesség. Áltudományos: cáfolhatatlan állítások, anekdotikus bizonyítékok, titkolt módszerek, a kritika elutasítása, változatlan tanok.

5. Mi a dihidrogén-monoxid példa tanulsága a tudományos kommunikáció szempontjából?

Válasz: A példa (a víz tudományos nevének használata ijesztő összefüggésben) rámutat, hogy: a) a tudományos terminológia félreérthető lehet megfelelő kontextus nélkül; b) a tények szelektív bemutatása torzíthat; c) a tudományos írástudás hiánya sebezhetővé tesz a manipulációval szemben; d) fontos a kritikus gondolkodás és a források ellenőrzése.

Ajánlott irodalom

Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health, 2*, 100051. https://doi.org/10.1016/j.glmedi.2024.100051

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173

Bartoš, F., Maier, M., Wagenmakers, E.-J., Nippold, F., Doucouliagos, H., Ioannidis, J. P. A., Otte, W. M., Sladekova, M., Deresssa, T. K., Bruns, S. B., Fanelli, D., & Stanley, T. D. (2024). Footprint of publication selection bias on meta-analyses in medicine, environmental sciences, psychology, and economics. *Research Synthesis Methods, 15*(3), 500–511. https://doi.org/10.1002/jrsm.1703

Blair, G., Coppock, A., & Humphreys, M. (2023). *Research design in the social sciences: Declaration, diagnosis, and redesign*. Princeton University Press.

Braga, A. A., Turchan, B. S., Papachristos, A. V., & Hureau, D. M. (2019). Hot spots policing and crime reduction: An update of an ongoing systematic review and meta-analysis. *Journal of Experimental Criminology, 15*(3), 289–311. https://doi.org/10.1007/s11292-019-09372-3

Bruckmaier, G., Krauss, S., Binder, K., Hilbert, S., & Brunner, M. (2021). Tversky and Kahneman's Cognitive Illusions: Who Can Solve Them, and Why? *Frontiers in Psychology, 12*, 584689. https://doi.org/10.3389/fpsyg.2021.584689

Carlson, S. (1985). A double-blind test of astrology. *Nature, 318*(6045), 419–425. https://doi.org/10.1038/318419a0

Cole, G. G. (2021). The Grievance Studies Affair; One Funeral at a Time: A Reply to Pluckrose, Lindsay, and Boghossian. *Sociological Methods & Research, 50*(4), 1937–1945. https://doi.org/10.1177/00491241211009949

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE.

Dean, G. O., & Kelly, I. W. (2003). Is astrology relevant to consciousness and psi? *Journal of Consciousness Studies, 10*(6-7), 175–198.

Feyerabend, P. (2010). *Against method*. Verso.

Forer, B. R. (1949). The fallacy of personal validation; a classroom demonstration of gullibility. *Journal of Abnormal Psychology, 44*(1), 118–123. https://doi.org/10.1037/h0059240

Hardwicke, T. E., Thibault, R. T., Clarke, B., Moodie, N., Crüwell, S., Schiavone, S. R., Handcock, S. A., Nghiem, K. an, Mody, F., Eerola, T., & Vazire, S. (2024). Prevalence of Transparent Research Practices in Psychology: A Cross-Sectional Study of Empirical Articles Published in 2022. *Advances in Methods and Practices in Psychological Science, 7*(4), Article 25152459241283477. https://doi.org/10.1177/25152459241283477

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900–1902. https://doi.org/10.1126/science.1128898

Kakamad, F. H., Abdalla, B. A., Abdullah, H. O., Omar, S. S., Mohammed, S. H., Ahmed, S. M., Mohammed, K. K., Hassan, H. A [Hemn A.], Baba, H. O., Ahmed, J. O., Mustafa, M. Q., Omar, D. A., Salih, R. Q., Rahim, H. M., Hussein, D. A., Hassan, M. N., Mikael, T. M., Hassan, H. A [Hunar A.], & Najar, K. A. (2024). Lists of predatory journals and publishers: a review for future refinement. *European Science Editing, 50*, Article e118119. https://doi.org/10.3897/ese.2024.e118119

Korbmacher, M., Azevedo, F., Pennington, C. R., Hartmann, H., Pownall, M., Schmidt, K., Elsherif, M., Breznau, N., Robertson, O., Kalandadze, T., Yu, S., Baker, B. J., O'Mahony, A., Olsnes, J. Ø.-S., Shaw, J. J., Gjoneska, B., Yamada, Y., Röer, J. P., Murphy, J., . . . Evans, T. (2023). The replication crisis has led to positive structural, procedural, and community changes. *Communications Psychology, 1*(1), 3. https://doi.org/10.1038/s44271-023-00003-2

Kuhn, T. S. (2012). *The structure of scientific revolutions*. The University of Chicago Press.

Lagerspetz, M. (2021). „The Grievance Studies Affair" Project: Reconstructing and Assessing the Experimental Design. *Science, Technology, & Human Values, 46*(2), 402–424. https://doi.org/10.1177/0162243920923087

Lim, W. M. (2025). What Is Qualitative Research? An Overview and Guidelines. *Australasian Marketing Journal, 33*(2), 199–229. https://doi.org/10.1177/14413582241264619

Ng, C. K. J., Kwan, L. Y. J., & Chan, W. (2024). A Note on Evaluating the Moderated Mediation Effect. *Structural Equation Modeling: A Multidisciplinary Journal, 31*(2), 340–356. https://doi.org/10.1080/10705511.2023.2201396

NISO CRediT Working Group. ANSI/NISO Z39.104-2022, CRediT, Contributor Roles Taxonomy. NISO.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology, 73*, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Pearl, J. (2013). *Causality*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Popper, K. R. (2002). *The logic of scientific discovery*. Routledge.

Richtig, G., Berger, M., Koeller, M., Richtig, M., Richtig, E., Scheffel, J., Maurer, M., & Siebenhaar, F. (2023). Predatory journals: Perception, impact and use of Beall's list by the scientific community-A bibliometric big data study. *PloS One, 18*(7), e0287547. https://doi.org/10.1371/journal.pone.0287547

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

II. INTERPRETATÍV FENOMENOLÓGIAI ELEMZÉS (IPA)

TURÓS MÁTYÁS

A fejezet a következő kérdéseket válaszolja meg:

a) Milyen elméleti alapokon nyugszik az IPA, és mit jelent a fenomenológiai, hermeneutikai és idiografikus elköteleződés a gyakorlatban?
b) Melyek az IPA-kutatás gyakorlati lépései a kutatási kérdés megfogalmazásától a szisztematikus adatelemzésig?
c) Milyen kutatási témák és kérdések vizsgálatára alkalmas az IPA, és mikor kell más módszert választanunk helyette?
d) Milyen minőségi ismérvek alapján különböztethető meg egy kiváló IPA-tanulmány egy átlagostól?
e) Hogyan terjeszthető ki az IPA a tapasztalatok időbeli változásának (longitudinális IPA) és kulturális beágyazottságának (participatív IPA) vizsgálatára?

1. Bevezetés

Az interpretatív fenomenológiai elemzés (Interpretative Phenomenological Analysis, IPA) azt vizsgálja, hogy miként értelmezik az emberek a személyes és társadalmi világukat, hogyan tesznek szert nagy jelentőségű tapasztalatokra. Az IPA egyértelműen a strukturálatlan adatokkal dolgozó módszerek közé tartozik, célja a megélt tapasztalatok mély, részletgazdag feltárása. A módszert Jonathan A. Smith pszichológus és munkatársai fejlesztették ki, és bár a pszichológiában, különösen az egészségpszichológiában a legismertebb, ma már széles körben alkalmazzák a társadalomtudományok más területein is. A módszer nevében is tükrözi kettős intellektuális gyökerét. Fenomenológiai, mert a résztvevők közvetlen, megélt tapasztalatát igyekszik megragadni, és interpretatív (értelmező), mert elismeri, hogy a kutató ehhez a tapasztalathoz mindig csak a résztvevő elbeszélésének, vagyis a tapasztalatról alkotott értelmezésének elemzésén keresztül férhet hozzá. Ez a kettősség – a résztvevő világához való hűségre való törekvés és a kutatói értelmezés elkerülhetetlenségének elfogadása – adja a módszer egyedi jellegét.

2. Elméleti alapok

Az IPA egyedi jellegét három, egymást kiegészítő tudományfilozófiai hagyomány alakítja, a fenomenológia, a hermeneutika és az idiográfia.

A fenomenológia az IPA filozófiai magja. A fenomenológia legismertebb képviselője Edmund Husserl. Filozófiája a megélt tapasztalat vizsgálatát helyezte a megismerés középpontjába. Husserl híres felhívása, hogy „térjünk vissza a dolgokhoz magukhoz", arra ösztönöz, hogy a jelenségeket a maguk tiszta formájában ragadjuk meg, ahogyan a tudatunkban megjelennek, mielőtt előzetes elméleteink vagy a hétköznapi gondolkodásunk (a „természetes beállítódás") kategóriáiba sorolnánk őket. Ezért a kutatónak munkája során törekednie kell saját előfeltevéseinek zárójelbe tételére. Az IPA képviselői szerint a teljes előfeltevés-mentesség, a husserli értelemben vett epoché a gyakorlatban elérhetetlen

ideál. Ehelyett a módszer sokkal inkább a reflexivitás fontosságát hangsúlyozza. A kutató nem zárja ki saját előfeltevéseit, hanem folyamatosan tudatosítja azokat, és reflektál azok szerepére az értelmezési folyamatban.

Az IPA nem a hétköznapi, automatikus cselekvésekre kíváncsi, hanem azokra a kitüntetett életeseményekre, amelyek az egyén számára különös jelentőségűek, és reflektív gondolkodást indítanak el. Az IPA célja nem általánosítható összefüggések feltárása, hanem a szubjektív jelentések és tapasztalatok megértése. Ezáltal a kutatás tárgya a résztvevő életvilága lesz.

A hermeneutika az értelmezés elmélete és gyakorlata. Azt hangsúlyozza, hogy minden tapasztalat már eleve értelmezett tapasztalat, a megértés mindig történetileg és kulturálisan kondicionált, és a jelentés dialogikus folyamatban alakul ki. Ezek az elméleti alapok konkrét kutatásmetodológiai megközelítésekben is megnyilvánulnak. Amikor a kutató egy elbeszéléssel találkozik, akkor már eleve egy értelmezett tapasztalattal dolgozik. A kutatási folyamat ezért kettős hermeneutika.[8] A kutató megpróbálja értelmezni azt a folyamatot, ahogyan a résztvevő értelmezi a saját világát. Ebből adódik, hogy a kutató nem befogadója az adatoknak, hanem résztvevője a jelentésalkotásnak. Az elemzés során a hermeneutikai kör logikáját követi: a rész (például egy mondat vagy egy szövegrészlet) megértése hozzájárul az egész (az interjú kontextusa) megértéséhez, és fordítva, az egész kontextusa ad jelentést a részeknek. Ez az oda-vissza mozgó, ismétlődő folyamat teszi lehetővé a jelentésrétegek fokozatos feltárását. A kutató saját előzetes tudása és tapasztalatai ugyan elkerülhetetlenül részét képezik a folyamatnak, mégis a cél az, hogy egy szorosan az adatokban gyökerező értelmezés alakuljon ki. A hermeneutikai hagyomány a törvényszerűségeket kereső (nomotetikus) megközelítéssel szemben az egyedi eset részletes vizsgálatára helyezi a hangsúlyt (idiografikus megközelítés).[9] Ezeknek az elveknek egyik jelentős megvalósítása az IPA, amely egyértelműen a hermeneutikai és fenomenológiai hagyományra épít. Az IPA idiografikus elköteleződése a kutatás minden szintjén megnyilvánul:

[8] Az IPA kettőshermeneutika-fogalma különbözik az Anthony Giddens által használt szociológiai terminustól. Míg Giddensnél a fogalom arra utal, ahogy a társadalomtudományi elméletek és a társadalmi gyakorlat folyamatosan alakítják egymást, addig Jonathan Smith nyomán az IPA kontextusában ez a kifejezés a kutatási folyamat kétszintű interpretatív jellegét jelöli. Az első szinten a résztvevő próbál értelmet adni saját élményeinek (értelmezi az életvilágát), a második szinten pedig a kutató próbál értelmet adni a résztvevő által megfogalmazott jelentéseknek (értelmezi a résztvevő interpretációját).

[9] A nomotetikus és idiografikus megközelítések szétválasztása Wilhelm Windelband német filozófus 1894-es rektori székfoglaló beszédére (Geschichte und Naturwissenschaft) vezethető vissza. Míg a természettudományok célja az általános törvények feltárása (nomotetikus), addig a szellemtudományoké az egyedi jelenségek megértése (idiografikus). Jonathan Smith Windelbandra hivatkozva emelte be az idiográfiát az IPA alapkövei közé.

a) Cél a kis mintán végzett mélyreható elemzés. A kutató célzottan választ ki egy homogén csoportot, vagyis olyan személyeket, akik egy adott tapasztalat szempontjából hasonló helyzetben vannak. A cél nem a statisztikai reprezentativitás, hanem a gazdag, összefüggésrendszerbe ágyazott adatok gyűjtése, amelyekből a jelenség mélyen megérthető.
b) Az elemzés mindig az egyes esettel kezdődik. Minden interjút a maga egyediségében, rendkívül részletesen kell elemezni, és csak ezt követően kerülhet sor az esetek közötti mintázatok, a konvergenciák és divergenciák feltárására.
c) Az IPA nem állít egyetemes törvényeket. Célja, hogy részletes, árnyalt betekintést nyújtson egy adott jelenségről adott körülmények között. Az általánosítás lehetősége nem statisztikai, hanem elméleti, vagyis az olvasó a gazdag leírás alapján döntheti el, hogy a kutatás eredményei mennyire alkalmazhatók más, hasonló helyzetekre.

3. A kutatás folyamata

3.1. Adatgyűjtés

Az IPA elsődleges adatgyűjtési módszere a félig strukturált mélyinterjú. Ez a technika biztosítja a szükséges rugalmasságot ahhoz, hogy a kutató kövesse a résztvevő gondolatmenetét és a számára fontos témákat, miközben egy interjúvázlat segítségével biztosítja, hogy a kutatás szempontjából lényeges területek mindegyike szóba kerüljön. Az adatfelvétel folyamata több lépésből áll:

a) Az előzetes megkeresésnek hangsúlyoznia kell a kutatás fontosságát és a résztvevő hozzájárulásának egyedi értékét.
b) A kutatónak alkalmazkodnia kell a résztvevőhöz az adatfelvétel helyét és idejét illetően. Az ideális egy nyugodt, zavartalan környezet, ahol a résztvevő kényelmesen és biztonságban érzi magát.
c) Az interjú elején a kutatónak szóban és írásban is tájékoztatnia kell a résztvevőt a kutatás céljáról, a hangfelvétel készítéséről, a névtelenség teljes körű biztosításáról és az adatok kezelésének módjáról. A tájékozott beleegyező nyilatkozat aláírása elengedhetetlen.

d) Az interjút minden esetben hangfelvétellel kell rögzíteni. Ez teszi lehetővé a szó szerinti átirat elkészítését, ami az IPA-elemzés alapfeltétele.

Bár az interjú rugalmas, nem rögtönzött. A kutatónak egy átgondolt interjúvázlattal kell készülnie, amely a kutatási kérdésből bomlik ki. Az interjúvázlatnak öt-nyolc főbb, nyitott kérdést kell tartalmaznia, amelyek alatt egy-három segítő vagy követőkérdés (prompt) helyezkedhet el. A kérdéssornak mindig egy általános, bevezető kérdéssel kell kezdődnie, amely elegendő teret ad a résztvevőnek a téma szabad kibontásához.

Minden interjúkérdésnek – típusától függetlenül – meg kell felelnie két alapvető követelménynek:

a) Nyitottság. A kérdés tegye lehetővé, hogy a résztvevő szabadon és részletesen kifejezze gondolatait, saját szavaival és saját tempójában. Kerülendők az eldöntendő (igen/nem) kérdések. Például: „Hogyan élte meg ezt a helyzetet?" vagy „Kérem, meséljen arról az időszakról…".
b) Semlegesség, nem irányító jelleg. A kérdés ne sugalmazzon választ, és ne befolyásolja a résztvevőt a kutató előfeltevéseinek irányába. Például: „Mit gondol erről a helyzetről?" – az „Ugye egyetért azzal, hogy ez nehéz volt?" típusú, sugalmazó kérdésekkel szemben.

Ezeken az alapelveken belül az IPA-interjúban többféle kérdéstípus alkalmazható a megcélzott tartalom függvényében. A leggyakrabban használtak:

a) Leíró kérdések, amelyek egy helyzet, esemény vagy tapasztalat konkrét bemutatására ösztönöznek. Például: „Mi történt azon a napon?"
b) Narratív kérdések, amelyek egy történet elmesélésére hívnak, és lehetővé teszik az események időbeli kibontását. Például: „Hogyan kezdődött ez az időszak az életében?"
c) Értékelő kérdések, amelyek a résztvevő értékelését, önreflexióját és érzelmi reakcióját segítik elő. Például: „Visszatekintve, hogyan értékeli ennek a döntésnek a hatását az életére?"
d) Kontrasztív kérdések, amelyek különböző helyzetek, időszakok vagy szempontok összehasonlítására ösztönöznek. Például: „Miben különbözött ez az időszak a korábbiaktól?"

Mivel az IPA célja a megélt tapasztalat mély, részletgazdag feltárása, a kiváló minőségű interjú elengedhetetlen. Ennek érdekében a kutató feladata egy olyan bizalmi légkör megteremtése, amelyben a résztvevő hajlandó és képes

megosztani személyes jelentéseit. A személyes jelentések megosztásával összefüggő fogalom az „önfeltárás szintje". Az önfeltárás szintje azt mutatja meg, hogy a résztvevő milyen mélységben és őszinteséggel beszél a tapasztalatairól. A válaszadás mélysége szerint általában három szintet különböztetünk meg:

a) Felszíni szint. A résztvevő általánosságokban, közhelyesen beszél, gyakran társadalmilag elvárt válaszokat ad (például: „A tanári pálya szép hivatás, sokat ad az embernek.").
b) Közepes mélység. A résztvevő már személyes példákat is hoz, de még tartózkodik a mélyebb érzelmek, dilemmák vagy belső konfliktusok megosztásától (például: „Volt egy nehéz osztályom tavaly, sok energiámat felemésztette.").
c) Mély szint. A résztvevő őszintén és részletesen beszél a személyes élményeiről, érzéseiről, belső küzdelmeiről és azok jelentéséről (például: „Amikor az az osztályom volt, komolyan elgondolkodtam a pályaelhagyáson. Minden este szorongva mentem haza…").

Az IPA-kutató aktívan törekszik a harmadik szint elérésére. Ennek érdekében a következő eszközöket alkalmazhatja:

a) empatikus, ítélkezésmentes hozzáállás,
b) cselekvő és figyelmes hallgatás,
c) türelem és a csend (a csend időt ad a résztvevőnek a gondolatai megfogalmazására, és gyakran a legmélyebb önreflexiókhoz vezet).

Az önfeltárás szintjének felmérése és írásban történő rögzítése az elemzés szempontjából is létfontosságú. A kutatónak az interjú után érdemes azonnal kutatói jegyzeteket készítenie a benyomásairól, rögzítve a beszélgetés dinamikáját, a kiemelkedő nem szóbeli jeleket vagy az esetleges hárító viselkedést. Ezek a jegyzetek többletjelentést adnak az elemzéshez.

Miután a jó minőségű, mély adatokat tartalmazó hangfelvétel elkészült, a következő lépés a szó szerinti átirat elkészítése, amely megalapozza a részletes fenomenológiai elemzést. Az interjú után a hanganyagot szó szerint kell átírni. Az IPA esetében nincs szükség a beszéd finom prozódiai jellemzőinek (például hangsúly, hanglejtés) részletes rögzítésére, de a szöveget ne tisztítsuk meg, mert jelentéshordozók a hezitálások, az ismétlések, a töltelékszavak és a félbehagyott mondatok is, és ezeket pontosan le kell jegyezni.

4. doboz: A mélyinterjúk előkészítése és levezetése IPA-kutatásban

Gyakorlati tanácsok az interjúk sikeres levezetéséhez

I. Előkészületek és technikai háttér

Az időpontot mindig a résztvevő igényeihez igazítsuk. A legjobb helyszín a résztvevő otthona vagy egy általa választott, nyugodt, zavartalan hely, például az irodája munkaidőn kívül. A beszélgetésen csak a kutató és a résztvevő legyen jelen.
Már az előzetes megkeresésben (e-mail, telefon) hangsúlyozzuk a nyugodt, zavartalan körülmények fontosságát. Ez azért fontos, mert a résztvevő felkészül egy elmélyült beszélgetésre.

Tájékoztassuk a résztvevőt, hogy az interjú 45–180 percet is igénybe vehet. Ez segít komolyan venni a helyzetet, és felkészíti a válaszadót arra, hogy ne siessen. Ne feledjük: ha az előzetes megkeresés során háromórás interjút jelzünk előre, kisebb az esélye annak, hogy a beszélgetés végül félórásra zsugorodik, mintha eleve csak egy órát említettünk volna.

A technikai hibák elkerülése érdekében mindig készítsünk egy póthangfelvételt. Egy diktafont helyezzünk jól láthatóan az asztalra (ez a szoftveres átíráshoz is kedvező), miközben egy másik eszköz diszkréten, tartalékként szintén rögzít.

Ha a résztvevő nem egyezik bele a hangfelvétel készítésébe, az interjút nem lehet lefolytatni, mivel kézzel az IPA-elemzéshez szükséges részletességgel jegyzetelni gyakorlatilag lehetetlen.

II. A kutató szerepe és viselkedése

A kutató viselkedése, öltözködése és szóhasználata teremti meg a bizalmat. Legyünk önazonosak és hitelesek, kerüljük a mesterkélt formalitást és az erőltetett, barátkozó stílust egyaránt.

A kutató előtt legyen ott a kérdéssor, de ne mechanikusan „lekérdezze” azt. Vagyis kövesse a beszélgetés természetes ívét, ha a helyzet úgy kívánja, térjen el a sorrendtől, és hagyjon ki egy témát, ha az nem odavágó – de törekedjen arra, hogy a főbb területeket érintse. Előfordulhat, hogy a válaszadó előre kifejt olyan témát, amelyre későbbi kérdés vonatkozott volna. Ilyen esetben a kutató ne tegye fel még egyszer a kérdést, egy ilyen hiba teljesen elkedvetlenítheti a válaszadót.

III. A beszélgetés levezetése

Az interjú érzékeny folyamat, amely éber, pártatlan figyelmet követel. A cél a résztvevő pszichológiai és szociális világának megértése empatikus, nem ítélkező módon.

Az interjú elején szánjunk három-öt percet a legfontosabb tudnivalók rövid ismétlésére (kutatás célja, anonimitás), és kérdezzünk rá újra a hangfelvételbe való beleegyezésre.

Egyszerre mindig csak egy kérdést tegyünk fel. A cél, hogy a résztvevő beszéljen, ne a kutató. Kevés, de célzott kérdést tegyünk fel. Egy-egy érdekes gondolatmenet továbbviteléhez gyakran elég ennyi: „Tudna erről többet mondani?" vagy „És mit érzett ezzel kapcsolatban?".

A kutatónak tudnia kell viselni a csendet. Egy 10 másodperces (akár hosszabb) szünet teret ad a résztvevőnek a gondolkodásra és a mélyebb önreflexióra. Ezt a szabályt nem szabad lebecsülni, de figyelni kell arra is, hogy a csend mikor válik feszültté, és mikor vár a résztvevő újabb kérdésre.

A kutató nem fejezheti ki tetszését vagy nemtetszését, sem szóban, sem gesztussal, sem arcjátékkal. Ugyanakkor az erőltetett pártatlanság és a folyamatos pozitív megerősítés is zavaró lehet. Az empátia a megértésre való törekvést jelenti, nem a semlegességet vagy az egyetértést.

Figyeljük meg, hogyan hat a beszélgetés a résztvevőre. Ha egy téma láthatóan kényelmetlen, tapintatosan lépjünk vissza, és közelítsünk más irányból. A résztvevőt a saját tapasztalatának szakértőjeként kell kezelnünk.

A gyakorlatban a kutató olyan kérdéseket használ, amelyek a résztvevő szubjektív világának feltárására irányulnak. Tipikus kérdésfordulatok lehetnek például: „El tudná mondani, hogy…?", „Hogyan érzi magát, amikor…?", „Hogyan jellemezné a helyzetet?", „Mit jelent az Ön számára a… kifejezés?", „Hogyan tekint most önmagára annak tükrében, hogy…?", „Miben látja most magát másnak szakmailag, mint korábban?".

3.2. Adatelemzés

Az IPA-elemzés módszeres, aprólékos és ismétlődő folyamat. Az iteráció az IPA-ban két szinten valósul meg. Egyrészt a hermeneutikai kör logikája szerint a kutató folyamatosan vált a részletek (szavak, mondatok) és az egész (teljes kontextus, témák) értelmezése között. Másrészt csoportos kutatás esetén az iteráció a konszenzuális validálás folyamatát jelenti. Ez nem átlagolást vagy kötelező egyetértést takar, hanem az eltérő kutatói perspektívák szisztematikus ütköztetését és finomítását, amíg egy gazdagabb, több szempontot integráló értelmezési keret létre nem jön.

Mindkét szinten érvényesül, hogy az iteráció révén a nyers, strukturálatlan szövegből fokozatosan feltárulnak a résztvevő tapasztalatainak mélyebb, pszichológiai jelentésrétegei. A hitelesség növelése érdekében a kutatócsoportok gyakran alkalmazzák a megtárgyalt konszenzus módszerét, ahol a kutatók egymástól függetlenül elemzik az adatokat, majd megvitatják az eredményeket. A cél ilyenkor nem a teljes egyetértés, hanem a több szempontú, gazdag értelmezés kialakítása. Az elemzés fő lépései a következők, amelyek minden egyes interjú esetében megismétlődnek:

Első lépés: elmélyülés az adatokban, olvasás és újraolvasás. Az elemzés a teljes interjúátirat többszöri elolvasásával kezdődik. A kutatónak legalább kétszer, lehetőség szerint egy-két nap kihagyással érdemes elolvasnia a szöveget, miközben párhuzamosan hallgatja a hangfelvételt. A cél a teljes elmerülés a résztvevő világában, és ezáltal megérteni a ritmust, a hanghordozást, átérezni az elbeszélés hangulatát – mielőtt a formális elemzés megkezdődne.

Második lépés: a kezdeti felfedezés, jegyzetelés. Ebben a szakaszban a kutató sorról sorra haladva, az átirat bal oldali margójára szabadon jegyzetel. Itt még minden adat, nincsenek jó vagy rossz részek, és a cél a lehető legnyitottabb hozzáállás. A jegyzetek három fő típusba sorolhatók, amelyek segítenek a tapasztalat különböző rétegeinek megragadásában:

a) Leíró kommentárok, amelyek a szöveg tartalmára és a „mit mond?” kérdésre összpontosítanak. Ezek lehetnek kulcsszavak, frázisok, események vagy érzelmek kiemelései. Példa: „Szorongás említése a vizsgahelyzet kapcsán”, „Családi viszály leírása”, „Pozitív munkahelyi élmény”.
b) Nyelvi kommentárok, amelyek a „hogyan mondja?” kérdésre összpontosítanak. A nyelvhasználat, a metaforák, az ismétlések, a hangsúlyok, a szünetek, a nevetés vagy a hangnem változásai tartoznak ide. Példa:

„Háborús metafora használata a betegséggel való küzdelem leírásakor", „Gyakori ‚talán' szó – bizonytalanságot jelez?", „Erős érzelmi töltetű szavak".

c) Fogalmi kommentárok, amelyek a kutató első, kezdeti értelmezései, kérdései, absztrakciói. Itt már megjelenik a kettős hermeneutika. Példa: „Ez az identitásválság jele lehet?", „Ambivalencia a karrierváltással kapcsolatban", „A társadalmi normák befolyása érződik a kijelentés mögött".

Harmadik lépés: a mintázatok kialakítása, témák azonosítása egy eseten belül. A kutató a jegyzeteket vizsgálva azonosítani kezdi az ismétlődő mintázatokat és kapcsolatokat. A hasonló jelentésű elemeket csoportosítja, és magasabb szintű, pszichológiai jelentéssel bíró kialakuló témákká (emergent themes) sűríti. Ezek a témák hierarchiába vagy klaszterekbe rendeződhetnek, ahol egy-egy főtéma köré szerveződnek a kapcsolódó gondolatok. Ebben a szakaszban alakulnak ki a fölérendelt témák (super-ordinate themes) és a hozzájuk tartozó altémák (sub-themes), amelyek az adott résztvevő tapasztalatának főbb aspektusait képviselik.

Negyedik lépés: rendszerezés, megjelenítés és az elemzési táblázat. A kialakított témákat egy táblázatban érdemes rendszerezni. Ez a táblázat nemcsak az átláthatóságot segíti, hanem az esetek közötti összehasonlítás alapjául is szolgál. Az alábbiakban egy kitalált példán keresztül mutatjuk be, hogyan nézhet ki egy ilyen elemzési táblázat egy pályakezdő tanárral készített interjú alapján. A kutatás azt vizsgálja, hogyan élik meg a kezdő pedagógusok az első tanévüket egy hátrányos helyzetű iskolában.

Fölérendelt téma (Super-ordinate Theme)	Altéma (Sub-theme)	Idézet az interjúból	Helye az átiratban
1. Szakmai és érzelmi küzdelmek	1.1. Konfliktusok és a fegyelmezés tehetetlensége	„Egyszerűen nem tudtam mit kezdeni a folyamatos bekiabálásokkal. Voltak napok, amikor teljesen tehetetlennek éreztem magam."	(4. oldal, 112–113. sor)

	1.2. Elszigeteltség a tantestületben	„Senki nem jött oda megkérdezni, hogy vagyok. Mindenki rohant a dolgára, mintha én ott se lennék, csak egy újabb átutazó.”	(7. oldal, 201–202. sor)
	1.3. Önbizalomhiány és a pályaelhagyás gondolata	„Volt olyan este, hogy sírva gondoltam arra, hogy felmondok. Hogy ez nekem nem fog menni, rossz döntést hoztam.”	(9. oldal, 255–256. sor)
2. A hivatástudat és a kis sikerek ereje	2.1. Egy-egy diák áttörése mint feltöltődés	„De amikor Peti, aki egész évben meg se mukkant, feltette a kezét és válaszolt egy kérdésre… az mindenért kárpótolt.”	(12. oldal, 310–311. sor)
	2.2. A felelősségérzet mint megtartóerő	„Tudtam, hogy ha én feladom, nekik talán senki más nem marad, aki hinne bennük. Ez tartott itt a legnehezebb napokon is.”	(14. oldal, 345–346. sor)

2. táblázat: IPA elemzési táblázat minta

Ötödik lépés: az elemzés kiterjesztése a következő esetre. A kutató megismétli az 1–4. lépéseket a minta következő tagjával. Ennek a lépésnek a kritikus pontja a reflexív zárójelezés. A kutatónak tudatosítania kell az előző eset elemzése során kialakult témákat és előfeltevéseket, és törekednie kell arra, hogy az új eset a maga egyediségében szólalhasson meg.

Hatodik lépés: az eseteken átívelő mintázatok keresése. Miután a kutató minden esetet külön-külön, a maga mélységében kielemzett, összeveti az egyes esetekhez készült tématáblázatokat. Keresi a közös pontokat, az eltéréseket, a konvergenciákat (ahol a résztvevők tapasztalatai hasonlóak) és a divergenciákat (ahol az élmények eltérőek vagy egyediek). Ebből a szintézisből emelkednek ki a teljes mintára jellemző, átfogóbb csoportos tapasztalati témák (Group Experiential Themes, GETs). Ezek a végső, magas szintű absztrakciók adják a kutatás fő eredményeit. A folyamatot összesítő táblázat segítségével tehetjük átláthatóvá.

Csoportos tapasztalati téma (GET)	Hozzájáruló egyéni témák a résztvevőktől	Illusztratív idézet (egy résztvevőtől)	Megjegyzések (konvergencia és divergencia)
1. GET: Az ideálok és a valóság ütközése	Anna: Szakmai és érzelmi küzdelmek; Béla: Túlélési stratégiák a káoszban; Cecília: Identitásválság: tanár vagy szociális munkás?	„Azt hittem, én majd megváltom a világot. Ehhez képest az első hetekben azért imádkoztam, hogy valahogy éljem túl a napot." (Cecília)	Konvergencia: Mindhárom résztvevőnél központi élmény volt a kezdeti sokk és a képzéseken tanult ideálok és a mindennapi gyakorlat közötti szakadék. Divergencia: Az ütközés más-más formában csapódott le: Annánál érzelmi kimerültségben, Bélánál a gyakorlati eszközök keresésében, Cecíliánál pedig a szakmai szerepének megkérdőjelezésében.

2. GET: A „mikro-győzelmek" mint a túlélés zálogai	Anna: Egy-egy diák áttörése mint feltöltődés; Béla: A működő tanóra mint sikerélmény; Cecília: Egy sikeres szülői megbeszélés	„De amikor Peti, aki egész évben meg se mukkant, feltette a kezét és megválaszolt egy kérdést… az mindenért kárpótolt." (Anna)	Konvergencia: Mindhárom résztvevő értelmezéseiben szerepet játszottak az apró, de jelentőségteljes sikerek, amelyek erőt adtak a folytatáshoz a nehézségek közepette. Ez egy egyetemes tapasztalat volt a mintában.
3. GET: Navigálás a tantestületi kapcsolatok útvesztőjében	Anna: Elszigeteltség a tantestületben; Béla: Viszály a régi gárdával; Cecília: Egy tanácsadó megtalálása	„Senki nem jött oda megkérdezni, hogy vagyok. Mindenki rohant a dolgára, mintha én ott se lennék, csak egy újabb átutazó." (Anna)	Konvergencia: A tantestületi beilleszkedés mindenkinek kihívást jelentett. Divergencia: A megküzdés módja és az élmény jellege eltérő volt: Anna tétlen elszigeteltséget élt meg, Béla éles konfliktusba keveredett a tapasztaltabb munkatársakkal, míg Cecília – egyedüliként – talált egy tanácsadót, ami segítette a beilleszkedését.

3. táblázat: IPA csoportos tapasztalati témák (GETs) táblázatminta

Hetedik lépés: a mélyebb jelentés feltárása és értelmezés. Az elemzés utolsó szakasza a puszta leíráson túlmutató értelmezés. A kutató a feltárt témákat és mintázatokat összefüggő értelmezési keretbe rendezi, és minden állítását az interjúból származó idézetekkel támasztja alá. Ebben a szakaszban a kutató a kettős hermeneutika második szintjén dolgozik, vagyis nemcsak azt rögzíti, amit a résztvevő elmond, hanem értelmezi a résztvevő önértelmezését is – miközben ügyel arra, hogy értelmezései szorosan az adatokban gyökerezzenek. Ebben a szakaszban érdemes mélyebb, értelmező kérdéseket feltenni a szöveggel

kapcsolatban: „Mit akar itt a résztvevő elérni ezzel a kijelentéssel?”, „Megjelenik-e itt valami, ami nem volt szándékos?”, „Úgy érzem-e, hogy itt valami olyasmi fogalmazódik meg, amivel talán maga a résztvevő sincs tisztában?”.

A kutatás végső eredménye gyakran egy összegző vizuális modellben (például egy hagymamodell, folyamatábra) is megjeleníthető, amely egyetlen ábrán, áttekinthetően mutatja be a vizsgált jelenség összetettségét.

5. doboz: Kvalitatív adatelemző (QDA) szoftverek szerepe az IPA-ban

Elsőre ellentmondásosnak tűnhet, hogy egy ennyire mélyen humanisztikus, interpretatív és a kutató személyes elmélyülésére építő módszert, mint az IPA, szoftverrel támogassunk. A szoftveres támogatás mégis az IPA-elemzés aprólékos folyamatának minden lépéséhez segítséget nyújthat. A kvalitatív adatelemző (QDA) szoftverek olyan digitális munkakörnyezetet biztosítanak, amely a kutatót megszabadítja az adminisztratív terhek nagy részétől, így több erőt fordíthat a jelentésrétegek feltárására. Az elemző szoftver az adatokkal való elmélyülés szakaszában lehetővé teszi az interjúátiratok és a hangfelvételek összekapcsolását, így az olvasás során bármikor visszahallgatható az eredeti hangsúly és hangnem. A kezdeti leíró, nyelvi és fogalmi kommentárok közvetlenül a szövegrészletekhez fűzhetők, digitális jegyzetekként rögzíthetők. A szoftverek legnagyobb előnye a témák fejlesztésénél és rendszerezésénél mutatkozik meg. Ahelyett, hogy a kutató papírcetlikkel és szövegkiemelőkkel dolgozna, a kezdeti kódokat rugalmasan csoportosíthatja kialakuló témákká. A témák könnyen átnevezhetők, egyesíthetők vagy szétválaszthatók, miközben a hozzájuk tartozó idézetek automatikusan követik ezeket a változásokat. Ez nagy mértékben megkönnyíti az iteratív elemzési folyamatot. A kutatói önreflexió fenntartását a szoftverekben jellemzően megtalálható „memó” vagy „jegyzet” funkció segíti, ami a kutatói napló digitális megfelelőjeként szolgál, és segít rögzíteni a kettős hermeneutika folyamatát. Végül, a kutatási jelentés megírásakor a szoftver egyetlen kattintással képes kigyűjteni az összes, egy adott témához tartozó idézetet, ami megkönnyíti az értelmezés felépítését és az állítások adatokkal való alátámasztását.

4. A minőség biztosítása

Az IPA minőségét nem technikai szabályok betartása garantálja, hanem az analitikus mesterségbeli tudás, az a képesség, hogy a kutató meggyőző, mély, adatokkal gazdagon alátámasztott és elméletileg koherens értelmezést alkosson. Az első, kifejezetten az IPA-ra szabott minőségi útmutatót Jonathan A. Smith publikálta egy átfogó irodalmi áttekintés részeként. Ez a keretrendszer ma is alapvető fontosságú a publikálhatóság megítélésében. A minőségi követelményrendszer a következő.

A tanulmánynak négy alapkövetelménynek kell megfelelnie:

a) a kutatásnak egyértelműen fenomenológiai, hermeneutikai és idiografikus elvekre kell épülnie,
b) a kutatási folyamatnak (a mintaválasztástól az elemzés utolsó lépéséig) követhetőnek és átláthatónak kell lennie az olvasó számára,
c) az elemzésnek észszerűnek, meggyőzőnek és érdekesnek kell lennie,
d) az állításokat elegendő számú idézettel kell alátámasztani.

A jó minőségű tanulmánynak az alapkövetelményeken felül három további jellemzője van:

a) sajátos témát vizsgál mélyrehatóan, nem pedig felszínesen érint sok területet,
b) a résztvevői idézetekre épülő kutatói értelmezés éleslátó,
c) a tanulmány olvasmányos, bevonja és elgondolkodtatja az olvasót.

A legmagasabb minőségű IPA-kutatásokat négy ismérv jellemzi.

a) A tanulmány koherens, meggyőző történetet mesél el. Az értelmezés fokozatosan épül fel, ahogy a kutató az idézeteket és a saját meglátásait váltogatja. Minden új idézet és elemzés hozzátesz valamit a történethez, egy lépéssel tovább viszi az olvasót a megértésben, megteremtve a hermeneutikai kör dinamikáját. Gyakorlati tanács: Az értelmezés felépítéséhez gondosan válogassa meg az idézetek sorrendjét. Ne csak

szemléltessen egy témát, hanem építsen fel egy gondolati ívet. Az első idézet bemutatja a téma alapját, a második árnyalja azt, a harmadik pedig egy új nézőpontot vagy mélységet ad hozzá.

b) A kutatás túllép a puszta élmény leírásán, és feltárja annak mélyebb, egzisztenciális jelentését. Nemcsak arról ír, hogy a résztvevő mit élt át, hanem arról is, hogy ez a tapasztalat hogyan hatott az identitására, az önmagáról, a világról, a halandóságról vagy az emberi kapcsolatokról alkotott képére. Gyakorlati tanács: Elemzés közben tegye fel a kérdést: „Mit jelent ez a tapasztalat a résztvevőnek mint emberi lénynek?" Keresse azokat a pontokat, ahol a beszámoló az identitás, az énkép, az elszigeteltség vagy az élettel való megküzdés alapvető kérdéseit érinti.

c) Az idézetek nem beszélnek önmagukért. A kutató feladata, hogy feltárja a szavak mögötti rejtett jelentésrétegeket. Ez a közeli olvasat a nyelvhasználat finom részleteire összpontosít, azaz a szóválasztásra, a metaforákra, az ismétlésekre, a hezitálásokra vagy a hangsúlyokra. Gyakorlati tanács: Ne csak azt írja le, mit mondott a résztvevő, hanem azt is, hogyan mondta. Például: „A résztvevő harcos metaforát használva (»meg kellett küzdenem vele«) írja le a betegségét, ami a helyzet tehetetlenséggel és agresszióval teli megélésére utal."

d) A cél nem pusztán az, hogy megmutassuk, egy közös téma hogyan jelenik meg különböző, egyedi élettörténetekben; az elemzésnek mesterien egyensúlyoznia kell a közös, eseteken átívelő mintázatok (konvergencia) bemutatása és az egyes résztvevők egyedi, megismételhetetlen tapasztalatainak (divergencia) tiszteletben tartása között. Mutassa be, hogyan kapcsolódik egy egyéni idézet a csoport egészének tapasztalatához. Hozzon példát arra, amikor egy résztvevő másként él meg valamit, mint a többiek, és értelmezze, miért lehet ez így (például az élettörténete, a személyisége miatt).

Minőségi követelmények	Mit jelent a gyakorlatban?
Módszertani átláthatóság	A kutatás minden lépése (mintaválasztás, interjúkérdések, elemzés) világosan és részletesen le van írva.
Elméleti hűség	A kutatás egyértelműen a fenomenológia, hermeneutika és idiográfia elveire épül.
Meggyőző értelmezés	Az eredmények egy észszerű, koherens történetet mesélnek el, ahol az idézetek és az értelmezés fokozatosan építik fel a jelentést.

Egzisztenciális mélység	Az elemzés túllép a felszínes leíráson, és a tapasztalatok mélyebb, az identitást és az emberi létezést érintő jelentéseit tárja fel.
Közeli analitikus olvasat	A kutató részletesen elemzi a nyelvhasználatot (metaforák, szóválasztás, ismétlések), nem hagyja az idézeteket magyarázat nélkül.
Konvergencia és divergencia	Az elemzés egyensúlyt teremt a közös mintázatok bemutatása és az egyéni, egyedi tapasztalatok tiszteletben tartása között.
Önreflexió	A kutató tudatosítja és átláthatóan kezeli saját előfeltevéseit és szerepét az értelmezési folyamatban.

4. táblázat: Gyakorlati ellenőrző lista a minőségi IPA-kutatáshoz

5. Mikor használjuk és mikor ne használjuk a módszert?

5.1. A kutatási kérdés

A módszer akkor hasznos, ha a kutatót valamely jelenség összetettsége, folyamatjellege vagy újdonsága foglalkoztatja. Az IPA kutatási kérdései nyitottak, feltáróak, és a jelentésre, a megélt tapasztalat lényegére összpontosítanak. Jellemzően a „hogyan", „mi" és „milyen" kérdőszavakkal kezdődnek. A cél a tapasztalatok, nézetek és élmények mély megértése, ezért a módszertan kerüli az ok-okozati összefüggésekre vagy a változók közötti különbségekre irányuló kérdéseket, mint például a „miért" vagy a „hatással van-e". A jó kutatási kérdés teret ad a résztvevőnek, hogy saját szavaival, saját struktúrájában bontsa ki a történetét.

Példa egy jó IPA-kérdésre: „Hogyan élik meg a nők az identitásuk változását az anyává válás során?"
Példa egy kerülendő, nem IPA-jellegű kérdésre: „Miért okoz identitásválságot az anyává válás?"

5.2. A minta

Az IPA-ban a mintaválasztás mindig célzott, nem véletlenszerű. A kutató nem az alapsokaság statisztikai reprezentálására törekszik, hanem olyan személyeket keres, akik lényeges tapasztalatokkal rendelkeznek a vizsgált jelenségről, és gazdag, részletes betekintést nyújthatnak abba. A minta két fő jellemzője:

a) Az IPA a „mélység a szélesség helyett" elvét követi. Míg a korábbi szakirodalom gyakran három-hat fős mintákat javasolt, a frissebb ajánlások szerint egy alapos, publikálható elemzéshez legalább öt-, de inkább hét-tíz fős minta szükséges. Bár egyetlen, rendkívül gazdag és alaposan elemzett esettanulmány is lehet érvényes, a csoportos tapasztalati mintázatok feltárásához több résztvevőre van szükség.
b) A kutató egy viszonylag homogén csoportot választ, vagyis olyan résztvevőket, akik egy adott tapasztalat (például egy betegség, egy életesemény, egy szakmai szerep) szempontjából hasonló helyzetben vannak. Ez a közös háttér azért fontos, mert lehetővé teszi, hogy a kutató az általános, felszínes különbségek helyett a közös tapasztalatokon belüli egyéni eltérésekre és a személyes jelentésadás finom árnyalataira összpontosítson. Az egyneműség tehát nem a változatosságot zárja ki, hanem éppen annak mélyebb vizsgálatát teszi lehetővé egy jól meghatározott kereten belül.

Mikor megfelelő a minta mérete?	Mikor NEM megfelelő a minta mérete?
Jó: öt-tíz fő. Ez teszi lehetővé az egyes esetek mélyreható, részletes elemzését, ugyanakkor elegendő alapot biztosít az esetek közötti mintázatok (konvergenciák és divergenciák) feltárásához.	Túl kicsi: kevesebb mint négy-öt fő. Bár egyetlen, rendkívül gazdag esettanulmány is lehet értékes, kis mintáknál az elemzés egyéni portrék sorozata marad, és nehéz az eseteken átívelő témákat hitelesen azonosítani.
Lehetséges: egy-négy fő (esettanulmányként). Ha a cél egy vagy néhány egyedi, különlegesen informatív eset rendkívül mély bemutatása, akkor elfogadható.	Túl nagy: több mint 15–20 fő. Az IPA elemzési folyamata annyira intenzív és időigényes, hogy egy nagy minta feldolgozása elkerülhetetlenül felszínességhez vezet. A kutató belefulladhat az adatokba, és elvész a módszer lényege, a részletgazdag elemzés.

5. táblázat: Útmutató az IPA-mintaméret meghatározásához

5.3. A kutatás jellege

Az IPA a legjobb választás, ha a kutatást a következők jellemzik:

a) A kutatás témája jelentőségteljes, megélt emberi tapasztalat. Az IPA azokra az eseményekre és folyamatokra összpontosít, amelyek az egyén számára különös jelentőséggel bírnak, gyakran sorsfordítóak vagy identitásalakítók. Példák: krónikus betegséggel való együttélés, szülővé válás, egy trauma feldolgozása, hivatásváltás, új technológia (például mesterséges intelligencia) bevezetése a szakmai életbe, gyászfolyamat.
b) A kutatási kérdés a „hogyan?" és a „milyen?" kérdése. A kutatási kérdés a megértésre összpontosít. Ahogy korábban is tárgyaltuk, a módszer a „hogyan" és „milyen" típusú kérdésekre keresi a választ, nem pedig a „miért" vagy az „okoz-e" kérdésekre.
c) A várt eredmény mély, részletgazdag megértése. A kutató nem általánosítható törvényszerűségeket, hanem mély, összefüggésrendszerbe ágyazott, sűrű leírást keres. A cél egy jelenség feltárása és megértése az azt átélők szemszögéből. Az IPA kiválóan alkalmas új, feltáratlan területek feltérképezésére, ahol még nincsenek szilárd elméletek.

Az IPA NEM jó választás, ha a kutatást a következők jellemzik:

a) A kutatás témája viselkedés, gyakoriság vagy elméletek tesztelése. Ha a kutató arra kíváncsi, hogy hányan tesznek valamit, milyen gyakran fordul elő egy jelenség, vagy egy beavatkozás hatékonyabb-e egy másiknál, az IPA nem megfelelő. Alkalmatlan témák: fogyasztói preferenciák mérése, oktatási program hatékonyságának tesztelése, szavazói viselkedés előrejelzése.
b) A kutatás kérdése a „miért?" és az „okoz-e?". Az IPA nem ok-okozati kapcsolatok feltárására vagy csoportok közötti statisztikai különbségek kimutatására szolgál. Kerülendő kérdések: „Miért vezet a kiégés pályaelhagyáshoz?" „A férfiak vagy a nők élik meg nehezebben a válást?"
c) A várt eredmény általánosítható, vagy gyors, egyértelmű válaszokat igényel a kutatás. Ha a kutatás célja, hogy a teljes populációra érvényes, statisztikailag általánosítható következtetéseket vonjon le, vagy egyértelmű, a szakpolitikai döntéshozatalban azonnal használható adatokat szolgáltasson, akkor az IPA nem megfelelő választás.

d) Nagy mennyiségű, strukturált adat áll rendelkezésre. Az IPA nem alkalmas nagy korpuszok, kérdőíves adatok vagy közösségimédia-posztok ezreinek elemzésére. Ezekhez más módszerek (például tartalomelemzés, természetesnyelv-feldolgozás) valók.

Előnyök	Hátrányok
Rendkívül gazdag, mély adatok. Az IPA betekintést nyújt a jelenségek emberi, szubjektív oldalába, amelyet más módszerek nem tudnak megragadni.	Idő- és munkaigényes. Az adatgyűjtés, az átírás és különösen az elemzés rendkívül sok időt és energiát emészt fel.
A résztvevő hangjának felerősítése. A módszer közvetlenül az érintettek tapasztalataira épít, és teret ad a személyes értelmezéseknek.	Nem általánosítható. Az eredmények a vizsgált kis, homogén mintára vonatkoznak, statisztikailag nem vetíthetők ki a tágabb populációra.
Új elméletek és felismerések forrása: feltáratlan területeken képes olyan új belátásokat generálni, amelyek későbbi kvantitatív kutatások alapjául szolgálhatnak.	Erősen interpretatív. A kutató értelmezői szerepe központi, ami felveti a kutatói elfogultság kockázatát. Ezt reflexivitással kell kezelni.
Az interjúk és az elemzés során is képes alkalmazkodni a felmerülő, váratlan témákhoz.	Érzelmileg megterhelő lehet. A kutatónak és a résztvevőnek is, különösen az érzékeny, traumatikus témák esetében.

6. táblázat: Az IPA előnyei és hátrányai

6. A módszer változatai és haladó alkalmazásai

A módszertan fejlődésének egyik iránya a tapasztalatok időbeli változásának megragadása. Míg a legtöbb IPA-kutatás egyetlen időpontban, pillanatfelvételként vizsgálja a megélt élményeket, a longitudinális IPA (LIPA) arra tesz kísérletet, hogy egy folyamatot dokumentáljon. Például a kutatók nem elégednek meg egyetlen interjúval, hanem ugyanazokkal a válaszadókkal három időpontban is beszélgetnek. Így például anyákkal a terhesség alatt, a szülést követő korai

hetekben, majd hónapokkal később. A longitudinális felépítés lehetővé teszi, hogy a kutatók ne csupán azt értsék meg, milyen a perinatális szorongás, hanem azt is, hogyan alakul az idő múlásával. A résztvevői (participatív) IPA pedig a résztvevőket nem pusztán adatforrásként, hanem a kutatási folyamat aktív közreműködőiként kezeli, vagyis bevonja őket az értelmezés alakításába és validálásába. Ez különösen fontos lehet akkor, ha a kutató más kultúrában vagy sérülékeny közösségben dolgozik, ahol a külső szemlélő pozíciója önmagában is problematikus. A longitudinális és participatív változatok azt mutatják, hogy az IPA képes megbirkózni az időben változó, bonyolult folyamatok és a mélyen, kulturálisan beágyazott tapasztalatok vizsgálatának kihívásaival is.

A módszer egy másik fontos kiterjesztése a multiperspektivikus IPA (multiperspectival IPA). A hagyományos IPA-kutatás egy homogén mintán belül, egyetlen nézőpontból vizsgál egy jelenséget – például krónikus betegek saját élményeit a betegséggel kapcsolatban. A multiperspektivikus felépítés ezzel szemben ugyanazt a jelenséget vagy élethelyzetet több, egymással kapcsolatban álló résztvevő szemszögéből közelíti meg.
Tipikus kutatási elrendezések lehetnek:

a) diádikus vizsgálatok, amikor egy pár vagy egy közeli kapcsolat két tagja kerül be a mintába (például daganatos beteg és a főápoló családtag vagy pszichoterápiás páciens és terapeuta);
b) családi vagy rendszerszintű vizsgálatok, amikor egy család vagy kisebb közösség több tagjának tapasztalatait hasonlítják össze (például egy krónikus gyermekbetegség kontextusában a szülő, a testvér és a gyermek perspektíváját),
c) szolgáltató-igénybevevő vizsgálatok, amikor ugyanannak a szolgáltatásnak vagy beavatkozásnak a megélését kutatják mindkét oldalról (például betegek és kezelőorvosok, diákok és pedagógusok, kliensek és szociális munkások nézőpontjából).

A módszer fontos jellemzője, hogy az idiografikus elkötelezettség ebben a változatban is érvényben marad. Minden résztvevő adatait először külön-külön, a maga mélységében kell elemezni. Csak ezt követően kerülhet sor az esetek közötti és a perspektívák közötti összehasonlításra. A multiperspektivikus IPA értékes lehet olyan jelenségek kutatásában, amelyek természetüknél fogva kapcsolati, rendszerszintű vagy intézményi beágyazottságúak, és egyetlen szereplő nézőpontjából nem érthetők meg teljesen. Emellett kifejezetten hasznosnak bizonyult az alkalmazott kutatás területén, például egészségügyi beavatkozások, szolgáltatásfejlesztési és co-design projektek megalapozásában,

ahol a különböző érintettek együttes perspektívája közvetlenül hozzájárulhat a gyakorlat formálásához.

7. Összefoglalás

Az interpretatív fenomenológiai elemzés a megélt emberi tapasztalat mély megértésére és értelmezésére törekszik. A módszer ereje abban rejlik, hogy a fenomenológiai, hermeneutikai és idiografikus elveket koherens kutatási gyakorlatban egyesíti, és lehetővé teszi a kutató számára, hogy a résztvevők szubjektív valóságához a lehető legközelebb kerüljön. Az IPA központi célja annak feltárása, hogy az egyének hogyan adnak jelentést a számukra fontos, gyakran sorsfordító életeseményeknek. Ennek érdekében a módszertan idiografikus fókuszt alkalmaz, és kis, célzottan kiválasztott, homogén mintákkal dolgozik, hogy a részletgazdag, mély elemzés megvalósulhasson. Az adatgyűjtés legfőbb eszköze a rugalmas keretet biztosító félig strukturált interjú, melynek sikere a gondos előkészítésen és a bizalmi légkör megteremtésén múlik. A módszer fenomenológiai gyökerei abban mutatkoznak meg, hogy a kutató a résztvevő tapasztalatát tekinti elsődlegesnek, és reflexív módon tudatosítja saját előfeltevéseit ahelyett, hogy azokat teljesen kizárná. Ugyanakkor az IPA hermeneutikai jellege elismeri a kettős értelmezési folyamatot, amelyben a kutató a résztvevő értelmezését értelmezi. Az IPA-elemzés aprólékos és iteratív folyamat, amely az egyes esetek alapos, sorról sorra történő vizsgálatától halad a kialakuló témák azonosításán át az esetek közötti átfogóbb mintázatok feltárásáig. A kutatás minőségét és hitelességét olyan követelmények biztosítják, mint a módszertani átláthatóság, az elméleti hűség és a kutatói reflexivitás. A legkiválóbb IPA-munkák ezenfelül meggyőző értelmezést építenek, egzisztenciális mélységet érnek el, a szavak közeli olvasatával dolgoznak, és ügyesen egyensúlyoznak a konvergencia és divergencia bemutatása között. Végül láthattuk, hogy az IPA élő, fejlődő módszertan, amely olyan haladó alkalmazásokat is lehetővé tesz, mint a longitudinális és a participatív megközelítés, bizonyítva a módszer használhatóságát az összetett, időben változó és kulturálisan beágyazott jelenségek esetében is.

8. Önellenőrző kérdések

1. Magyarázza el saját szavaival, mit jelent a „kettős hermeneutika” az IPA szempontjából, és miért központi eleme a módszernek!

Válasz: A kettős hermeneutika azt jelenti, hogy a kutató kétszintű értelmezési folyamatot végez. Egyrészt megpróbálja értelmezni azt, ahogyan a résztvevő a saját élményeit és világát értelmezi. Másrészt a résztvevő értelmet ad a tapasztalatainak, a kutató pedig ennek az értelemadási folyamatnak próbál értelmet adni. Azért központi elem, mert az IPA elismeri, hogy a kutató sosem a nyers tapasztalathoz fér hozzá, hanem mindig a résztvevő elbeszéléséhez, ami már önmagában is egy értelmezés.

2. Fogalmazzon meg egy kutatási kérdést, amely kifejezetten alkalmas IPA-val való vizsgálatra, és egy olyat, amelyik NEM az! Indokolja meg a választását!

Válasz: Alkalmas kérdés: „Hogyan élik meg a pályakezdő tanárok az első tanévüket egy hátrányos helyzetű iskolában?” Ez a kérdés a megélt tapasztalat és a jelentésadás folyamatának mély, feltáró jellegű vizsgálatára irányul. Alkalmatlan kérdés: „Milyen hatással van a tanácsadói program a pályakezdő tanárok pályaelhagyási arányára?” Ez a kérdés ok-okozati összefüggést és mérhető hatást vizsgál, amihez matematikai módszerek (statisztikai elemzés) szükségesek, nem pedig az IPA.

3. Sorolja fel a magas minőségű IPA-kutatás négy ismérvét!

Válasz: A magas minőségű IPA-kutatás négy ismérve: 1. meggyőző, kibontakozó értelmezés felépítése; 2. erőteljes élményszerű és/vagy egzisztenciális elemzés kidolgozása; 3. a résztvevői szavak közeli, analitikus olvasata; 4. a konvergenciára (közös mintázatok) és divergenciára (egyedi tapasztalatok) való odafigyelés.

4. Mi jellemzi az ideális mintát egy IPA-kutatásban? Röviden magyarázza el a három legfőbb jellemzőt!

Válasz: Az ideális minta három legfőbb jellemzője a kis méret, a célzott kiválasztás és a homogenitás. Kis méretű (jellemzően 5-10 fős), hogy lehetővé tegye a mély, esetenkénti elemzést. Célzottan kiválasztott, vagyis a kutató olyan személyeket keres, akik lényeges tapasztalattal rendelkeznek a vizsgált

jelenséggel kapcsolatban. Egynemű, hogy az elemzés a közös tapasztalaton belüli finom árnyalatokra összpontosíthasson, ne pedig a csoportok közötti nagy különbségekre.

5. Miben különbözik a longitudinális IPA (LIPA) a hagyományos, egy időpontban végzett IPA-kutatástól? Milyen többletinformációt nyújthat ez a megközelítés?

Válasz: A hagyományos IPA egyetlen időpontban, pillanatképként vizsgálja a tapasztalatot, míg a longitudinális IPA (LIPA) ugyanazokat a résztvevőkkel több különböző időpontban is készít interjút. A LIPA többletinformációja abban rejlik, hogy képes megragadni a tapasztalatok folyamatát, időbeli változását és fejlődését, így nemcsak azt mutatja be, hogy milyen egy élmény, hanem azt is, hogyan alakul az idő múlásával.

Ajánlott irodalom

Finlay, L. (2021). Thematic Analysis. *European Journal for Qualitative Research in Psychotherapy(11)*, 103–116. https://doi.org/10.24377/EJQRP.article3062

Folliard, K. J., Crozier, K., & Kamble, M. M. W. (2024). A very, very lonely, unmagical time. The lived experience of perinatal anxiety: A longitudinal interpretative phenomenological analysis. *Midwifery*, 136, 104070. https://doi.org/10.1016/j.midw.2024.104070

King, N. (2019). *Interviews in qualitative research*. SAGE.

McCammon, L. C., Gillen, P., McLaughlin, D., & Kernohan, W. G. (2025). An Interpretative Phenomenological Analysis That Seeks to Describe and Understand the Personal Experience of Burnout in General Practitioners. *Qualitative Health Research, 35*(1), 118–131. https://doi.org/10.1177/10497323241260738

Nizza, I. E., Farr, J., & Smith, J. A. (2021). Achieving excellence in interpretative phenomenological analysis (IPA): Four markers of high quality. *Qualitative Research in Psychology, 18*(3), 369–386. https://doi.org/10.1080/14780887.2020.1854404

Silver, C., & Lewins, A. (2014). *Using Software in Qualitative Research: A Step-by-Step Guide*. SAGE. https://doi.org/10.4135/9781473906907

Smith, J. A. (2011). Evaluating the contribution of interpretative phenomenological analysis. *Health Psychology Review, 5*(1), 9–27. https://doi.org/10.1080/17437199.2010.510659

Smith, J. A. (2024). Travelling in time: Using interpretative phenomenological analysis (IPA) to examine temporal aspects of the personal experience of issues in health and well-being. *European Journal of Psychotherapy & Counselling, 26*(3-4), 235–252. https://doi.org/10.1080/13642537.2024.2388663

Smith, J. A., Flowers, P., & Larkin, M. (2022). *Interpretative phenomenological analysis: Theory, method and research*. SAGE.

Smith, J. A., & Nizza, I. E. (2022). *Essentials of interpretative phenomenological analysis*. American Psychological Association.

Taylor, N., Maduesesi, O., Vasiliou, V. S., & Thompson, A. R. (2025). The experience of living with vitiligo in Nigeria: A participatory Interpretative Phenomenological Analysis. *Journal of Health Psychology, 30*(5), 1120–1135. https://doi.org/10.1177/13591053241261684

Tomán, E. (2024). Genetikai örökség – családi test: Egy öröklődő betegség embodiment-tapasztalata. *Alkalmazott Pszichológia, 26*. https://doi.org/10.17627/ALKPSZICH.2024.4.37

Zahavi, D. (2021). Applied phenomenology: why it is safe to ignore the epoché. *Continental Philosophy Review, 54*(2), 259–273. https://doi.org/10.1007/s11007-019-09463-y

III. REFLEXÍV TEMATIKUS ELEMZÉS (RTA)

TURÓS MÁTYÁS

A fejezet a következő kérdéseket válaszolja meg:

a) Mi az a reflexív tematikus elemzés (RTA), és miben különbözik a hagyományosabb, objektivitásra törekvő kvalitatív tartalomelemzéstől?
b) Milyen kutatásfilozófiai alapokon nyugszik az RTA, és miért fontos ez a gyakorlatban?
c) Hogyan értelmezi újra az RTA a kutatói szubjektivitás és a reflexivitás szerepét az elemzési folyamatban?
d) Mi a reflexív tematikus elemzés hat, egymásra épülő, de önmagába visszatérő szakasza?
e) Milyen elméleti döntéseket kell meghoznia a kutatónak az elemzés megkezdése előtt?
f) Milyen minőségi követelmények mentén értékelhető egy RTA-kutatás, és miért veti el a módszer a kódolók közötti megbízhatóság mérését?
g) Milyen kutatási kérdések és célok esetén ideális választás az RTA, és mikor érdemes más hermeneutikai módszert választani?

1. Bevezetés

Ahogyan a kötet bevezető szakaszában láttuk, a „tematikus elemzés" ernyőfogalom, amely több, filozófiájában eltérő irányzatot foglal magában. Jelen fejezet tárgya a reflexív tematikus elemzés (RTA), amely az irányzatok között kifejezetten interpretatív megközelítést képvisel. Ahhoz, hogy megértsük az RTA sajátosságait és a hagyományosabb módszerektől való eltéréseit, először a strukturálatlan adatok elemzésének legalapvetőbb eszközét, a kódolást kell áttekintenünk.

A kódolás analitikus címkézési folyamat, amely során a kutató a nyers adatokat (szöveget, képet, bármilyen nem strukturált adatot) kisebb, értelmes egységekre bontja, és mindegyikhez egy-egy rövid, a jelentés lényegét megragadó címkét, azaz kódot rendel. A kód nem pusztán kulcsszó vagy egyszerű tartalmi összefoglaló, hanem analitikus absztrakció. Olyan címke, amely a kutató első interpretációját tükrözi arról, hogy az adott adatrészlet miért érdekes vagy lényeges a kutatási kérdés szempontjából. A kódok létrehozásának logikája szerint három fő megközelítést különböztethetünk meg.

Az induktív (adatvezérelt) kódolás alulról építkező folyamat, ahol a kutató törekszik arra, hogy előzetes elméleti kategóriáktól mentesen, közvetlenül az adatból építse fel a kódokat, a résztvevők világának megértésére törekedve. Ez a logika jellemzi a konvencionális kvalitatív tartalomelemzést, amelyet akkor alkalmaznak, ha egy jelenségről még kevés tudás áll rendelkezésre. A kutató itt úgy dolgozik, ahogyan egy nyomozó, aki a helyszínen gyűjtött nyomokból építi fel az elméletét.

Ezzel szemben az elméletvezérelt (deduktív) kódolás logikája fentről lefelé halad; a kutató egy már meglévő elméleti keretből származó kategóriákat keres és értelmez az adatokon. Ez az irányított megközelítés akkor ideális, amikor egy elméletet szeretnénk egy új környezetben vizsgálni, hasonlóan egy szerelőhöz, aki egy diagnosztikai ellenőrző lista segítségével keresi a hibát.

Végül a harmadik, az abduktív kódolás[1] párbeszédet folytat az adatok és az elmélet között: a kutató egy meglepő jelenségre figyel fel, majd arra törekszik,

[1] Az abdukció fogalma Charles Sanders Peirce nevéhez fűződik. Míg a dedukció egy meglévő szabályt alkalmaz az egyedi esetre, az indukció pedig az egyedi esetekből próbál általános szabályt alkotni, az abdukció a „meglepetésből" indul ki. A kutató egy olyan adatra bukkan, amely a jelenlegi elméleti kereteivel nem magyarázható, és keresi azt a hipotézist, amelynek igazsága esetén az adott jelenség magától értetődővé válna. A módszertanban ezt gyakran az elméleti érzékenység és az adatok közötti

hogy megtalálja vagy megalkossa azt a legjobb lehetséges magyarázatot, amely értelmet ad a rejtélynek.

6. doboz: Az induktív, deduktív és abduktív kódolási logikák gyakorlati különbségei

A három kódolási logika közötti különbséget a legjobban egyetlen, közös példán keresztül lehet szemléltetni. Vegyünk egy interjúrészletet, amelyben egy egyetemi hallgató a beilleszkedési nehézségeire emlékszik vissza: „Az első évben borzasztóan elveszettnek éreztem magam. Mindenki olyan okosnak tűnt, én meg csak ültem ott, mintha nem is oda való lennék. Aztán a második félévben csatlakoztam egy tanulókörhöz, és minden megváltozott. Rájöttem, hogy mások is ugyanígy éreznek." Induktív kódolás során a kutató az élményből kiindulva hoz létre kódokat, például: „Imposztorszindróma", „Társas támogatás mint fordulópont" vagy „Kollektív tapasztalat felismerése". Deduktív kódolás esetén, ha a kutató például Tinto hallgatói integrációs modelljét használja, amely megkülönbözteti az akadémiai és társas integrációt, a kódjai az elmélet fogalmaihoz igazodnának: „Hiányzó akadémiai integráció" és „Sikeres társas integráció". Abduktív kódolásnál a kutató a rejtélyre összpontosít (miért oldott meg egy akadémiai eszköz egy szociális-érzelmi problémát?), és egy új hipotézist fogalmaz meg a kódban: „Akadémiai tevékenység mint a szociális beilleszkedés motorja".

A hagyományos kvalitatív tartalomelemzés (legyen az induktív, deduktív vagy akár a szavak gyakoriságából kiinduló szummatív) jellemzően a szabálykövető, objektivitásra törekvő szemléletet képviseli. A kutatói szubjektivitást kontrollálandó tényezőnek tekinti, és jól kialakított, megbízható kategóriarendszer létrehozására összpontosít. Ennek érdekében gyakran hangsúlyt fektet a kódolók közötti megbízhatóság mérésére, amely azt számszerűsíti, hogy több független kódoló mennyire ért egyet az adatok kódolásában. Erre a célra több elterjedt mutató is létezik, mint a százalékos egyezés, a véletlenre is korrigáló Cohen-féle vagy Fleiss-féle kappa (κ), valamint a robusztus Krippendorff-féle alfa (α).

oda-vissza mozgásként írják le. (Lásd: Peirce, C. S. [1960]. *Collected Papers of Charles Sanders Peirce*. Harvard University Press.)

7. doboz: A kódolói megbízhatóság mutatói (Percent Agreement, Kappa, Krippendorff-alfa)

A kódolói megbízhatóság (inter-coder reliability, IRR) számszerűsítése azt a célt szolgálja, hogy a kódolási folyamat objektivitását és következetességét biztosítsuk, túllépve egyetlen kutató szubjektív interpretációján. A legegyszerűbb mutató az egyszerű százalékos egyezés (Percent Agreement), amely azonban félrevezetően magas értéket adhat, mivel nem veszi figyelembe a véletlennek betudható egyezéseket. Hasonló korlátokkal rendelkezik a Søren-sen–Dice-együttható is, amely a kódok közös előfordulására összpontosít, de figyelmen kívül hagyja azt, amikor a kódolók abban értenek egyet, hogy egy kód nincs jelen, és szintén nem korrigál a véletlenre. E probléma kiküszöbölésére a társadalomtudományi gyakorlatban olyan mutatók terjedtek el, amelyek kiszűrik a véletlen szerepét. Nominális (kategoriális) adatok és két kódoló esetén a legelterjedtebb a Cohen-kappa (κ), amelynek kettőnél több kódoló esetén használható megfelelője a Fleiss-kappa. Amikor azonban a kódolók nem csupán kategóriákat rendelnek az adathoz, hanem ordinális (például egyetértés mértéke 1–5 skálán) vagy folytonos skálán (például egy viselkedés intenzitása) értékelnek, az osztályon belüli korrelációs együttható (Intraclass Correlation Coefficient, ICC) a megfelelő választás. Az ICC azt méri, hogy a teljes varianciából mekkora hányadot tesz ki a mért alanyok közötti tényleges különbség, szemben a kódolók közötti inkonzisztenciával. A legrugalmasabb és módszertanilag legfejlettebb mutató a Krippendorff-alfa (α), amely tetszőleges számú kódolóval, bármilyen mérési szintű adattal (nominális, ordinális, intervallum-, arányskála) és hiányzó adatokkal is megbízhatóan működik. Míg tehát a Dice-együttható gyors, de felületes becslést ad, a magas minőségű kutatásokban a véletlenre korrigált mutatók, mint a Cohen-kappa és különösen a Krippendorff-alfa tekinthetők szabványnak, mivel ezek adják a legpontosabb képet a kódolás tényleges megbízhatóságáról.

Ahogy azonban a későbbiekben látni fogjuk, az RTA éppen ezeket a kvantitatív megbízhatósági mutatókat utasítja el mint a kvalitatív elemzés minőségének hamis mércéjét – és helyettük egészen más minőségkritériumokat állít a középpontba.

A cél tehát a valóságot a lehető legpontosabban leképező rendszer létrehozása, ahol a kódokból felépülő kategóriák és témák szisztematikus rendszerezési és adatcsökkentési célokat szolgálnak. Ezzel a strukturált, a valóság leképezésére törekvő szemlélettel szemben a reflexív tematikus elemzés alapvetően eltérő filozófiát kínál. Az RTA szerint a kutató nem tárgyilagos, semleges, pártatlan eszköz, aki felfedi a már létező jelentést az adatokban. Ehelyett a kutató az elemzési folyamat tevékeny és szubjektív résztvevője, aki a témákat az adatokkal való elmélyült, alkotó és reflexív munka során hozza létre. A jelentés az adat és az

azt értelmező kutató összjátékában létesül, így a kutatói szubjektivitás itt nem hibaforrás, hanem az elemzés elengedhetetlen és értékes erőforrása. Az alábbi táblázat a két megközelítés legfontosabb különbségeit foglalja össze.

Szempont	Kvalitatív tartalomelemzés	Reflexív tematikus elemzés
Fő cél	Szövegek tartalmának rendszerszerű leírása, gyakran kategóriák gyakoriságának mérése.	A résztvevők tapasztalatainak mély, kontextuális és interpretatív megértése. A témák létrehozása.
Kutatói szerep	Tárgyilagosabb, távolságtartóbb. A kutató a kategóriák megbízható alkalmazója.	Központi, alkotó és önértékelő. A kutató a jelentés tevékeny létrehozója, aki elismeri és tudatosan használja saját szubjektivitását.
Fókusz	Jellemzően a szöveg felszíni, explicit tartalmára, a kifejezett jelentések rendszerezésére koncentrál.	A nyílt és a rejtett, mögöttes jelentésekre is összpontosít, a jelentésalkotás folyamatát vizsgálja.
Struktúra	Erősen strukturált, gyakran előre meghatározott kódolási sémával és lineáris elemzési folyamattal dolgozik.	Rugalmas és rekurzív. Bár a folyamatnak vannak jól meghatározott szakaszai, azok nem feltétlenül lineárisak.
Kulcskérdés	Mit és milyen gyakran mondanak?	Hogyan jön létre a jelentés ebben a viszonyrendszerben, és én hogyan vagyok ennek a folyamatnak a része?

7. táblázat: A kvalitatív tartalomelemzés és a reflexív tematikus elemzés összehasonlítása

2. Elméleti alapok

A reflexív tematikus elemzés a kutató szerepét és a tudás természetét tehát – az objektivista metodológiákhoz képest – radikálisan újraértelmezi. Az RTA a kutatást nem tényfeltáró folyamatnak, hanem mélyen interpretatív és kontextuális tevékenységnek tekinti. Főként ez különbözteti meg a hagyományosabb, kvalitatív tartalomelemzéshez közel álló módszerektől. Ezért emeli ki Virginia Braun és Victoria Clarke a – Kidder és Fine nyomán bevezetett – „Big Q" és „small q" megkülönböztetést. E felosztás nem a kutatás jelentőségére utal, hanem a kvalitatív jelző mögött meghúzódó alapvető kutatásfilozófiai különbségre. A szerzők szerint míg a „small q" a kvalitatív (strukturálatlan) adatok használatát jelöli gyakran pozitivista kereten belül, addig a „Big Q" egy teljes kutatási filozófiát, egy világnézetet takar, amely a valóságot társadalmilag konstruáltnak és szubjektívnek tekinti. Az alábbi táblázat részletesen bemutatja a két szemlélet közötti alapvető különbségeket.

Szempont	small q (kvalitatív mint adat)	Big Q (kvalitatív mint paradigma)
Kutatásfilozófia	Jellemzően (poszt) pozitivista, realista. Feltételezi, hogy létezik objektív, megismerhető valóság, amelyet a kutatásnak a lehető legpontosabban le kell képeznie.	Konstruktivista, interpretativista, kritikai. A valóságot társadalmilag konstruáltnak, szubjektívnek és sokrétűnek tekinti. A tudás nem felfedezett, hanem a kutató és a résztvevők kölcsönhatásában jön létre.
A kutatás célja	Leírás, kategorizálás, megbízható kódolás. Gyakran a cél a strukturálatlan adatok csökkentése, rendszerezése, esetenként a mintázatok számszerűsítése.	Mély megértés, interpretáció, a jelentésalkotás folyamatainak feltárása. A cél a jelenségek gazdag, viszonyrendszerbe ágyazott, árnyalt megértése, nem pedig egyetlen „igazság" megtalálása.

A kutató szerepe	Tárgyilagos, távolságtartó kódoló. A kutató semleges eszköz, akinek feladata a kódolási szabályok következetes alkalmazása és a saját szubjektivitásának minimalizálása.	Alkotó, szubjektív és reflexív. A kutató a legfőbb elemzési eszköz. A személyes tapasztalatai, elméleti tudása és értékei az elemzés szerves és értékes részét képezik.
Minőségi követelmények	Megbízhatóság (reliability), objektivitás, kódolói egyezés (inter-coder reliability). A minőség záloga, ha különböző kódolók ugyanazt az eredményt kapják.	Hitelesség (credibility), átláthatóság (transparency), reflexivitás. A minőség záloga a jól argumentált, az adatokkal alaposan alátámasztott, a kutatói döntéseket átláthatóan bemutató, mély és árnyalt értelmezés.
Tipikus módszerek	Strukturált kvalitatív tartalomelemzés, nyílt végű kérdőíves kérdések egyszerű kódolása.	Reflexív tematikus elemzés, interpretatív fenomenológiai elemzés (IPA), megalapozott elmélet (Grounded Theory), diskurzuselemzés.

8. táblázat: A kvalitatív kutatás két paradigmája, a „small q"- és a „Big Q"-megközelítés összehasonlítása

Braun és Clarke szerint, ha a kutató az RTA rugalmas folyamatát pozitivista minőségi követelményekkel – például a kódolói egyezés számításával – próbálja ötvözni, az eredmény metodológiailag zűrzavaros lesz. Ezt a jelenséget a szerzők a „pozitivizmus beszivárgásának" nevezik. Az ilyen felemás, hibrid megoldásokkal szemben az RTA egyértelműen a „Big Q"-paradigmába tartozik, és ennek alapvető következményei vannak a kutatás teljes folyamatára.

Az RTA filozófiájának sarokköve a kutatói szubjektivitás újragondolása. Míg a pozitivista hagyományban a kutató személyes tapasztalatai és értékei kiküszöbölendő torzító tényezők, addig az RTA-ban ezek az elemzés elsődleges és megkerülhetetlen erőforrásai. A kutató nem megfigyelő, aki az adatokból hozza a felszínre az ott rejlő témákat, hanem alkotó, aki saját elméleti és személyes

hátterét felhasználva, az adatokkal folytatott párbeszéd során érti meg a kutatott jelenséget. Erre utal a híres mottó, miszerint a „témák nem bukkannak fel".

A reflexivitás során a kutató kritikusan szemléli saját helyzetét, döntéseit és az elemzési folyamatra gyakorolt hatását. A reflexivitás legfontosabb eszköze a kutatói napló, amelyben a kutató saját előfeltevéseit, az elemzés, a kódolás és a témaalkotás során hozott döntéseit és azok indokait rögzíti. A dokumentáció biztosítja az elemzés átláthatóságát és ellenőrizhetőségét; ez teszi lehetővé az olvasó számára, hogy ne csak a végeredményt, hanem az odáig vezető gondolati utat is megértse.

Az RTA újradefiniálja a téma fogalmát is. A téma nem „tárolóedény", hanem összetett, analitikus történetet elmesélő központi szervezőelv. Ezzel összefügg, hogy az elemzés folyamata szerves és rekurzív. A kutató újra meg újra visszatér a korábbi elemzési szakaszokhoz, és folyamatosan alakítja a témát. Összefoglalva, a minőségbiztosítás nem a kódolói egyezésen, hanem a gondos, átlátható és jól kialakított elemzésen nyugszik. Az RTA szándékosan elveti a kódolók közötti megbízhatóság mérését, mivel a cél nem az egyetértés, hanem a kutató egyedi nézőpontjának átlátható érvényesülése.

Az RTA tehát szisztematikus és alapos módszer, de a kódolást nem szabályozott tevékenységként, hanem alkotó, intellektuális és mélyen interpretatív folyamatként fogja fel. A kódolás az RTA esetében is nélkülözhetetlen az elmélyüléshez, az adatok redukciójához és a témák megalapozásához, a folyamat azonban nyitott. Nincs előre meghatározott kódkönyv, a kódok az adatokkal való összjáték során születnek és fejlődnek. A kódok továbbá interpretatívak, nem pedig leírók, már egyetlen kód is analitikus lépésnek, apró értelmezési mozzanatnak számít. Például a „családi háttér" helyett sokkal erősebb kód lehet az, hogy „a származás mint a jelenlegi identitás magyarázata".

3. A kutatás folyamata

Az RTA szerint a kutató az elemzésbe mindig magával viszi lételméleti és a megismerés természetével kapcsolatos előfeltevéseit. Ezért az elemzés megkezdése előtt a kutatónak döntést kell hoznia a kutatás elméleti alapjairól. Saját feltevéseit négy szempont mentén kell mérlegelnie.

3.1. A kutatás tervezése és az elméleti pozicionálás

Először az esszencialista és a konstruktivista episztemológia között kell választani. A központi kérdés az, hogy a nyelv tükrözi-e a valóságot (esszencialista nézet), vagy a nyelv maga alkotja meg azt (konstruktivista nézet). Ez azért lényeges, mert míg az esszencialista kutató az interjúalany „teljesen kiégtem" kijelentését a valós, belső állapot leírásaként értelmezi, addig a konstruktivista kutató azt vizsgálja, hogyan használja a résztvevő a kiégés kulturálisan rendelkezésre álló fogalmát a saját valóságának megalkotására.

Az esszencialista szemlélet szerint a nyelv olyan, mint egy ablak, amelyen keresztül betekinthetünk a résztvevő valódi, belső világába. A kutató ebben az esetben azt feltételezi, hogy a szavak a tapasztalatok, érzések és a belső élmények hű leképezései, és a cél a jelentés pontos megragadása. A konstruktivista megközelítés szerint a nyelv azonban nem ablak, hanem eszköz, amellyel a valóságot építjük. A kutató nem azt kérdezi, hogy „mit érez valójában a résztvevő?", hanem azt, hogy „hogyan beszél a kiégésről, és ez a beszédmód milyen jelentéseket és valóságokat hoz létre?". További vizsgálódási kérdések lehetnek a következők: „Hogyan használja a résztvevő a kiégés fogalmát?" „Mire szolgál ez az értelmezés?" „Lehet, hogy a kiégés mint diagnosztikus címke segít neki legitimálni a tapasztalatát, felmentést adni a teljesítménykényszer alól, vagy éppen orvosi segítséget kérni?" Ebben az esetben tehát az elemzés a kiégésdiskurzus működésére és társadalmi következményeire összpontosít.

Másodszor az élményközpontú és a kritikai orientáció között kell választani. Az élményközpontú megközelítés a résztvevők szubjektív tapasztalatainak minél hívebb megragadására törekszik. Ezzel szemben a kritikai megközelítés egy külső elméleti keretből (például feminizmus) elemzi a tapasztalatokat alakító társadalmi-hatalmi struktúrákat, a feminizmus példájánál maradva például azt, hogyan alakítják a női vezetők tapasztalatait a patriarchális normák. Az élményközpontú elemzés ebben az esetben arra összpontosítana, hogy milyennek élik meg a nők a vezetői létet, milyen kihívásokkal szembesülnek, milyen örömöket találnak benne, és hogyan egyeztetik össze a munkát a magánélettel. A kritikai elemzés ezzel szemben arra kérdezne rá, hogyan formálják a patriarchális normák és a hatalmi struktúrák a női vezetők élményeit, milyen rejtett elvárásoknak kell megfelelniük, hogyan reprodukálják vagy kérdőjelezik meg ezeket a normákat, és milyen szerkezeti korlátokkal szembesülnek a férfi vezetőkhöz képest.

Harmadszor az induktív és a deduktív elemzés logikája között kell választani. Bár az RTA használja ezeket a kutatásmódszertanból ismert kifejezéseket,

azokat a saját, „Big Q"-paradigmáján belül egyedi módon értelmezi. A deduktív logika alkalmazása az RTA-ban nem jelent visszatérést a pozitivista, szabálykövető módszerekhez. A cél nem egy elmélet tesztelése előre meghatározott kódkönyvvel, hanem az elmélet kreatív használata analitikus lencseként, amely mélyebb értelmezést tesz lehetővé. A kutató reflexív és szubjektív szerepe mindkét logika esetében központi és megkérdőjelezhetetlen marad.

Induktív (adatvezérelt) megközelítés esetén például a kutató a fiatalok közösségimédia-használati szokásait vizsgálja. Nincs konkrét elmélete, egyszerűen csak arra kíváncsi, mit mondanak a résztvevők. Az elemzés során olyan témákat dolgozhat ki, mint az „online önbemutatás", a „csoporthoz tartozás élménye" vagy a „digitális szorongás" – mindezeket a kutató közvetlenül a résztvevők beszámolóihoz kapcsolva alakítja ki. Deduktív (elméletvezérelt) megközelítés esetén a kutató sajátos elméleti keretből (például Bourdieu tőkeelmélete, Foucault hatalomelmélete) indul ki, és az elemzés során azt vizsgálja, hogyan jelennek meg ezek az elméleti fogalmak az adatokban. Az elmélet itt olyan fogalomkészletként működik, amely irányítja a kutató figyelmét. Például a kutató ugyanazt a témát (közösségi média) vizsgálja, de Bourdieu tőkeelméletén keresztül. Az elemzés során kifejezetten arra keresi a választ, hogy a fiatalok hogyan használják a közösségi médiát a szociális, kulturális és szimbolikus tőkéjük felhalmozására és bemutatására. A kódok és témák már eleve az elmélet fogalmi készletére épülnek, például „Szociális tőke építése like-okkal" vagy „Kulturális tőke demonstrálása utazási fotókkal".

Végül a szemantikus és a látens fókusz között kell választani. A szemantikus elemzés a szöveg felszíni, kimondott jelentésére összpontosít (például a klímaváltozással kapcsolatos aggodalmak megfogalmazására), míg a látens elemzés a sorok között megbúvó, rejtett jelentéseket, ideológiákat és feltételezéseket keresi (például a természet mint kizsákmányolható erőforrás elgondolását). A klímaváltozással kapcsolatos attitűdöket vizsgáló interjúkban a szemantikus elemzés során olyan témák alakulhatnak ki, mint az „Aggodalom a jövő miatt", „A kormányzati intézkedésekkel való elégedetlenség" vagy a „Személyes felelősségvállalás". Látens fókusszal ugyanazokat az interjúkat elemezve a kutató olyan témákat dolgozhatna ki, mint „A természet mint közömbös, kizsákmányolható erőforrás" (ha a résztvevők sokat beszélnek a gazdasági növekedésről) vagy „Apokaliptikus jövőkép mint a cselekvés akadálya" (ha a reménytelenség és a tehetetlenség értelmezései dominálnak). Ezek a témák nem hangzanak el kimondva, de a kutató értelmezése szerint e mögöttes gondolati keretek szervezik a résztvevők beszédét.

3.2. Az adatelemzés

Miután a kutató tisztázta elméleti pozícióját és összegyűjtötte az adatokat, következhet a hat, egymásra épülő, de gyakran rekurzív szakaszból álló elemzési folyamat.

Első szakasz: ismerkedés az adatokkal
Az első szakasz az adatokkal való ismerkedés, amely a teljes adathalmazban való mély elmerülést jelenti az interjúátiratok többszöri elolvasásával, a hangfelvételek meghallgatásával és kezdeti jegyzetek készítésével. A kutató az olvasás során kezdeti, reflexív jegyzeteket készít, amelyek már előrevetítik a lehetséges mintázatokat. Ez nem csupán szöveges, hanem képi, alkotó formát is ölthet, például vázlatok, rajzok vagy fogalmi firkák formájában.

Második szakasz: kezdeti kódok létrehozása
A második szakaszban a kutató kezdeti kódokat hoz létre. Végighalad a teljes adathalmazon, és minden lényeges szövegrészlethez egy-egy interpretatív kódot rendel. Egyetlen szövegrészlethez több, akár egymást kiegészítő kód is tartozhat. A kutató az interjúrészlet mellett a margón vagy kvalitatív elemzőszoftverben rögzíti a kódokat.

Harmadik szakasz: a kezdeti témák kialakítása
A harmadik szakaszban, a kezdeti témák kialakítása során a fókusz a kódokról a témákra helyeződik át. A kutató a kódok csoportosításával és képi eszközök, például tematikus térképek segítségével alakítja ki a lehetséges témákat.

Negyedik szakasz: a témák felülvizsgálata
A negyedik szakaszban a kutató felülvizsgálja a témákat, és ellenőrzi, hogy azok mennyire illeszkednek egyrészt a kódolt adatrészletekhez, másrészt a teljes adathalmazhoz. Ez az elemzés legnehezebb, de leginkább alkotó része; ebben a szakaszban a témák sokat változhatnak. Egyesek összeolvadhatnak, mások szétválhatnak, újak pedig születhetnek.

Ötödik szakasz: a témák meghatározása és elnevezése
Az ötödik szakaszban a kutató minden végleges témához definíciót és analitikus nevet rendel. E szakasz a negyedik, felülvizsgálati szakaszban véglegesített tematikus térkép finomhangolásáról és részletes kidolgozásáról szól. Több mint egyszerű címkézés; inkább olyan analitikai lépés, amely során a kutató minden egyes témához tiszta, lényegre törő értelmezést rendel. A cél az, hogy

minden téma önállóan is megálló, koherens történetet meséljen el az adatok egy-egy fontos aspektusáról. A munka két elemből áll.

a) A kutatónak minden témához rövid, de velős elemzői definíciót kell írnia, egy-két bekezdésben. A definíciók nem csupán összefoglalják a téma tartalmát, hanem több kérdésre is választ adnak. Mi a téma központi gondolata? (Mi az az egyedi történet, amelyet ez a téma elmesél?) Pontosan miről szól? (Milyen adatrészletek, kódok tartoznak ide, és mi a közös bennük?) Hogyan járul hozzá a kutatási kérdés megválaszolásához? (Milyen szempontból világítja meg a nagyobb képet?) A definíció lényegében a végső jelentés adott fejezetének a magja lesz. Segít a kutatónak rendezni és élesíteni a gondolatait, mielőtt elkezdené a szövegírás hatodik szakaszát.
b) A definíció megalkotása után a kutató végleges, analitikus nevet ad a témának. A jó témanév az RTA-ban nem csupán leíró címke (például „Családi kapcsolatok"), hanem olyan rövid, figyelemfelkeltő és informatív kifejezés, amely már önmagában is hordozza a téma lényegét és a kutató interpretációját. Egy kezdeti, leíró név lehet például a „Társas támogatás", amely a végleges, analitikus változatban így szólhat: „Tudtam, hogy nem vagyok egyedül – a sorstársi közösség mint a megküzdés kulcsa". Látható, hogy a második, analitikus név sokkal gazdagabb. Nemcsak megnevezi a jelenséget (társas támogatás), hanem megmutatja annak tágabb jelentését (sorstársi közösség), a résztvevők számára betöltött funkcióját (megküzdés), és egy idézettel azonnal érzelmileg is bevonja az olvasót. Összefoglalva, az ötödik szakasz az a híd, amely az elemzés nyers, feltáró szakaszát (kódolás, csoportosítás) összeköti a végső, csiszolt, értelmező kutatási jelentéssel. Itt nyeri el minden téma a végleges formáját, mélységét és analitikus erejét.

Hatodik szakasz: a jelentés megírása

Végül a hatodik szakasz a jelentés megírása, amely az elemzési folyamat olyan befejező mozzanata, amelyben a kutató meggyőző, adatokkal alátámasztott analitikus elbeszélést hoz létre. A kutató feladata, hogy olyan meggyőző értelmezést alkosson, amelyben az idézetek nemcsak példázzák az állításokat, hanem az analitikus érvelés szerves részei, vagyis az elemzés az idézet nyelvi és tartalmi finomságait is boncolja.

8. doboz: A szoftveres környezet előnyei a reflexív elemzésben és a témák fejlesztésében

A reflexív tematikus elemzés (RTA) egyszerre szerves és szigorúan módszeres. A kvalitatív adatelemző (QDA) szoftverek éppen ezt a kettősséget támogatják. Struktúrát adnak a szisztematikus munkához, és teret engednek a kutatói krea-tivitásnak és reflexivitásnak. Az RTA esetében különösen fontos hangsúlyozni, hogy a szoftver nem a kódolás mechanizálására vagy a kvantitatív megbízhatóság biztosítására szolgál – hiszen éppen ezt utasítja el a módszer –, hanem a kutató reflexív, alkotó munkájának támogatására. Az RTA-ban a kutató az elsődleges elemzési eszköz, a szoftver pedig a tevékeny, alkotó munka segítője. A szoftver nem megtalálja a témákat, hanem a kutató számára teszi átláthatóvá és kezelhetővé a folyamatot.

A szoftveres környezet emellett ideális a témák rekurzív fejlesztéséhez. Segíti a teljes adathalmaz következetes kódolását, majd a kódok rugalmas csoportosítását kezdeti témákká. A kutató a szoftverben könnyen felülvizsgálhatja, összevonhatja vagy újradefiniálhatja a témákat, és gyakran képi eszközökkel, például fogalmi térképekkel modellezheti a kialakuló tematikus szerkezetet.

Az RTA központi eleme a kutatói reflexivitás, amelynek a memóírás vagy napló funkció biztosít digitális teret. A reflexív napló rögzíti az elméleti pozicionálással kapcsolatos döntéseket, az elemzés során felmerülő gondolatokat és azokat az interpretációs lépéseket, amelyek a témák végső formájához vezettek. A dokumentáció teszi a kutatói szubjektivitást nyomon követhető analitikai erőforrássá. A kutatási nyomvonal (audit trail) a minőségi RTA-kutatás feltétele, hiszen a kódok és a témák fejlődése, valamint a hozzájuk fűzött memók mind a kutatói döntések átláthatóságát szolgálják. A kutatónak ezért olyan szoftveres megoldást érdemes választania, amely a kódolási és keresési funkciók mellett hangsúlyt fektet a memóírásra és a képi modellezési lehetőségekre.

4. A minőség biztosítása

A magas minőségű RTA-kutatás jellemzői a következők:

a) Szisztematikus elemzés. Bár az RTA rugalmas és szerves, ez nem jelenti, hogy zavaros vagy fegyelmezetlen lenne. A minőség alapja a hatfázisú folyamat, ami biztosítja, hogy a kutató a teljes adathalmazzal elmélyülten foglalkozzon, és ne csak a legérdekesebb, leginkább kiugró részekre összpontosítson.
b) Mély interpretáció a leírás helyett. A minőségi RTA túllép az adatok összegzésén vagy leírásán. Míg a 2024-ben Mitchell Nicmanis által leírt reflexív tartalomelemzés (RCA) elsősorban a manifeszt tartalom leírására és redukálására összpontosít, addig az RTA a rejtett jelentések feltárására törekszik. A minőségi elemzés nemcsak azt mutatja be, mit mondanak a résztvevők, hanem azt is, ahogyan ez a jelentés létrejön, és hogy az milyen tágabb társadalmi jelentéseket hordoz.
c) Reflexivitás mint analitikai erőforrás. A reflexivitás a minőségbiztosítás központi eleme. Ez az a folyamatos és szándékos gyakorlat, amely során a kutató kritikusan tekint saját szerepére, előfeltevéseire, döntéseire és az elemzési folyamatra gyakorolt hatására. A cél nem a kutatói szubjektivitás kiküszöbölése, hanem annak megértése, hogy a kutatói szubjektivitás hogyan járul hozzá az egyedi és értékes elemzési meglátásokhoz. Ennek legfontosabb eszköze a kutatói napló (reflexive journal) vezetése, amelyben a kutató rögzíti az elemzés során felmerülő kétségeit, döntéseit és interpretációs lépéseit.
d) Átláthatóság és ellenőrizhetőség. A minőségi kutatás lehetővé teszi az olvasó számára, hogy ne csak a végeredményt, hanem az odáig vezető gondolati utat is megértse. Az átláthatóság azt jelenti, hogy a kutató részletesen és őszintén bemutatja az elemzés folyamatát, beleértve a kódolási és témaalkotási döntéseket. A kutatási nyomvonal magában foglalja a kódok iteratív fejlődésének feljegyzését, a tematikus térképek vázlatait és a kutatói napló lényeges bejegyzéseit.

Egy RTA-munka értékeléséhez a következő ellenőrzési lista javasolható.

Szakasz	Minőségi követelmények
Adat-előkészítés	Az adatok átírása megfelelő részletességgel történt, és az átiratokat ellenőrizték a hangfelvételhez képest.
Kódolás és témafejlesztés	Minden adatelem (például interjú) alapos és ismételt figyelmet kapott a kódolási folyamat során.
	A kódolás alapos és inkluzív volt; a témákat nem csupán néhány kiragadott, élénk példa alapján alakították ki.
	Az összes fontos adatrészletet összegyűjtötték minden egyes témához.
	A jelölt témákat összevetették a kódolt adatokkal és az eredeti adathalmazzal is.
	A témák belsőleg koherensek, konzisztensek és egymástól megkülönböztethetők; minden téma jól meghatározott központi szervezőelv köré épül.
Jelentés és interpretáció	Az adatokat elemezték – interpretálták, értelmezték –, nem csupán összefoglalták vagy átfogalmazták.
	Az elemzés és az adatok illeszkednek egymáshoz; az idézetek alátámasztják az analitikus állításokat.
	Az elemzés egy meggyőző és jól szervezett történetet mesél el az adatokról, és válaszol a kutatási kérdésre.
	Egyensúly van az analitikus értelmezés és az azt alátámasztó idézetek között.
Átfogó szempontok	Elegendő időt szántak az elemzés minden szakaszának elvégzésére anélkül, hogy siettették volna a folyamatot.
	A választott tematikus elemzési megközelítést és annak részleteit (beleértve az elméleti pozíciókat) egyértelműen kifejtették.
	Jó illeszkedés van a megfogalmazott célok és a ténylegesen elvégzett elemzés között.
	A jelentésben használt nyelvezet és fogalmak összhangban vannak a kutatás lételméleti és megismerésfilozófiai pozícióival.
	A kutató cselekvő szereplő a kutatási folyamatban; a témák nem csupán „felbukkannak”.

9. táblázat: Az RTA minőségi szempontjai

5. Mikor használjuk és mikor ne használjuk a módszert?

5.1. A kutatási kérdés

A reflexív tematikus elemzés arra a központi kérdésre keresi a választ, hogy a résztvevők hogyan alkotják meg egy adott jelenség jelentését, és milyen közös mintázatok, témák rajzolódnak ki az adathalmaz egészében. A módszer akkor a leghasznosabb, ha a kutatót a megosztott tapasztalatok, a társadalmi reprezentációk vagy a diskurzusok mögött rejlő feltételezések érdeklik. Az RTA-kutatási kérdések nyitottak, és a jelentésmintázatokra összpontosítanak. Céljuk, hogy feltárják azokat a témákat, amelyek egy adott csoport számára fontosak. A módszer kerüli az egyedi, személyes élettörténetre (idiografikus fókusz) vagy a szigorú ok-okozati összefüggésekre irányuló kérdéseket. Példa egy jó RTA-kérdésre: „Hogyan konstruálják meg a Z-generációs munkavállalók a munkahelyi lojalitás jelentését a szakmai közösségi oldalakon?" Példa egy kerülendő, nem RTA-jellegű kérdésre: „Hogyan éli meg a három legnagyobb budapesti PR-cég vezetője a digitális átállás okozta személyes identitásválságot?" (Ez inkább IPA-kérdés.)

5.2. A minta

Az RTA-ban a mintaválasztás célzott, de a módszer nagyobb rugalmasságot enged, mint az IPA. A cél olyan résztvevők vagy adatforrások kiválasztása, akik vagy amelyek gazdag és érdemi információt nyújtanak a vizsgált témáról. A minta két fő jellemzője a következő:

a) Az RTA nem követi a „mélység a szélesség helyett" elvét olyan szigorúan, mint az IPA. Mivel a cél a megosztott mintázatok feltárása, a minta általában nagyobb. Kisebb, egyetemi szintű kutatásokhoz elegendő lehet 6-10 interjú, de nagyobb projektek esetében a 15–30 fős minta a kedvező. Rövidebb szöveges adatok (például online hozzászólások) elemzésekor a minta akár több száz is lehet. Ez a rugalmasság különbözteti meg az RTA-t más, kötöttebb szabályrendszerrel dolgozó hermeneutikai módszerektől. Míg az interpretatív fenomenológia elemzés (IPA) kizárólag

egyének megélt tapasztalatainak mély, idiografikus feltárására összpontosít kis, homogén mintán, addig az RTA szélesebb körű témák és nagyobb, heterogénebb minták elemzésére is alkalmas. A megalapozott elmélettel (Grounded Theory) szemben az RTA célja nem feltétlenül egy formális elmélet felépítése, hanem az adatokkal való párbeszéd során kialakuló jelentésmintázatok szisztematikus kidolgozása és interpretációja.

b) Míg az IPA a homogén mintát részesíti előnyben, az RTA hatékonyan képes kezelni a heterogénebb mintákat is. A kutató szándékosan választhat különböző hátterű résztvevőket, hogy összehasonlítsa a nézőpontokat, és feltárja a témák különböző megnyilvánulási formáit egy tágabb jelenségen belül.

9. doboz: Adatgyűjtési technikák és az adatszaturáció kérdése

Az RTA nem kötődik egyetlen sajátos adatgyűjtési technikához sem; a lényeg, hogy az adatok gazdagok, részletesek és árnyaltak legyenek. Bár leggyakrabban félig strukturált interjúk elemzésére használják, RTA-val sikeresen elemezhetők fókuszcsoportos beszélgetések, nyílt végű kérdőíves válaszok, naplók, online fórumok és médiaanyagok is. Az adatszaturáció az a pont, amelyet akkor érünk el, amikor az újabb adatok már nem hoznak új információt. Braun és Clarke kritikusan viszonyulnak ehhez a fogalomhoz, mivel szerintük a kvalitatív elemzés minősége nem a minta méretével, hanem a kutató elemzői munkájának mélységével mérhető. Helyette azt javasolják, hogy a kutató a kutatási kérdés és az adatok minősége alapján döntsön a minta nagyságáról.

5.3. A kutatás jellege

Az RTA a legjobb választás, ha a kutatást a következők jellemzik:

a) A kutatás a megosztott jelentésmintázatokra összpontosít. A cél nem az egyéni, egyedi tapasztalat mély feltárása, hanem az adathalmazon átívelő közös témák, értelmezések és reprezentációk azonosítása.

b) A felszíni és a rejtett jelentések egyaránt érdekesek. A módszer keretrendszere kifejezetten bátorítja a kutatót, hogy a manifeszt tartalmak mellett a rejtett jelentésrétegeket is feltárja.

c) Szisztematikus, de rugalmas elemzési folyamatra van szükség. Az RTA hatfázisú folyamata egyértelmű struktúrát kínál, de teret enged a kutatói kreativitásnak, interpretációnak és a folyamat önmagába visszatérő jellegének.

d) Közepes vagy nagyobb adathalmazok elemzése a cél. A módszer szisztematikus jellege lehetővé teszi, hogy a kutató nagyobb mennyiségű strukturálatlan adatot elemezzen anélkül, hogy elveszne a részletekben.
e) Rugalmas keretrendszerre van szükség. Az RTA nem kötődik egyetlen sajátos episztemológiai kerethez sem, így a kutató a saját elméleti elköteleződésének megfelelően használhatja.

A módszer erősségeinek megértéséhez érdemes azt is tisztázni, hogy milyen kutatási célokra nem alkalmas. Ne használjon RTA-t, ha:

a) A fókusz az egyedi, személyes megélt tapasztalat mély, idiografikus feltárásán van. Ha a cél egyetlen személy vagy egy nagyon kis, homogén csoport szubjektív világának rendkívül részletes megértése, válassza az interpretatív fenomenológiai elemzést.
b) A cél egy teljesen új elmélet szisztematikus felépítése az adatokból. Ha a kutatás kifejezett célja ez, a megalapozott elmélet (Grounded Theory) kínál ehhez egy jóval kidolgozottabb módszertani keretrendszert.
c) A kutatás a nyelvhasználat finom részleteire vagy annak performatív jellemzőire irányul. Ilyen kutatási célokhoz a különböző diskurzuselemzési irányzatok sokkal alkalmasabbak.

6. A módszer változatai és haladó alkalmazásai

Az RTA más analitikai eljárásokkal is kombinálható, ami növelheti az elemzési folyamat átláthatóságát és módszertani szigorát. Kiváló példa erre a kvalitatív összehasonlító elemzés (Qualitative Comparative Analysis, QCA) integrálása a témafejlesztés folyamatába. A hibrid megközelítésben az RTA kódolási és témafejlesztési szakaszai után a QCA logikája és eszköztára használható fel a témák közötti összetett, konfigurációs viszonyok feltárására.

Egy másik haladó példa a reflexív tematikus elemzés és a nagy nyelvi modellek (LLM-ek) vegyes alkalmazása. Ebben az esetben a kutatás a kutató és a mesterséges intelligencia erősségeit ötvözi. A kutató végzi el a mély, hermeneutikai munkát, vagyis kialakítja a kódrendszert, meghatározza a kódokat, és reflektál

saját értelmezési kereteire. Miután a kutató a kódkönyvet a modell számára is érthetővé tette és tesztelte, a nagy nyelvi modell végzi el a kódok alkalmazását a több ezer dokumentumból álló adathalmazon. Ez a vegyes megközelítés lehetővé teszi a hermeneutikai kutatás nagy léptékű kiterjesztését anélkül, hogy feláldozná annak interpretatív mélységét. A kutató marad az elemzési folyamat alakítója, a technológia segítségével azonban olyan mintázatokat is kidolgozhat, amelyek kisebb minták alapján nem lennének azonosíthatók.

7. Összefoglalás

A reflexív tematikus elemzés (RTA) a tematikus elemzés korszerű, interpretatív és elméletileg rugalmas válfaja, amely mélyen a „Big Q"-paradigmában gyökerezik. Alapvetően eltér a hagyományos, pozitivista szemléletű tartalomelemzéstől, mivel a kutatót nem tárgyilagos, semleges adatfeldolgozónak, hanem az elemzési folyamat alkotó és szubjektív résztvevőjének tekinti. Az RTA szerint a jelentés nem az adatokban rejlő, felfedezésre váró igazság, hanem a kutató és az adatok kölcsönhatásában jön létre. Ennek megfelelően a kutatói szubjektivitás nem kontrollálandó hibaforrás, hanem az elemzés legfőbb és legértékesebb erőforrása. Az RTA ennek megfelelően a témákat nem az adatokban már meglévő, felbukkanó egységeknek tekinti, hanem a kutató aktív értelmező munkájának eredményeként felfogott központi szervezőelveknek.

A módszer központi eleme a reflexivitás. A kutatónak folyamatosan és kritikusan vizsgálnia kell saját előfeltevéseit, döntéseit és az elemzésre gyakorolt hatását; ezt a munkát a kutatói napló vezetése segíti. Az RTA hat jól meghatározott, de nem mereven lineáris, hanem rekurzív szakaszból álló folyamatot kínál az adatokkal való ismerkedéstől a kódoláson, a témafejlesztésen és -felülvizsgálaton át a kutatási jelentés megírásáig. A módszer minőségi ismérvei nem a kódolói egyezésen, hanem a rendszerezett, átlátható, mélyen interpretatív és jól érvelő elemzésen nyugszanak. Az RTA leginkább olyan feltáró kutatási kérdések megválaszolására alkalmas, amelyek a résztvevők által megalkotott jelentésmintázatok megismerésére irányulnak.

8. Önellenőrző kérdések

1. Mi a legfőbb filozófiai különbség a reflexív tematikus elemzés és a hagyományos kvalitatív tartalomelemzés között a kutató szerepét illetően?

Válasz: A legfőbb különbség a kutatói szubjektivitáshoz való viszonyban rejlik. A hagyományos tartalomelemzés a kutatót tárgyilagos, távolságtartó eszköznek tekinti, akinek a szubjektivitása kontrollálandó hibaforrás. Ezzel szemben a reflexív tematikus elemzés a kutatót az elemzés tevékeny, szubjektív résztvevőjének tekinti, akinek a személyes háttere, értékei és elméleti tudása az elemzés központi és értékes erőforrása.

2. Mit jelent a gyakorlatban az a sokat idézett mondat, hogy „a témák nem bukkannak fel”?

Válasz: Ez a mottó azt jelenti, hogy a témák nem az adatokban rejlő, felfedezésre váró jelenségek. Ehelyett a témák a kutató tevékeny, intellektuális és interpretatív munkájának eredményeként jönnek létre, vagyis a kutató alkotja meg őket az adatokkal folytatott mély, reflexív párbeszéd során.

3. Sorolja fel az RTA hat elemzési szakaszát! Miért nevezzük ezt a folyamatot inkább rekurzívnak, mint lineárisnak?

Válasz: Az RTA hat szakasza a következő. 1. Ismerkedés az adatokkal; 2. Kezdeti kódok létrehozása; 3. A kezdeti témák kialakítása; 4. A témák felülvizsgálata; 5. A témák meghatározása és elnevezése; 6. A jelentés megírása. A folyamatot azért nevezzük rekurzívnak, mert a kutató nem mereven, lépésről lépésre halad előre, hanem gyakran visszatér korábbi szakaszokhoz, ahogyan gondolatai mélyülnek.

4. Miért nem alkalmazzák az RTA-ban a kódolók közötti megbízhatóság (például Cohen-kappa) mérését a minőségbiztosítás eszközeként? Milyen követelmények lépnek a helyébe?

Válasz: Az RTA azért veti el a kódolók közötti megbízhatóság mérését, mert az ilyen eljárás az objektivitásra törekvő, pozitivista paradigmához kötődik, ahol a cél az egyetértés. Az RTA ezzel szemben a „Big Q"-paradigmához tartozik, ahol a kutató egyedi nézőpontja és szubjektív interpretációja a cél, nem pedig az, hogy több kódoló ugyanarra az eredményre jusson. A megbízhatóság mérésének helyébe a hitelesség, az átláthatóság és a reflexivitás követelményei lépnek.

5. Milyen négy elméleti döntést kell meghoznia a kutatónak az RTA-elemzés megkezdése előtt?

Válasz: A kutatónak négy szempont mentén kell tisztáznia elméleti pozícióját. Először az esszencialista és a konstruktivista episztemológia között kell választani, vagyis arról kell döntenie, hogy a nyelvet a valóság hű tükrözőjének vagy a valóság megalkotójának tekinti-e. Másodszor az élményközpontú és a kritikai orientáció között kell választani attól függően, hogy a résztvevők szubjektív tapasztalatait vagy az azokat alakító társadalmi-hatalmi struktúrákat helyezi-e a középpontba. Harmadszor az induktív és a deduktív elemzési logika között kell döntenie, vagyis hogy az adatokból építkezve vagy egy meglévő elméleti keretet analitikus lencseként használva közelít-e az anyaghoz. Végül a szemantikus és a látens fókusz között kell választani, vagyis hogy a szöveg felszíni, kimondott jelentésére vagy a sorok között megbúvó, rejtett jelentésekre és ideológiákra összpontosít-e.

6. Milyen típusú kutatási kérdés megválaszolására NEM alkalmas az RTA? Mondjon olyan módszereket, amelyek ilyen esetekben jobb választások lehetnek!

Válasz: Az RTA nem alkalmas akkor, ha a cél az egyedi, személyes megélt tapasztalat mély, idiografikus feltárása, egy új elmélet szisztematikus felépítése vagy a nyelvhasználat finom részleteinek vizsgálata. Ilyen esetekben jobb választás lehet az interpretatív fenomenológiai elemzés (IPA) az egyedi élményekre, a megalapozott elmélet (Grounded Theory) az elméletalkotásra vagy a diskurzuselemzés a nyelvhasználat vizsgálatára.

Ajánlott irodalom

Braun, V., & Clarke, V. (2021). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research, 21*(1), 37–47. https://doi.org/10.1002/capr.12360

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health, 13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

Braun, V., & Clarke, V. (2023). Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. *Health Psychology Review, 17*(4), 695–718. https://doi.org/10.1080/17437199.2022.2161594

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity, 56*(3), 1391–1412. https://doi.org/10.1007/s11135-021-01182-y

Campbell, K., Orr, E., Durepos, P., Nguyen, L., Li, L., Whitmore, C., Gehrke, P., Graham, L., & Jack, S. (2021). Reflexive Thematic Analysis for Applied Qualitative Health Research. *The Qualitative Report*. Advance online publication. https://doi.org/10.46743/2160-3715/2021.5010

Conaghan, C., Daly, E., & Ryan, L. (2024). Reflexive thematic analysis of interviews with postprimary educators' experience and attitudes toward concussion. *Psychology in the Schools, 61*(9), 3714–3727. https://doi.org/10.1002/pits.23257

Jenkins, R. (2025). *Pierre Bourdieu*. Routledge.

Krippendorff, K. (2019). *Content Analysis: An Introduction to Its Methodology*. SAGE. https://doi.org/10.4135/9781071878781

Neuendorf, K. A. (2017). *The Content Analysis Guidebook*. SAGE. https://doi.org/10.4135/9781071802878

Nicmanis, M. (2024). Reflexive Content Analysis: An Approach to Qualitative Data Analysis, Reduction, and Description. *International Journal of Qualitative Methods, 23*, Article 16094069241236603. https://doi.org/10.1177/16094069241236603

Ragin, C. C. (2014). *The comparative method : moving beyond qualitative and quantitative strategies*. University of California Press.

Silver, C., & Lewins, A. (2014). *Using Software in Qualitative Research: A Step-by-Step Guide*. SAGE. https://doi.org/10.4135/9781473906907

Terry, G., & Hayfield, N. (2021). *Essentials of thematic analysis*. American Psychological Association. https://doi.org/10.1037/0000238-000

Tinto, V. (2017). Through the Eyes of Students. *Journal of College Student Retention: Research, Theory & Practice, 19*(3), 254–269. https://doi.org/10.1177/1521025115621917

Trainor, L. R., & Bundon, A. (2021). Developing the craft: reflexive accounts of doing reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health, 13*(5), 705–726. https://doi.org/10.1080/2159676X.2020.1840423

Wen, C., Clough, P., Paton, R., & Middleton, R. (2025). Leveraging large language models for thematic analysis: a case study in the charity sector. *AI & SOCIETY*. Advance online publication. https://doi.org/10.1007/s00146-025-02487-4

IV. SZEMANTIKUS TARTALOMELEMZÉS (SCA)

TURÓS MÁTYÁS

A fejezet a következő kérdéseket válaszolja meg:

a) Mi a szemantikus tartalomelemzés (SCA), és hogyan viszonyul a tágabb értelemben vett kvalitatív tartalomelemzéshez (QCA)?
b) Miben különbözik az SCA elméleti alapja és célja a reflexív tematikus elemzéstől (RTA) és az interpretatív fenomenológiai elemzéstől (IPA)?
c) Milyen lépésekből áll egy szemantikus tartalomelemzésen alapuló kutatás folyamata a kutatási kérdéstől az eredmények közléséig?
d) Hogyan biztosítható a kutatás tudományos minősége és megbízhatósága a hitelesség, átvihetőség, konzisztencia, megerősíthetőség követelményei mentén?
e) Milyen helyzetekben és kutatási kérdések esetén érdemes az SCA módszerét választani, és mikor jelent jobb alternatívát más eljárás?
f) Melyek az SCA alapvető változatai, és milyen haladó elemzési lehetőségeket kínál a módszer?

1. Bevezetés

Az interpretatív fenomenológiai elemzés (IPA) és a reflexív tematikus elemzés (RTA) után a szakirodalomban kevésbé egységesen értelmezett területre érkezünk, a szemantikus tartalomelemzéshez (SCA). Míg az IPA a megélt tapasztalatra, az RTA pedig a kutatói reflexivitásra és a jelentés megalkotására összpontosít, addig az SCA olyan megközelítést képvisel, amelynek pontos helye és szerepe gyakran vita tárgya. A dilemma abból fakad, hogy a „szemantikus tartalomelemzés" kifejezést (és a kvalitatív tartalomelemzést is) a tudományos diskurzusban ernyőfogalomként használják. A „szemantikus tartalomelemzés" kifejezés szorosan összefonódik a kvalitatív tartalomelemzés tágabb kategóriájával, amelyet gyakran a tematikus elemzés szabálykövetőbb, a szöveg kimondott jelentéstartalmának rendszerezésére törekvő formájaként írnak le.

Mielőtt rátérnénk a szemantikus tartalomelemzésre (SCA), fontos elhelyeznünk azt a módszertani térképen és tisztáznunk néhány gyakori fogalmi félreértést. Fontos megkülönböztetni az SCA-t a tisztán matematikai alapú eljárásoktól, mint a topikmodellezés (például LDA). Bár ezek is a szövegek rejtett fogalmi struktúráit tárják fel, módszertanuk alapvetően algoritmikus, míg az SCA a kvalitatív, kódolásalapú elemzéshez áll közelebb. A fogalmi tisztázáshoz tartozik a „QCA" rövidítés problémája is. Ez a betűszó egyszerre jelölheti a kvalitatív tartalomelemzést (Qualitative Content Analysis) – ami lényegében az SCA szinonimája – és a következő fejezetben tárgyalt, teljesen eltérő kvalitatív összehasonlító elemzést (Qualitative Comparative Analysis) is. Mivel a két módszer összekeverése súlyos félreértésekhez vezet, a rövidítés alapjául szolgáló módszerről mindig érdemes tájékozódni.

Ez a fogalmi sokszínűség minden módszertani kötetet kihívás elé állít. Melyik definíciót kövessük, és hogyan illesszük azt koherens logikai ívbe? Az érvek alapján a fejezet címe kvalitatív tartalomelemzés és szemantikus tartalomelemzés is lehet, a megnevezés tekintetében mindkettő előnyeivel és hátrányaival.

Kötetünk, ahogy azt a bevezetőben is lefektettük, a strukturálatlan adatoktól a strukturált adatok felé haladó logika mentén épül fel. E belső koherencia megőrzésének érdekében ebben a tankönyvben a szemantikus tartalomelemzést (SCA) a kvalitatív tartalomelemzés sajátos, a jelentés szisztematikus feltárására és kódolására összpontosító, szabálykövetőbb válfajaként határozzuk meg. Ennek oka, hogy ez a megközelítés hidat képez a korábbi, tisztán hermeneutikai módszerek és a későbbi, már strukturált adatokkal dolgozó modellezési

eljárások között. A meghatározás alapján az SCA alapvetően különbözik az előző fejezetben tárgyalt reflexív tematikus elemzéstől. Míg az RTA a „Big Q"-paradigmában gyökerezik, ahol a kutató a jelentés tevékeny alkotója, és a szubjektivitás értékes erőforrás, addig az általunk bemutatott SCA a „small q"-szemlélethez áll közelebb. Itt a kutató szerepe inkább a szövegben rejlő jelentések szisztematikus feltárása és megbízható kódolása. A figyelem a kutatói önreflexióról a kódolási folyamat konzisztenciájára és a kategóriák átláthatóságára helyeződik át. A cél nem egy mély, interpretatív történet elmesélése, hanem a szöveges adatok tartalmának redukálása, megbízható rendszerezése és egyértelmű kategóriákba sorolása. Ez a meghatározás tudatosan elkülöníti a tárgyalt módszert a tisztán matematikai eljárásoktól. A látens szemantikai elemzés (LSA), a topikmodellezés és más, a strukturálatlan szöveget numerikus mátrixszá alakító technikák kötetünk logikája szerint a természetesnyelv-feldolgozás (NLP) témakörébe tartoznak, ezeket későbbi fejezetben mutatjuk be. Ebben a fejezetben tehát az SCA-t mint a strukturálatlan adatok strukturálása felé tett első lépést vizsgáljuk. Az SCA gyakorlatias és szisztematikus eszköz, amely lehetővé teszi a szöveges adatok megbízható elemzését, és előkészíti a terepet az olyan, már a kódolt adatok közötti logikai kapcsolatokat vizsgáló módszerek számára, mint amilyen a következő fejezetben tárgyalt kvalitatív összehasonlító elemzés (QCA). Végül megjegyezzük, hogy a fogalomhasználatot érintő választásunk (QCA helyett SCA) ellenére e fejezetben az SCA kifejezés felcserélhető a QCA (mint kvalitatív tartalomelemzés) fogalmával.

2. Elméleti alapok

A szemantikus tartalomelemzés elméleti alapjait pragmatikus, módszertani szempontrendszer határozza meg. Az SCA nem sajátos filozófiai iskolából nőtt ki, hanem a hagyományos tartalomelemzésből fejlődött ki azzal a céllal, hogy a strukturálatlan adatok elemzését rendszerezetté, átláthatóvá és megbízhatóvá tegye.

Ahogyan azt már említettük, a nemzetközi szakmai diskurzusban a kvalitatív tartalomelemzés ernyőfogalom, amely többféle megközelítést foglal magában. Ezen belül három alapvető típust különíthetünk el, amelyek abban különböznek, hogy a kódok és kategóriák honnan származnak.

a) Konvencionális (conventional) tartalomelemzés. Ez a megközelítés induktív. A kutató előzetes elméletek vagy kategóriák nélkül merül el az adatokban, és hagyja, hogy a kódok és témák közvetlenül a szövegből, alulról felfelé bontakozzanak ki. Akkor ideális, ha a jelenségről kevés előzetes tudás áll rendelkezésre, és a cél a feltárás.
b) Irányított (directed) tartalomelemzés. Ez a megközelítés deduktív. Az elemzés már meglévő elméletből, modellből vagy a témához tartozó szakirodalomból levezetett kategóriákkal kezdődik. A kutató célja, hogy tesztelje, validálja vagy kiterjessze a meglévő elméleti keretet. Az adatok elemzése során az előre meghatározott osztályozás finomodhat, és új kódok is megjelenhetnek.
c) Összegző (summative) tartalomelemzés. Ez a típus a szavak vagy kifejezések előfordulásának számszerűsítésével kezdődik, majd a mögöttes kontextus értelmezésével folytatódik. Ez a megközelítés hidat képez a tisztán matematikai és tisztán hermeneutikai elemzés között, feltárva a szavak használatának mintázatait és rejtett jelentéseit.

Az itt bemutatott szemantikus tartalomelemzés leginkább a konvencionális és az irányított tartalomelemzés elveiből építkező, de azoknál szigorúbb, szabálykövetőbb és a megbízhatóságra erősebben összpontosító eljárás. Az SCA központi elméleti tétele a kutató szerepének és a jelentés természetének újrafogalmazása az RTA-hoz képest. Míg az RTA a kutatói szubjektivitást az elemzés értékes erőforrásának tekinti, addig az SCA a szubjektivitást kontrollálandó, a kutatás megbízhatóságát esetleg veszélyeztető tényezőként kezeli. A kutató itt nem a jelentés tevékeny, interpretatív alkotója, hanem képzett és tárgyilagos elemző, akinek feladata, hogy a szövegben rejlő jelentéseket előre meghatározott vagy szigorú szabályok mentén kialakított kategóriarendszer segítségével következetesen azonosítsa és rendszerezze. Ebből a szempontból válik fontossá a nyílt és a rejtett jelentéstartalom megkülönböztetése.

A nyílt tartalom (amelyet gyakran manifeszt tartalomnak is neveznek) a felszíni, kimondott, szó szerinti jelentésre utal. A nyílt tartalom vizsgálatával arra a kérdésre lehet válaszolni, hogy „Mit mondtak?". A rejtett tartalom (amelyet gyakran látens tartalomnak is neveznek) a mögöttes, interpretációt igénylő jelentésrétegre utal. A rejtett tartalom vizsgálatával arra a kérdésre lehet válaszolni, hogy „Mit jelentett az, amit mondtak?". Az SCA elsődleges célja a nyílt, szemantikus tartalom megbízható leírása és redukálása, a szövegben nyíltan megjelenő adatok rendszerezése.

Ebből következik az SCA két legfontosabb elméleti-módszertani alapelve, a szisztematikusság és a megbízhatóság. A szisztematikusság azt jelenti, hogy az

elemzési folyamat minden lépése világosan meghatározott szabályokon alapul. Az elemzés fontos eszköze a részletes kódkönyv, amely pontosan meghatározza a kategóriákat és azok alkalmazási szabályait. A folyamat lineárisabb, lépésről lépésre haladó logikát követ, amelynek célja az adatok következetes feldolgozása. Az RTA-val ellentétben, amely szándékosan elveti a kódolók közötti egyezés mérését, az SCA ezt kifejezetten fontosnak tartja.

Összefoglalva, a szemantikus tartalomelemzés elméleti alapját olyan pragmatikus és szabálykövető paradigma képezi, amely a strukturálatlan adatok elemzését a megbízhatóság, az átláthatóság és a szisztematikusság elvei mentén szervezi. Az SCA a kutatót tárgyilagos kódolónak tekinti, a jelentést pedig a szöveg nyílt, szemantikus tartalmában keresi. Az SCA robusztus, jól dokumentálható és a strukturálatlan adatokat hatékonyan rendszerező módszer, amely hidat képez a mélyinterpretációra építő és a modellező, matematikai alapú eljárások között.

Az SCA változatai

A szemantikus tartalomelemzés nem egységes, hanem többféle változatban létező eljárás. Az SCA három leggyakoribb célja a leírás (strukturálás), az értékelés és a típusalkotás. A gyakorlatban a kutatási cél határozza meg, hogy melyik változatot alkalmazzuk.

a) Strukturáló szemantikus tartalomelemzés (structuring QCA). Ez az SCA alapvető és leggyakrabban használt formája. Célja az adatokban rejlő témák és altémák azonosítása, rendszerezése és elemzése. A középpontban az áll, hogy az anyagot a kutatási kérdés szempontjából lényeges témák szerint rendszerezze, és átfogó képet adjon a vizsgált jelenségről. A strukturáló elemzés megvalósulhat induktívan (amikor a kategóriák tisztán az adatokból bontakoznak ki, megfelelve a konvencionális SCA logikájának) vagy deduktívan (amikor egy meglévő elméleti keret adja a fő kategóriákat, megfelelve az irányított SCA logikájának). A leggyakrabban azonban a deduktív-induktív kevert megközelítést alkalmazzák. Tipikusan többciklusos kategóriafejlesztési és kódolási folyamaton alapul, ahogyan azt az SCA folyamatát bemutató szakaszban részletesen kifejtjük.
b) Értékelő szemantikus tartalomelemzés (evaluative QCA). Ez a változat túllép a témák azonosításán, és a tartalom értékelésére, osztályozására és felmérésére összpontosít. A kutatók egy előre meghatározott dimenzió mentén értékelik az adatokat. Például nemcsak azt kódolják, hogy az interjúalany beszél-e a személyes felelősségvállalásról, hanem azt is, hogy

az általa elmondottak alacsony, közepes vagy magas szintű felelősségvállalást fejeznek-e ki. Mivel az értékelési dimenzió (skála) szinte mindig elméleti megfontoláson vagy korábbi kutatáson alapul, ez a változat tipikusan erősen deduktív, irányított megközelítésnek felel meg. Általában ordinális skálákat használó kategóriákat épít. Holisztikusabb, az egész esetre vagy egy-egy szövegrészletre vonatkozó értékelést igényel, és nagyobb szakértelmet követel a kódolóktól, mivel a besorolás gyakran interpretatívabb. Különösen alkalmas elméletorientált kutatásokra.

c) Típusalkotó szemantikus tartalomelemzés (type-building QCA). A strukturálatlan adatok elemzésének egyik célja a tipológiák létrehozása. Ez a módszerváltozat nem a témákat, hanem magukat az eseteket (például személyeket, szervezeteket) csoportosítja a több dimenzió mentén mutatott jellegzetes mintázataik alapján. A cél olyan típusok azonosítása, amelyek belsőleg egyneműek (homogének), egymástól pedig a lehető leginkább különböznek (heterogének). Ez a legösszetettebb és módszertanilag legigényesebb változat, amely szinte mindig a strukturáló vagy értékelő SCA előzetes munkájára épül. A típusalkotás alapját azok a központi kategóriák (dimenziók) képezik, amelyeket korábbi elemzési szakaszban, akár induktívan, akár deduktívan, azonosítottunk. Ezek a dimenziók adják a tipológia koordinátarendszerét.

3. A kutatás folyamata

A következőkben a strukturáló SCA folyamatát mutatjuk be az alábbi fázisokon keresztül.

3.1. A kutatási kérdés megfogalmazása

A kutatási folyamat elején fontos tisztázni a kutató céljait és reflektálni a saját előfeltevéseire. A nyitottság azt jelenti, hogy a kutató elfogadja az adatokból feltáruló új nézőpontokat, és hajlandó felülvizsgálni saját előzetes elképzeléseit. Ennek megfelelően az SCA kutatási kérdése a kutatás során finomodhat. A kutató az adatokkal való kezdeti ismerkedés során új, váratlan szempontokat fedezhet fel, amelyek a kérdés pontosítását vagy módosítását tehetik szükségessé.

3.2. A kezdeti munka a szöveggel: olvasás, jegyzetelés és eset-összefoglalók készítése

A formális elemzés megkezdése előtt elengedhetetlen az adatokkal való alapos, elmélyült megismerkedés. Ezt a szakaszt a hermeneutikai hozzáállás jellemzi, amely nem egyszerű feladat, hiszen több száz oldalnyi, látszólag rendezetlen információdarab állhat rendelkezésre. A cél ekkor még nem a rendszerezés, hanem a türelmes elmélyülés, amely során időt kell hagyni arra, hogy többször is végigolvassuk a szöveget anélkül, hogy azt azonnal kategóriákba akarnánk szorítani. A megértés ebből a kezdeti befogadó hozzáállásból születik. Ekkor a cél tehát még nem a kódolás, hanem a szöveg jelentésének átfogó befogadása. A folyamat lépései a következők.

a) Mélyreható olvasás. A szövegeket (például interjúátiratokat) az elejétől a végéig figyelmesen el kell olvasni a kutatási kérdés szempontjait szem előtt tartva.
b) Kiemelés. A központi fogalmakat, fontos részeket és a nehezen érthető szakaszokat érdemes megjelölni.
c) Jegyzetelés. A szöveg olvasása közben felmerülő gondolatokat, ötleteket, hipotéziseket és értelmezési lehetőségeket megjegyzések formájában kell rögzíteni. A megjegyzések lehetnek leíró jellegűek (az adatokat foglalják össze), értelmező jellegűek (a kutató saját interpretációit tartalmazzák, például „ennek az oka az lehetne, hogy…") és módszertani jellegűek (a további elemzés lépéseire vonatkozó döntéseket rögzítik, például „ezt a témát érdemes lenne összevetni más esetekkel is").
d) Eset-összefoglalók készítése. Minden esetről (például interjúról) rövid, tényszerű összefoglalót kell írni. Az eset-összefoglaló a kutatási kérdés szempontjából legfontosabb információkat sűríti, de nem tartalmazza a kutató saját hipotéziseit. Célja, hogy gyors áttekintést adjon az adott esetről, és megalapozza a későbbi összehasonlító elemzést. Az eset-összefoglalóhoz adható mottó vagy cím is, amely megragadja az eset lényegét (például „R1 interjú: Pozitív hozzáállás ambíciók nélkül"). Ez a szakasz alapozza meg az esetorientált szemléletet, és biztosítja, hogy a kutató ne veszítse el a teljes képet a későbbi, részletekre összpontosító kódolás során.

3.3. A kategóriarendszer fejlesztése és a kódolás folyamata

Az elemzés fontos lépése a kategóriarendszer és az azt leíró kódkönyv kialakítása. A kategóriák olyan címkék, amelyek alá az oda tartozó szövegrészleteket besoroljuk. Kialakításukra többféle eljárás létezik. A három fő módszert (induktív, deduktív, abduktív kódolás) e könyv III. fejezetében, a reflexív tematikus elemzés bevezető részében, példával a 6. dobozban már ismertettük.

A kódolás a kategóriarendszer gyakorlati alkalmazása, amelynek során a kutató a szövegrészleteket (szegmenseket) hozzárendeli a megfelelő kategóriákhoz. A folyamat jellemzően két fő ciklusban zajlik.

a) Első kódolási ciklus. A teljes adatállományt végigkódolják a tág főkategóriákkal.
b) Második kódolási ciklus. Az egyes főkategóriákon belül a már kódolt szövegrészleteket hozzárendelik a részletesebb, újonnan képzett alkategóriákhoz.

A szemantikus tartalomelemzésben a kódolás alapegysége nem nyelvtani vagy formai egység (például szó, mondat, bekezdés), hanem téma vagy önálló gondolat. A kódolt szegmens méretét tehát szemantikai követelmények határozzák meg. Egyetlen téma kifejeződhet egyetlen szóban, frázisban, teljes mondatban, de akár egész bekezdésben is. A lényeg, hogy a kódolt egységnek teljes, környezetéből kiemelve is érthető gondolatot kell tartalmaznia, amely lényeges a kutatási kérdés szempontjából. Egy szövegrészlet több témát is érinthet, ezért lehetséges és gyakran szükséges is több kategóriához rendelni.

A kódolás megbízhatósága növelésének érdekében javasolt a konszenzuális kódolás. Ennek során legalább két kódoló egymástól függetlenül végzi el a kódolást, majd összevetik munkájukat, és a véleménykülönbségeket megvitatva jutnak egyezségre. Ez a folyamat nem csupán hibajavítás, hanem a kategóriameghatározások finomításának és pontosításának aktív eszköze.

10. doboz: A konszenzuális kódolás lépései és a kódkönyv fejlesztése

A magas minőségű kódolás a gyakorlatban egy iteratív, együttműködő konszenzusépítési folyamat, amelynek célja nem a nézeteltérések egyszerű eldöntése, hanem egy közösen elfogadott, következetes értelmezési keret kialakítása. A folyamat, ahogy azt a tudományos közlemények is tükrözik, jellemzően a részletes kódkönyv kidolgozásával kezdődik, amely minden kategóriát meghatároz, és példákkal támaszt alá. Ezt követi a próbakódolás vagy kalibrációs szakasz, ahol a kódolók egymástól függetlenül feldolgoznak egy kisebb, reprezentatív adatmintát. Ennek a folyamatnak a lényeges lépése a megegyezésre törekvő megbeszélés. A kódolók itt nem csupán a nézeteltéréseket azonosítják, hanem megvitatják azok okait, ami gyakran a kódkönyv meghatározásainak vagy szabályainak pontatlanságára világít rá. A vita eredményeképpen a kódolók közösen finomítják a kódkönyvet, megerősítve a közös értelmezési keretet. Csak e ciklusok után kerül sor a teljes adathalmaz feldolgozására, ahol a nézeteltérések száma már jóval alacsonyabb. A megmaradó vitás esetekben a végső döntés születhet többségi elv alapján vagy a kutatásvezető bevonásával. A folyamat végén számított megbízhatósági mutató (például Krippendorff-alfa) annak bizonyítéka, hogy a kutatócsoport sikeresen végigment a szisztematikus és reflektív konszenzusépítési folyamaton.

3.4. Az elemzés

A kódolás után kezdődik az elemzés, amelynek célja a mintázatok, összefüggések és jelentések feltárása. Az elemzés tehát több, mint a kategóriák gyakoriságának összeszámolása. Főbb irányai a következők.

a) Kategóriaalapú elemzés. A kérdés: „Mit mondtak a válaszadók egy adott témáról?" Az elemzés során a kutató bemutatja az egyes kategóriákban és alkategóriákban megjelenő tartalmakat, hasonlóságokat, különbségeket és jellegzetes megnyilvánulásokat, prototipikus idézetekkel szemléltetve.
b) Kategóriák közötti kapcsolatok vizsgálata. Bonyolultabb, több kategóriát is érintő mintázatok keresése. Például: „Van összefüggés a válaszadók által említett világproblémák és a személyes felelősségvállalásuk mértéke között?"
c) Esetalapú és csoport-összehasonlítások. A cél az esetek közötti szisztematikus összehasonlítás, gyakran táblázatos esetáttekintések (mátrixok) segítségével, ahol a sorok az eseteket, az oszlopok a kategóriákat képviselik.
d) Kiválasztott esetek mélyelemzése. A mintázatok azonosítása után érdemes visszatérni egy-egy tipikus vagy éppen a mintából kiugró (deviáns)

eset részletes, mélyreható értelmezéséhez. Erre két fő stratégia létezik. Az egyik a reprezentatív eset interpretációja, amikor egyetlen, az adott típusra legjellemzőbb (prototipikus) esetet választunk ki, és azt mutatjuk be részletesen, mintegy példaként a típus összes többi tagjára. A másik a modelleset megalkotása, amely nem egyetlen személyre összpontosít. A kutató a típusba tartozó több válaszadótól származó legjellemzőbb szövegrészletekből montázst vagy szinopszist hoz létre, így a modelleset sűrített, ideális képét adja a típusnak.

e) Típusalkotás. Az elemzés egyik legmagasabb szintű célja a tipológiák létrehozása lehet. Ennek során a kutató az eseteket (személyeket) az érdemi dimenziók mentén mutatott hasonlóságaik alapján csoportokba (típusokba) rendezi. A cél olyan típusok azonosítása, amelyek belsőleg homogének, egymástól pedig heterogének.

A típusalkotás részletes folyamata

A típusalkotás a képzett kódok modern, haladó elemzési formája, és az SCA egyik kiemelt fontosságú felhasználási módja. Ennek megfelelően részletesebben is áttekintjük a többlépcsős folyamat lépéseit.

a) Adat-előkészítés. A típusalkotás szinte mindig korábban elvégzett tematikus vagy értékelő kódolásra épül. Előfordulhat, hogy a típusalkotáshoz a meglévő kategóriák skáláját egységesíteni vagy egyszerűsíteni kell. Például az ötfokú skálán mért álláskeresési aktivitást egyszerűsíthetjük dichotóm (aktívan keres kontra nem keres) változóra a tisztább típusok képzése érdekében.

b) A típusok számának meghatározása. Mielőtt a csoportosítás megkezdődne, mérlegelni kell, hány típus létrehozása lenne a legcélszerűbb. Ezt befolyásolja a minta mérete (százfős mintánál több típus különíthető el, mint húsz főnél), a kutatási kérdés és az eredmények kommunikálhatósága (egy négy típusból álló tipológia gyakran érthetőbb és átláthatóbb, mint egy nyolcból álló).

c) Az esetek típusokba sorolása. A kutató a kiválasztott dimenziók alapján minden esetet (például interjúalanyt) besorol a létrehozott típusok

egyikébe. Fontos szabály, hogy egy eset csak egy típushoz tartozhat. Például a környezettudatosság és a környezetvédő viselkedés dimenziói alapján létrehozható a „Következetes védelmező", a „Környezeti szónok", az „El nem kötelezett" és az „Ignoráns" típus.

Az így kialakított tipológia nemcsak rendszerezi az adatokat, hanem alapot ad az esetek mélyelemzéséhez is: minden típusból kiválasztható egy reprezentatív eset vagy létrehozható egy modelleset.

3.5. Az eredmények bemutatása és a folyamat dokumentálása

A kutatási folyamat utolsó szakasza a feltárt eredmények és tanulságok írásos rögzítése egy kutatási jelentésben, tanulmányban vagy disszertációban. A jelentés nem a kategóriák száraz leírása. Egy olyan észszerű értelmezést kell felépíteni, amely a kutatási kérdésből kiindulva a kategóriákon és típusokon keresztül vezeti végig az olvasót a legfontosabb eredményekig. A jelentésnek egyensúlyt kell teremtenie a nyers adatok (idézetek) bemutatása és a kutatói értelmezés között. Az idézetek alátámasztják és szemléltetik az elemzést, de az értelmezés adja meg a jelentésüket. Grafikonok, táblázatok, térképek és a típusalkotást bemutató ábrák segíthetik az eredmények átlátható és hatékony bemutatását. Továbbá a kutatás hitelességének alapja az átláthatóság. A kutatónak részletesen rögzítenie kell az elemzés folyamatát, hogy az olvasó számára követhető és ellenőrizhető legyen, hogyan jutott a következtetéseire. A dokumentációnak tartalmaznia kell a kategóriarendszer részletes leírását (kódkönyv), a kódolási szabályokat és az elemzési lépések bemutatását, melyeket a kutatási jelentés függelékében szokás elhelyezni.

11. doboz: A QDA-szoftverek specifikus funkciói a tartalomelemzés támogatására

Habár továbbra is van létjogosultsága a kézi feldolgozásnak, a cetlikre és színes kiemelőkre épülő munkafolyamatok mára szinte teljesen átadták a helyüket a kvalitatív adatelemző (QDA) szoftvereknek.

Az elemzés első lépése az adatok digitális környezetbe való importálása. A korszerű szoftverek itt nem csupán tárolási, hanem szervezési funkciókat is kínálnak. A hang- és videófelvételek szöveges átiratokkal való szinkronizálása ilyen funkció, amely lehetővé teszi, hogy a szöveg bármely pontjáról közvetlenül visszajátszható legyen az eredeti hang- vagy képanyag. Ez elengedhetetlen a paralingvisztikai jelek (hangsúly, szünet, hanglejtés) és a kontextus elemzésbe való bevonásához, így az interpretáció mélysége jelentősen növelhető. A szoftverek gyakran kínálnak olyan beépített szövegszerkesztői felületeket is, amelyek speciális funkciókkal (például a lejátszási sebesség állításával, időbélyegek automatikus beillesztésével) segítik az átírás folyamatát, biztosítva a szöveg és a forrásanyag közötti szoros kapcsolatot.

A kódolás folyamán létrehozott kategóriák és alkategóriák hierarchikus vagy hálózatos rendszerben, képként is megjeleníthetők. A kutató a teljes kódrendszert folyamatosan maga előtt látja, és a kategóriákat könnyedén átnevezheti, összevonhatja vagy újrarendezheti anélkül, hogy a már elvégzett kódolások elvesznének – a szoftver ugyanis automatikusan frissíti a kódok és a szövegrészletek közötti kapcsolatokat. A szoftverek emellett lehetővé teszik, hogy a kutató jegyzeteket fűzzön a dokumentumokhoz, egyes szövegrészletekhez vagy magukhoz a kódokhoz. Ez elengedhetetlen a kódkönyv létrehozásához (a kódokhoz fűzött jegyzetek tartalmazzák a definíciókat és szabályokat), valamint az elemzés során felmerülő interpretatív gondolatok és hipotézisek rögzítéséhez.

A kódolás befejezése után a szoftverek lehetővé teszik az adatok szisztematikus és sokrétű feltárását. A kutatónak nem kell keresnie és kivágnia a szövegrészleteket, hiszen a szoftver ezt a feladatot automatizálja. Minden QDA-szoftver alapvető funkciója, hogy képes ehy adott kategóriához rendelt összes szövegrészletet a teljes adatbázisból összegyűjteni és egyetlen listában megjeleníteni. A kategóriaorientált elemzés lehetővé teszi a téma mélyreható, az összes esetre kiterjedő vizsgálatát.

A haladóbb szoftverek lehetőséget adnak összetett, logikai feltételeken alapuló keresésekre is. Például lekérdezhetők azok a szövegrészletek, amelyekben két adott kategória együtt vagy egymás közelében fordul elő, illetve amelyekben az egyik kategória jelen van, a másik viszont nem. Ez a funkció a kategóriák közötti kapcsolatok és rejtett mintázatok feltárásának hatékony eszköze.

A szoftverek képesek olyan táblázatokat (mátrixokat) létrehozni, amelyek az esetek és a kategóriák kereszteződésében mutatják be az adatok szerveződését. Az ilyen mátrixból gyorsan leolvasható, hogy mely témák dominálnak bizonyos esetekben, vagy mely esetek hasonlítanak egymásra a tematikus profiljuk alapján. A cellákban megjelenő információ lehet egyszerű jelenlét/hiány, a kódolások gyakorisága vagy akár a kódolt szöveg terjedelme. Végül lehetőség van fogalmi térképek vagy hálózati ábrák készítésére, amelyek képként jelenítik meg a kategóriák, esetek és a kutatói jegyzetek közötti logikai kapcsolatokat. Ezek a vizuális ábrázolások hasznosak az elméletalkotási folyamat támogatásához és az eredmények szemléletes bemutatásához.

4. A minőség biztosítása

A szemantikus tartalomelemzés tudományos hitelességének és megbízhatóságának alapja a módszertani szigor és a kutatási folyamat átláthatósága. A megfelelő minőség nem utólagos ellenőrzéssel biztosítható, hanem a módszer minden szakaszába beépített, tudatosan alkalmazott eljárások eredménye. Míg a kvantitatív kutatások a validitás és reliabilitás fogalmait használják, a hermeneutikai kutatásokban – az eltérő tudományfilozófiai alapok miatt – a „trustworthiness" (magyarul gyakran „kutatási megbízhatóság" vagy „hitelesség") keretrendszere terjedt el. Ez a keret négy feltételre épül: a hitelességre (credibility), az átvihetőségre (transferability), a konzisztenciára vagy más néven „folyamati megbízhatóságra" (dependability) és a megerősíthetőségre (confirmability). A minőségbiztosítás elemeit e négy szempont mentén mutatjuk be.

a) A hitelesség arra a kérdésre ad választ, hogy a kutató interpretációi mennyire felelnek meg a résztvevők által megélt valóságnak. A cél annak biztosítása, hogy az eredmények ne a kutató fantáziájának szüleményei legyenek, hanem szilárdan az adatokban gyökerezzenek. Ennek legfontosabb eszköze a konszenzuális kódolás, amely kizárólag a „small q"-paradigmában alkalmazott eljárás. A folyamat lépései a következők. Először a kutatócsoport legalább két tagja egymástól függetlenül kódolja ugyanazokat az adatokat (például néhány interjút) jól leírt és példákkal megerősített kategóriarendszer alapján. Ezután a kódolók összevetik a kódolásaikat,

megvizsgálják az egyezéseket és az eltéréseket – vagy megbízhatósági mutatót számítanak, vagy megegyezésre törekednek. Végül ha nem jutnak megegyezésre, újabb kódolót vonnak be a döntés érdekében.

b) Az átvihetőség azt vizsgálja, hogy a kutatás eredményei mennyiben relevánsak más körülmények között vagy más csoportok számára. A hermeneutikai kutatás célja nem a statisztikai általánosítás, hanem az, hogy a kutató olyan részletes leírást adjon, amelynek alapján az olvasó maga tudja eldönteni, hogy az eredmények átvihetők-e az őt érdeklő helyzetre. Ennek kulcsa az elemzési folyamat átlátható lejegyzése. A kutatónak részletesen rögzítenie kell az elemzés teljes folyamatát, hogy az olvasó számára követhető legyen, hogyan jutott el a következtetéseihez. A dokumentációnak több elemet kell tartalmaznia. Először a kutatási jelentésben vagy annak függelékében részletesen le kell írni a kategóriarendszert, beleértve az összes fő- és alkategóriát és azok pontos meghatározásait. Másodszor a kódolási szabályokat és a prototipikus példákat is írásban rögzíteni kell, mivel ezek szemléltetik a módszertani megközelítést, és segítik az olvasót annak megértésében, hogy egy-egy kategória alá milyen típusú szövegrészletek kerültek. Harmadszor a kutatónak be kell mutatnia a folyamatleírást, vagyis az elemzés lépéseit – elmagyarázva, hogyan jöttek létre a kategóriák, és milyen mértékben alapultak az adatokon.

c) A konzisztencia, amelyet a kvantitatív kutatások megbízhatóságának hermeneutikai megfelelőjeként vezettek be, a kutatási folyamat stabilitására és logikai következetességére utal. A cél annak biztosítása, hogy ha egy másik kutató követné a leírt eljárást, hasonló eredményekre és következtetésekre jutna. Ezt a rendszerszintű és szabályvezérelt folyamat biztosítja. E tulajdonság különbözteti meg az SCA-t a szabad, művészi interpretációktól. A kutatás lépései – a kategóriák fejlesztésétől a kódoláson át az elemzésig – pontosan leírhatók és elsajátíthatók. A részletes dokumentáció biztosítja a folyamat követhetőségét és reprodukálhatóságát.

d) A megerősíthetőség arra összpontosít, hogy az eredmények valóban az adatokból származzanak, ne pedig a kutató elfogultságából, előfeltevéseiből vagy motivációiból. Bár a teljes objektivitás a hermeneutikai kutatásban nem értelmezhető cél, a törekvés az, hogy a kutatói szubjektivitás hatása ellenőrzött és átlátható legyen. A megerősíthetőség nagymértékben az előző három követelmény teljesülésének eredménye. Ha a kutatás hiteles (azaz az adatokból fakad), átvihető (azaz részletesen leírt) és konzisztens (azaz észszerű és szabálykövető), akkor nagy valószínűséggel megerősíthető is. A fentebb részletezett konszenzuális kódolás itt is fontos, mivel csökkenti az egyetlen kutató szubjektív torzításának esélyét.

5. Mikor használjuk és mikor ne használjuk a módszert?

5.1. A kutatási kérdés

Az SCA egyaránt alkalmas feltáró és leíró kutatásokhoz. A módszer jól működik olyan nyitott, feltáró jellegű kutatás esetében, ahol a cél a jelenség részletes leírása, a benne rejlő témák, szempontok és mintázatok azonosítása. Amikor a fő cél a leírás, a módszer strukturáló változata a legmegfelelőbb. Az SCA alkalmas elméletorientált kutatásokhoz is, mivel használható a meglévő elméletek tesztelésére (elsősorban deduktív kategóriafejlesztéssel), elméletek finomítására vagy akár új elméletek létrehozására (induktív megközelítéssel). Az értékelő SCA alkalmas olyan elméletorientált kutatásokra, ahol a cél a szövegek értékelése, besorolása valamilyen elméleti dimenzió mentén.

5.2. A minta

Az SCA különböző típusú és méretű mintákkal egyaránt alkalmazható. A módszer alkalmazható vezérfonal-központú, problémaközpontú, fókuszcsoportos és narratív interjúk elemzésére. Emellett kiválóan használható újságcikkek, hivatalos dokumentumok (pártprogramok, vállalati brosúrák), tudományos irodalom, tankönyvek és egyéb szöveges adatok feldolgozására is. Az SCA jellemzően előre meghatározott mintákkal dolgozik, ellentétben a Grounded Theory (GT) elméleti mintavételezésével, ahol az adatgyűjtés és az elemzés párhuzamosan, egymást alakítva zajlik. Az SCA esetében a szisztematikus mintavételi tervek lehetővé teszik kellően nagy csoportok összehasonlítását is. Nagyon nagy adatmennyiség (például több száz interjú) esetén lehetséges egy alminta kiválasztása kvótaalapú vagy véletlenszerű módszerrel.

5.3. A kutatás jellege

Akkor használjunk SCA-t, ha az alábbi feltételek valamelyike teljesül:

a) A cél a szöveges adatok szisztematikus, kategóriaalapú strukturálása és elemzése.
b) A kutatás célja a jelenség részletes leírása.
c) A teljes adatállományt szisztematikusan, azonos módon szeretnénk feldolgozni és kódolni.
d) A kutatás előre meghatározott mintával dolgozik, és fontosak a csoportok közötti összehasonlítások.
e) Elméletorientált kutatásban a válaszadók saját szavaikkal kifejtett, árnyalt véleményét szeretnénk elemezni ahelyett, hogy előre meghatározott válaszokkal dolgozó standardizált kérdőíveket használnánk.

Akkor NE használjunk SCA-t, ha az alábbi feltételek valamelyike teljesül:

a) A kutatás kizárólagos és elsődleges célja egy új, adatokban megalapozott elmélet (közepes hatókörű elmélet) kidolgozása. Bár az SCA is alkalmas elméletalkotásra, a Grounded Theory teljes módszertana erre specializálódott.
b) A kutatás folyamatorientált, rugalmas megközelítést igényel, ahol az adatgyűjtés és az elemzés szorosan összefonódik (elméleti mintavételezés).
c) A cél nem a teljes anyag szisztematikus kódolása, hanem az elméleti telítődés elérése – ilyenkor a Grounded Theory megfelelőbb választás, mivel ott az elemzés végpontja maga a telítődés, nem pedig a teljes anyag feldolgozása.

6. A módszer változatai és haladó alkalmazásai

Az SCA számos összetett elemzési lehetőséget kínál, amelyek az adatok mélyebb feltárását és az összefüggések megértését célozzák.

a) Kategóriák közötti bonyolult kapcsolatok elemzése. Az elemzés összpontosíthat több kategória együttes előfordulására, mintázatok és konfigurációk feltárására. Egy női életútkutatás keretében például megvizsgálhatjuk, hogy a különböző típusokba (például „Meggyökerezett", „Ingázó" stb.) tartozó nők hogyan vélekednek a családról, a karrierről vagy a jövőterveikről.
b) Strukturálatlan és strukturált adatok integrálása. Az SCA kiválóan alkalmas a strukturálatlan (kódolt szövegrészletek, összefoglalók) és strukturált (például gyakoriságok, háttérváltozók) adatok összekapcsolására. Miután a kutató létrehozza a tipológiát, felteheti a kérdést. „Van kapcsolat a típusokba való tartozás és a válaszadók szociodemográfiai jellemzői (például kor, iskolai végzettség, jövedelem) között?" Ekkor kereszttáblát hozhat létre, amelyben összeveti a típusokba tartozók arányát például az iskolai végzettségük szerint. Ezt követően akár statisztikai próbákat is végezhet, hogy megvizsgálja, a talált összefüggés statisztikailag szignifikáns-e.
c) Módszerváltozatok vegyítése. A különböző SCA-változatok hatékonyan összefűzhetők egyetlen kutatáson belül. Az észszerű és gyakori folyamat a következő. Először strukturáló SCA során a kutató azonosítja a jelenség legfontosabb témáit és dimenzióit. Ezekre a dimenziókra építve típusalkotó SCA keretében létrehozza az esettípusokat. Végül értékelő SCA segítségével értékelheti, hogy az egyes típusokon belül egy adott jelenség (például a változással szembeni attitűd) milyen mértékben van jelen.
d) Vizuális adatelemzés és bemutatás. A bonyolult kapcsolatok és tipológiák bemutatására olyan haladó megjelenítési megoldások állnak rendelkezésre, mint például a fogalmi térképek, mátrixok vagy kétdimenziós diagramok. Ezek az eszközök nemcsak az eredmények hatékony közlését segítik, hanem magát az elemzési folyamatot is, mivel láthatóvá teszik a rejtett mintázatokat.

7. Összefoglalás

A szemantikus tartalomelemzés (SCA) a strukturálatlan adatok elemzésének szisztematikus, szabálykövető és pragmatikus módszere. Ellentétben a mélyinterpretációra összpontosító eljárásokkal, az SCA a „small q"-paradigmában gyökerezik, ahol a kutatói szubjektivitás kontrollálandó tényező, a cél pedig a szövegben rejlő kimondott, manifeszt jelentéstartalom megbízható és átlátható rendszerezése. A módszer alapja a részletes kódkönyv és a következetes kódolási folyamat, amelynek minőségét a hitelesség, átvihetőség, konzisztencia és megerősíthetőség („trustworthiness") követelményei biztosítják. A kutatás folyamata jól szervezett szakaszokból áll, az adatokkal való ismerkedéstől a kategóriák fejlesztésén és a kódoláson át az elemzésig és az eredmények bemutatásáig. Az SCA-nak több, sajátos célú változata létezik, mint a strukturáló, az értékelő és a típusalkotó elemzés, amelyek lehetővé teszik a jelenségek leírását, értékelését és bonyolult tipológiák létrehozását is. Az SCA robusztus és sokoldalú eszköz, amely hidat képez a tisztán interpretatív és a strukturált adatokkal dolgozó modellező eljárások között.

8. Önellenőrző kérdések

1. Mi a legfőbb különbség az SCA és az RTA között a jelentés természetét és a kutatói szerepet illetően?

Válasz: A legfőbb különbség a fókuszban van. Az RTA a „Big Q"-paradigmába tartozik, és a rejtett jelentésrétegek mély, interpretatív feltárására törekszik; ebben a megközelítésben a kutató szubjektivitása értékes erőforrás. Ezzel szemben az SCA a „small q"-szemléletet követi, és elsődlegesen a felszíni, kimondott, manifeszt tartalom rendszerező és megbízható kódolására összpontosít. Itt a kutató inkább tárgyilagos elemző, akinek szubjektivitását kontrollálni kell.

2. Ismertesse a kvalitatív tartalomelemzés három alapvető típusát (Hsieh & Shannon alapján), és magyarázza el, miben különböznek!

Válasz: A három típus a következő. Az első a konvencionális (induktív) tartalomelemzés, amelyben a kategóriák közvetlenül az adatokból, előzetes elméletek nélkül jönnek létre. A második az irányított (deduktív) tartalomelemzés, amelyben egy meglévő elmélet vagy modell adja a kezdeti kategóriarendszert, amelyet a kutatás tesztel vagy kiterjeszt. A harmadik az összegző tartalomelemzés, amely a szavak előfordulási gyakoriságának számszerűsítésével kezdődik, majd a kontextus értelmezésével folytatódik. A legfőbb különbség a kódok és kategóriák eredetében rejlik – abban, hogy adatvezérelt vagy elméletvezérelt logika alapján születnek.

3. Mit jelent a „trustworthiness" fogalma a hermeneutikai kutatásban, és melyik négy kritériumból áll?

Válasz: A „trustworthiness" a hermeneutikai jellegű (különösen a „small q") kutatások minőségének és tudományos szigorának mércéje. A négy követelmény a következő: A hitelesség (credibility) arra utal, hogy az eredmények mennyire felelnek meg a valóságnak. Az átvihetőség (transferability) azt vizsgálja, hogy az eredmények mennyire alkalmazhatók más körülményekre. A konzisztencia (dependability) a kutatási folyamat stabilitására és reprodukálhatóságára vonatkozik. Végül a megerősíthetőség (confirmability) arra utal, hogy az eredmények mennyire az adatokból, és nem a kutató elfogultságából származnak.

4. Míg az RTA elveti, az SCA miért tartja fontosnak a kódolók közötti megbízhatóság ellenőrzését?

Válasz: Az SCA azért tartja ezt fontosnak, mert célja a kutatói szubjektivitás minimalizálása és olyan megbízható, következetes elemzés létrehozása, amely megismételhető. A konszenzuális kódolás biztosítja, hogy a kategóriák ne egyetlen kutató interpretációjától függjenek, hanem átlátható, szabályvezérelt folyamat eredményei legyenek, ami növeli a kutatás konzisztenciáját és megerősíthetőségét.

5. Milyen kutatási cél esetén választaná az SCA-t a megalapozott elmélet (Grounded Theory, GT) helyett?

Válasz: Akkor választanám az SCA-t, ha a célom egy előre meghatározott mintán alapuló szöveganyag szisztematikus leírása, tematikus strukturálása vagy egy meglévő elmélet tesztelése. Ezzel szemben a GT-t akkor választanám, ha a kizárólagos cél egy teljesen új, adatokban megalapozott elmélet felépítése olyan folyamatban, ahol az adatgyűjtés és az elemzés párhuzamosan, egymást alakítva (elméleti mintavételezéssel) zajlik.

Ajánlott irodalom

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

Hill, C. E., & Knox, S. (2021). *Essentials of consensual qualitative research*. American Psychological Association. https://doi.org/10.1037/0000215-000

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *The European Journal of General Practice, 24*(1), 120–124. https://doi.org/10.1080/13814788.2017.1375092

Krippendorff, K. (2019). *Content Analysis: An Introduction to Its Methodology*. SAGE. https://doi.org/10.4135/9781071878781

Kuckartz, U., & Rädiker, S. (2023). *Qualitative content analysis: Methods, practice and software*. SAGE.

Mayring, P. (2022). *Qualitative content analysis: A step-by-step guide*. SAGE.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook*. SAGE.

O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods, 19*, Article 1609406919899220. https://doi.org/10.1177/1609406919899220

Preiser, R., García, M. M., Hill, L., & Klein, L. (2021). Qualitative content analysis. In R. Biggs, R. Preiser, A. de Vos, M. Schlüter, K. Maciejewski, & H. Clements (Eds.), *The Routledge Handbook of Research Methods for Social-Ecological Systems* (pp. 270–281). Routledge. https://doi.org/10.4324/9781003021339-23

Silver, C., & Lewins, A. (2014). *Using Software in Qualitative Research: A Step-by-Step Guide*. SAGE. https://doi.org/10.4135/9781473906907

Stapley, E., O'Keeffe, S., & Midgley, N. (2022). Developing Typologies in Qualitative Research: The Use of Ideal-type Analysis. *International Journal of Qualitative Methods, 21*, Article 16094069221100633. https://doi.org/10.1177/16094069221100633

V. KVALITATÍV ÖSSZEHASONLÍTÓ ELEMZÉS (QCA)

TURÓS MÁTYÁS

A fejezet a következő kérdéseket válaszolja meg:

a) Mi a kvalitatív összehasonlító elemzés (QCA), és milyen kutatási problémák megoldására alkalmas?
b) Melyek a QCA alapvető logikai elvei?
c) Hogyan néz ki egy QCA-kutatás teljes folyamata a kutatási kérdés megfogalmazásától az eredmények értelmezéséig?
d) Mi a kalibráció, és melyek a logikai maradékok?
e) Hogyan azonosítjuk a szükséges és az elégséges feltételeket, és hogyan értelmezzük a kapott formulákat?
f) Milyen szoftveres támogatásra van szükség egy QCA-elemzés elvégzéséhez, és mi a kutató szerepe ebben a folyamatban?
g) Melyek a magas minőségű QCA-kutatás legfontosabb ismérvei?
h) Milyen haladóbb változatai és alkalmazási területei vannak a módszernek?

1. Bevezetés

A társadalomtudományi kutatások egyik központi feladata az összefüggések feltárása. Miért sikeres az egyik társadalmi mozgalom, míg a másik elbukik? Milyen tényezők vezetnek egy gazdasági reform bevezetéséhez? A hagyományos tudományos gondolkodás, amelynek alapjait John Stuart Mill fektette le, azt az okot keresi, amelynek jelenléte kiváltja a hatást, vagy amelynek hiánya a hatás elmaradásához vezet. Azonban ahogyan azt a kötet első fejezetében már említettük, a valóság rendkívül összetett, és gyakran több, egymással is kölcsönhatásban lévő tényező együttes jelenléte vagy hiánya szükséges az eredmény bekövetkezéséhez. A társadalmi jelenségeket jellemzően nem egyetlen, elkülönült ok magyarázza, hanem tényezők bonyolult összjátéka.

A Mill-féle módszerek alkalmazása – különösen eredeti, szigorú értelmezésükben – a társadalomtudományi gyakorlatban két fő probléma miatt nehézkes. Egyrészt ritkán találunk két olyan esetet, amelyek csupán egyetlen lényeges tényezőben térnek el egymástól, és ez megnehezíti az okok pontos azonosítását. Másrészt Mill módszerei nem alkalmasak az oksági komplexitás minden formájának kezelésére.

Bár a korszerű statisztikai eljárások (például a többszörös regresszió interakciós tagokkal) képesek több tényező együttes hatását vizsgálni, a Mill-féle logikára épülő hagyományos megközelítések jellemzően nehezen ragadják meg az olyan jelenségeket, ahol nem egyetlen tényező nettó hatása számít, hanem a tényezők összjátéka. Ráadásul e tényezők különböző utakon is elvezethetnek ugyanahhoz az eredményhez, és a pozitív és negatív kimenet okai aszimmetrikusak lehetnek.

Ezekre a kihívásokra válaszul fejlesztette ki Charles Ragin az 1980-as években a kvalitatív összehasonlító elemzést (Qualitative Comparative Analysis, QCA; a magyar nyelvű szakirodalomban „kvalitatív komparatív analízis” megnevezéssel is). A QCA ötvözi az esetalapú, mélyreható megközelítést a formális, szisztematikus elemzéssel. Alapvető újítása az, hogy a változók közötti korrelációk keresése helyett a Boole-algebra és a halmazelmélet logikájára épít olyan eljárást, amely képes a bonyolult, több tényezőből álló oksági konfigurációk azonosítására. Nem azt kérdezi, hogy a tényező milyen mértékben befolyásolja a kimenetet, hanem azt, hogy adott feltétel szükséges és/vagy elégséges-e az adott eredmény bekövetkezéséhez. A QCA így lehetővé teszi a logikailag helytálló következtetések levonását olyan helyzetekben is, ahol a hagyományos,

változóorientált módszerek nem a legalkalmasabbak az oksági komplexitás feltárására.

Képzeljünk el egy országot, ahol több politikai párt verseng a választásokon. A Mill-féle logika (és a hagyományos regressziós elemzés) azt kérdezné: „A karizmatikus vezető növeli a szavazatok számát?" vagy „A gazdasági helyzet javulása hozzájárul a kormánypárt sikeréhez?" (tényezők nettó hatása). A valóság azonban ennél összetettebb lehet. Lehet, hogy a párt nem akkor lesz sikeres, ha egyetlen tényező önmagában erős, hanem akkor, ha több tényező megfelelő együttese van jelen. Például a karizmatikus vezető ÉS a jól szervezett kampány ÉS a gazdasági növekedés ÉS az erős médiajelenlét együttese vezethet választási sikerhez. Ugyanakkor előfordulhat az is, hogy nincs karizmatikus vezető, de van erős ideológiai alap ÉS nagyon elégedetlen társadalom ÉS erős, egységes ellenzék. Két különböző konfiguráció vezethet ugyanahhoz az eredményhez, a választási sikerhez. A hagyományos statisztika, amely egy-egy tényező nettó hatását nézi (miközben a többit állandónak feltételezi), nehezen ragadja meg ezt a vagy-vagy logikát, ahol különböző utak vezethetnek ugyanahhoz az eredményhez. Ezt a jelenséget a QCA-szakirodalom ekvifinalitásnak nevezi.

Az is előfordulhat, hogy a választási siker okai nem ellentétei a választási kudarc okainak. A siker oka lehet a karizmatikus vezető ÉS a jó gazdaság ÉS a hatékony kampány együttese. A kudarc oka viszont nem feltétlenül ezen tényezők hiánya. Lehet, hogy a párt akkor bukik el, ha korrupciós botrány robban ki ÉS az ellenzék hiteles alternatívát kínál – függetlenül attól, hogy volt-e karizmatikus vezető, vagy sem. A győzelem és a vereség okai tehát nem tükrözik egymást; ezt a jelenséget a QCA oksági aszimmetriának nevezi, és a Mill-féle tudományos logika ezt szintén nehezen kezeli.

A regresszió azt feltételezi, hogy a különböző tényezők hatásai összeadódnak. Ha A hatása 5, B hatása 10, akkor A és B együtt 15-öt ad hozzá az eredményhez. Még az interakciós tagok is additív módon írják le, hogy két változó hatása hogyan módosítja egymást, de az alapvető szemlélet az egyes változók hatásának mérésére összpontosít. Ezzel szemben a QCA nem feltételezi az összeadhatóságot. Azt mondja, hogy néha A önmagában semmit nem ér, B önmagában semmit nem ér, de A ÉS B együtt minőségileg új hatást hoz létre, amely nem írható le A és B különálló hatásainak összegeként. Például az üzem bezárása önmagában nem elegendő a tiltakozás kirobbantásához, és önmagában a karizmatikus szónok sem, de ha az üzem bezárásának hírét karizmatikus szónok tálalja elégedetlen közösségnek, az vezethet tiltakozáshoz. Itt a két tényező együttes jelenléte az, ami számít, nem a nettó hatásuk.

A QCA tehát kifejezetten alkalmas arra, hogy azonosítsa azokat a sajátos feltételeket, amelyek szükségesek és/vagy elégségesek az eredmény bekövetkezéséhez, és külön azokat, amelyek nem.

A halmazelméleti módszereknek csak akkor van értelme, ha a kutatónak jó oka van azt gondolni, hogy az általa vizsgált jelenséget a halmazkapcsolatok szempontjából lehet a legjobban megérteni. A QCA akkor megfelelő eszköz, ha a kutatót a halmazok közötti kapcsolatok és nem a változók közötti korrelációk érdeklik. A módszer használata továbbá feltételezi, hogy a kutató elfogadja az oksági szempontból összetett magyarázatok keresését. Ha e módszert választja, a helyes gyakorlat megköveteli, hogy megmagyarázza, miért a halmazelméleti logika a legalkalmasabb a kutatási kérdése megválaszolására.

2. Elméleti alapok

A QCA a következő három elvre épül:

a) Konjunkturális okság (conjunctural causation). Ez az elv kimondja, hogy a feltételek nem önmagukban, elszigetelten fejtik ki hatásukat, hanem egymással kombinálódva, meghatározott konfigurációkban. Egy-egy tényező hatása más tényezők jelenlététől vagy hiányától függ. Ahogy a sütemény receptjénél sem a liszt vagy a cukor önmagában, hanem azok megfelelő aránya és a többi hozzávalóval való együttes jelenléte vezet a kívánt eredményhez, úgy a társadalmi jelenségeknél is a feltételek együttes fennállása számít.

b) Ekvifinalitás. Ez a fogalom arra utal, hogy ugyanazt az eredményt több, különböző oksági úton (konfiguráción) is el lehet érni. Nincs egyetlen, mindent megmagyarázó ok. Például a politikai párt választási győzelméhez vezethet a karizmatikus vezető és az erős médiatámogatás összjátéka, de vezethet a jól szervezett mozgósítási kampány és a népszerű program együttes megléte is.

c) Aszimmetria. Ez az elv azt jelenti, hogy az esemény bekövetkezésének és be nem következésének okai nem feltétlenül egymás tükörképei. Azok a feltételek, amelyek a sikerhez vezetnek, nem feltétlenül azonosak (csupán ellentétes előjellel) azokkal, amelyek a kudarcot magyarázzák. Például a

gazdasági növekedés egyik lehetséges receptje lehet a magas szintű innováció és a képzett munkaerő. A gazdasági válságot azonban nem feltétlenül csupán az innováció hiánya és a képzetlen munkaerő okozza, hanem más tényezők (például politikai instabilitás, külső piaci sokk) is szerepet játszhatnak benne. A siker hiánya nem egyenlő a kudarccal.

Ezen elvek alapján a QCA az oksági kapcsolatokat a szükséges és elégséges feltételek logikáján keresztül vizsgálja. A feltétel akkor szükséges a kimenetelhez, ha a kimenet nem következhet be a feltétel hiányában. Más szóval: „Nincs Y X nélkül." Halmazelméleti nyelven ez azt jelenti, hogy a kimenet (Y) halmaza a feltétel (X) halmazának részhalmaza. Például a magyar állampolgárság szükséges feltétele a magyarországi választásokon való szavazásnak, de önmagában nem elégséges, hiszen az állampolgárság önmagában még nem garantálja, hogy az illető szavaz is (ehhez például érvényes névjegyzékre, tényleges részvételre stb. is szükség van). A feltétel (vagy feltétel-összetétel, más néven konfiguráció) akkor elégséges a kimenetelhez, ha a jelenléte önmagában garantálja az eredmény bekövetkezését. Más szóval: „Ha X, akkor Y." Halmazelméleti nyelven ez azt jelenti, hogy a feltétel (X) halmaza a kimenet (Y) halmazának részhalmaza. A társadalomtudományokban ritkák az önmagukban elégséges feltételek; inkább a feltételek bizonyos konfigurációja az, ami elégséges egy adott kimenetelhez.

12. doboz: A szükséges és elégséges feltételek logikája a tűz keletkezésének analógiáján

Milyen feltételek vezetnek ahhoz, hogy tűz keletkezzen (Y), vagy ne keletkezzen (~Y) egy helyiségben?

a) Szükséges, de nem elégséges feltétel. Az oxigén jelenléte. Ahhoz, hogy tűz keletkezzen, mindenképpen kell oxigén (szükséges feltétel). Az oxigén jelenléte azonban önmagában nem elegendő – hiába van levegő a szobában, ha nincs éghető anyag vagy gyújtóforrás, nem lesz tűz. Az oxigén hiánya viszont garantáltan megakadályozza a tűz kialakulását.

b) Elégséges, de nem szükséges feltétel. A benzinnel leöntött papír és az égő gyufa összeállítása elégséges a tűz keletkezéséhez. Ha ez a feltétel teljesül, biztosan lesz tűz. De ez nem az egyetlen út a tűzhöz (nem szükséges) – ugyanúgy keletkezhet tűz elektromos zárlattól, gázrobbanástól vagy akár spontán öngyulladástól is. Tehát ez a konkrét útvonal garantálja a tüzet, de nélküle is lehet tűz.

c) Sem nem szükséges, sem nem elégséges feltétel. A magas hőmérséklet (például 50 °C a szobában) önmagában nem elegendő a tűzhöz, és nem is szükséges – hideg helyen is keletkezhet tűz. Mégis bizonyos körülmények között hozzájárulhat a tűz kialakulásához vagy terjedéséhez, például megkönnyítheti egyes anyagok meggyulladását.

d) Szükséges és elégséges feltétel. A tűzháromszög teljesülése, vagyis az éghető anyag ÉS az oxigén ÉS a gyújtási hőmérséklet együttes jelenléte. Ha mindhárom elem jelen van, mindig lesz tűz (elégséges). Ha bármelyik hiányzik, soha nem lesz tűz (szükséges). Ez a társadalomtudományokban ritka, tökéletes determinisztikus kapcsolat – a természettudományokban viszont gyakrabban találkozunk ilyen egyértelmű törvényszerűségekkel.

A társadalmi jelenségeknél tehát általában nem egyetlen feltétel, hanem különböző konfigurációk vezetnek ugyanahhoz az eredményhez (ekvifinalitás), ahogy a tűzhöz is többféle út vezet. A QCA-módszer ezeket az alternatív útvonalakat képes feltárni.

A QCA tehát esetorientált módszer, amely nem változókkal, hanem halmazokkal dolgozik. Megőrzi az esetek integritását, és azokat holisztikus konfigurációkként elemzi. A központi kérdés nem a „mennyire?", hanem a „tagja-e, vagy sem?". A kutatás tárgyát képező eseteket aszerint osztályozzuk, hogy tagok-e egy adott, a kutató által meghatározott halmazban. Az ország GDP-je például lehet folytonos változó, de a QCA-ban inkább a „gazdag országok" halmazában való tagságát vizsgáljuk. Ez a szemléletváltás teszi lehetővé, hogy a kvalitatív kategóriákat formális logikai elemzésnek vessük alá. A QCA kétféle halmaztípust különböztet meg.

a) Éles halmazok (crisp sets). A tagság egyértelmű és dichotóm. Az eset vagy teljes jogú tagja a halmaznak (tagsági érték: 1), vagy egyáltalán nem tagja (tagsági érték: 0). Nincsenek átmenetek. Például az ország vagy tagja az Európai Uniónak (1), vagy nem (0).
b) Életlen halmazok (fuzzy sets). A tagság foka 0 és 1 között változhat, lehetővé téve az árnyaltabb megkülönböztetést. Ez a megközelítés jobban illeszkedik a társadalomtudományi fogalmak természetéhez, amelyek viszonylag ritkán írhatók le egyszerű igen/nem kategóriákkal. Az ország demokráciájának szintje például nem dichotóm. A „demokratikus országok" halmazában az ország lehet teljes mértékben tag (1), teljes mértékben nem tag (0) vagy valamilyen mértékben tag (például 0,8 vagy 0,3). A fuzzy

halmazoknál lényeges a 0,5-ös átfordulási pont (crossover point), amely a tökéletes kétértelműség pontja – határvonal aközött, hogy az eset inkább tagja, vagy inkább nem tagja a halmaznak.

3. A kutatás folyamata

A kvalitatív összehasonlító elemzés folyamata nem feltétlenül lineáris, hanem gyakran iteratív, ami azt jelenti, hogy a kutató az elemzés során szerzett újabb belátások alapján visszatérhet korábbi lépéshez, hogy finomítsa a feltételeket, a kalibrációt vagy akár a vizsgált esetek körét. A folyamat lépései a következők.

3.1. A kutatási kérdésnek megfelelő esetek és feltételek kiválasztása

Példánkban azt vizsgáljuk, hogy milyen feltétel-összeállítások vezetnek úttörő vállalkozások kialakulásához európai városokban. Ez a kérdés azért megfelelő a QCA-hoz, mert három feltételnek is eleget tesz. Először bonyolult okságot feltételez, vagyis nem egyetlen tényező számít. Másodszor több út vezethet ugyanahhoz az eredményhez, tehát különböző városok különböző módon lehetnek sikeresek. Harmadszor közepes számú esetet vizsgálunk (a példában 20 európai várost).

Válasszunk tehát 20 európai várost, amelyek elég változatosak ahhoz, hogy különböző konfigurációkat képviseljenek. Például: Stockholm, Berlin, London, Párizs, Amszterdam, Barcelona, München, Milánó, Koppenhága, Dublin, Helsinki, Bécs, Zürich, Oslo, Lisszabon, Prága, Varsó, Budapest, Tallinn, Ljubljana. A szakirodalom vagy szakértői interjúk alapján azonosítsunk öt olyan lehetséges feltételt, amely az innováció mint kimenet létrejöttéhez vezethet.

a) EGYETEM. Erős technológiai egyetemek jelenléte.
b) TŐKE. Kockázati tőke elérhetősége.
c) KORMÁNY. Támogató kormányzati politikák.
d) KULTÚRA. Vállalkozói kultúra.
e) INFRASTRUKTÚRA. Fejlett digitális infrastruktúra.

A QCA-kutatásban tehát mindenekelőtt az elméleti háttér vagy külön kutatás alapján azonosítani kell azokat a lehetséges oksági tényezőket (QCA-terminológiával: feltételeket), amelyek lényegesek lehetnek az eredmény szempontjából. A feltételek számát továbbá mérsékelt szinten kell tartani (jellemzően 4–7 feltétel a megfelelő), mivel a túl sok feltétel diszfunkcionális; a korlátozott sokféleség[2] problémájához vezet. A lehetséges konfigurációk száma ugyanis exponenciálisan növekszik a feltételek számával (K feltétel esetén 2^K összetétel létezik). Ezért minél nagyobb a K, annál nagyobb az esély arra, hogy több összetételhez nem fog tartozni egyetlen valós eset sem. Ezeket a hiányzó eseteket nevezzük majd logikai maradékoknak.

A túl sok logikai maradék miatt a kapott megoldási formulák kétféle problémához vezethetnek. Egyrészt túlságosan bonyolulttá válhatnak, gyakran csak egy-egy esetre vonatkoznak, és elveszítik az általánosító erejüket. Másrészt a megoldás nagymértékben a logikai maradékokkal kapcsolatos feltételezéseken alapulhat, ami növeli a téves következtetések kockázatát. Ezért a feltételek számának csökkentése és finomítása a jó QCA szerves része, például elméletileg megalapozott, magasabb szintű makrokategóriák létrehozásával.

3.2. Adatgyűjtés és kalibráció

Ahogyan azt már említettük, a QCA halmazelméleti logikán alapul, nem korrelációkon. Míg a hagyományos statisztikai módszerek (például a regresszió) közvetlenül dolgoznak a nyers adatokkal és azok varianciájával, a QCA azt vizsgálja, hogy az esetek mennyire tagjai különböző halmazoknak. A QCA-ban három fő ok miatt szükséges a kalibráció, vagyis három ok miatt nem használhatók közvetlenül elemzésre a nyers adatok.

a) Különböző mérési skálák összemérhetetlensége. Példánkban az egyetemi rangsor pontszáma (0–100), a kockázati tőkebefektetés (VC, venture capital; millió €) és a kormányzati index (1–10 skála) nem hasonlíthatók össze közvetlenül. A kalibráció azonban mindegyiket 0–1 közötti halmaztagsági értékké alakítja majd.

[2] A kifejezés a QCA alapművéből származik, és a társadalomtudományi adatok ritkaságára, azaz a lehetséges feltételkombinációk hiányára utal. Lásd: Ragin, C. C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press.

b) Kvalitatív küszöbök jelentősége. A QCA-ban nem az számít, hogy egy város 2800 vagy 2900 millió € befektetést kap, hanem az, hogy átlépte-e azt a küszöböt, amely felett már magas befektetésű városnak számít. Vagy például a 45-ös és 50-es egyetemi pontszám között lehet, hogy nincs érdemi különbség, de a 75 és 85 között már van (az egyik még nem világszínvonalú, a másik már az).
c) Boole-algebrai műveletek követelménye. Az igazságtábla és a logikai minimalizálás (jelentésüket lásd később) csak akkor működnek, ha meg tudjuk mondani, hogy egy feltétel jelen van (1) vagy hiányzik (0). A nyers adatok ezt nem teszik lehetővé, ezért a kalibráció szükséges.

A kalibráció során tehát a nyers adatokat halmaztagsági értékekké alakítjuk. Ehhez először érdemes elkészíteni a nyers adatok táblázatát (meglévő adatbázisból vagy saját adatgyűjtésből). A példában ez a következőképpen néz ki.

Város	Egyetemi rangsor pontszám	VC befektetés (millió €)	Kormányzati támogatás index	Vállalkozói kultúra pontszám	Digitális infrastruktúra index	Vállalkozások száma
Stockholm	85	1200	8,5	7,8	9,2	4
Berlin	78	2800	7,2	8,5	8,7	6
Budapest	45	150	5,5	5,2	7,1	0
London	96	3500	7,0	9,4	9,1	12
Prága	42	250	5,8	6,1	7,2	0
…	…	…	…	…	…	…

10. táblázat: A vizsgált európai városok nyers adatai (részlet)

A kalibrációnál használhatunk például elméleti küszöbértékeket szakirodalom alapján, természetes töréspontokat (például EU-átlag alapján), kvalitatív ismereteket az esetekről vagy szakértői véleményeket. Példánkban a nyers adatok halmaztagsági értékké alakítása elméleti tudáson és külső szabványokon alapul. Ezek alapján három horgonypontot határozunk meg minden feltételhez és kimenetelhez.

a) A teljes tagság küszöbe (0,95). Az a nyers adatérték, amely felett az esetet már teljes mértékben a halmaz tagjának tekintjük.

b) Átfordulási pont (0,5; crossover point). A tökéletes kétértelműség pontja, vagyis az az érték, amelynél az eset éppúgy tagja a halmaznak, mint amennyire nem.
c) A teljes nem tagság küszöbe (0,05). Az a nyers adatérték, amely alatt az esetet már egyértelműen nem tekintjük a halmaz tagjának.

Nézzük meg ezt a gyakorlatban a példánk EGYETEM feltételének kalibrációján keresztül. A számszerűen is fiktív példában a horgonypontokat az európai egyetemi rangsorok elemzése alapján határoztuk meg. A 80 pont a nemzetközi technológiai rangsorokban a top 100-as helyezésnek felel meg, amely jellemzően a világszínvonalú kutatóegyetemek küszöbe. A 60 pont az európai technológiai egyetemek mediánja. A 40 pont alatt pedig már statisztikailag szignifikánsan kevesebb spin-off vállalat és szabadalom születik, ami a gyenge innovációs teljesítmény indikátora. E három érték alapján a horgonypontokat a következőképpen határozzuk meg.

a) 0,95 (teljesen bent). 80+ pont, vagyis világszínvonalú technológiai egyetemek.
b) 0,5 (átfordulás). 60 pont, vagyis közepes minőségű felsőoktatás. (A 60 pont nem a minta mediánja, hanem az európai egyetemek rangsorának mediánja, azaz külső, elméleti szempont alapján meghatározott érték.)
c) 0,05 (teljesen kint). 40 pont, vagyis gyenge technológiai oktatás.

A három horgonypont meghatározása után ezeket az értékeket sigmoid (logisztikus) függvény kalibrálására használjuk. Ez a függvény biztosítja a sima átmenetet a teljes nem tagság (0) és a teljes tagság (1) között ahelyett, hogy éles határvonalakat húznánk. A sigmoid függvény automatikusan áthalad a három horgonyponton. 40 pontnál közel 0,05 értéket ad, 60 pontnál pontosan 0,5-öt (ez a crossover point definíciója), 80 pontnál pedig közel 0,95-öt. A halmaztagsági értéket a következő képlettel számoljuk ki:

$$\text{halmaztagsági érték} = 1 / (1 + \exp(-k \cdot (x - c)))$$

A képletben x a nyers adat, c az átfordulási pont (esetünkben 60), k pedig a meredekség.

13. doboz: A fuzzy halmazok kalibrációja és a sigmoid függvény paraméterei

A képletben szereplő k paraméter, vagyis a meredekség azt szabályozza, hogy a függvény milyen meredeken vagy fokozatosan tér át a teljes nem tagságból (0) a teljes tagságba (1) az átfordulási pont (c) körül. A szerepe a fényerőszabályzó analógiájával érthető meg a legkönnyebben. A magas k érték (nagy meredekség) olyan, mint az egyszerű villanykapcsoló. A változás nagyon hirtelen történik az átfordulási pont körül. Az esetek többsége vagy egyértelműen „bent" (közel 1), vagy egyértelműen „kint" (közel 0) lesz, és csak nagyon kevés eset kap köztes értéket. Ezt akkor használjuk, ha a fogalmunk határai elméletileg élesek. Az alacsony k érték (kis meredekség) olyan, mint a finoman szabályozható dimmer. Az átmenet sokkal lassabb, fokozatosabb. Több eset fog köztes, fuzzy értéket kapni (például 0,3; 0,4; 0,6), amely azt jelzi, hogy a fogalom határai elmosódottak. Fontos, hogy a k értékét – a horgonypontokhoz hasonlóan – nem magukból az adatokból számoljuk. A legtöbb QCA-szoftver rendelkezik alapértelmezett beállítással, amely közepesen meredek függvényt eredményez, és a legtöbb társadalomtudományi elemzéshez ez tökéletesen megfelel. Hacsak a kutatónak nincs erős elméleti indoka különösen éles vagy különösen elmosódott határvonal feltételezésére, az alapértelmezett érték használata a bevett gyakorlat.

Például Stockholm 85-ös pontszáma jóval a teljes tagságot jelentő 80 pont felett van, így a kalibrált értéke magas, 0,92 lesz. Ezzel szemben Budapest 45-ös pontszáma a 40 pontos „teljesen kint" és a 60 pontos „átfordulási pont" között helyezkedik el, de jóval közelebb az előbbihez, így a kalibrált értéke alacsony, 0,15 lesz. Miután ezt a logikát minden feltételre és a kimenetelre is alkalmaztuk az összes város esetében, megkapjuk a kalibrált adattáblát.

Város	EGYETEM	TŐKE	KORMÁNY	KULTÚRA	INFRA	INNOVÁCIÓ
Stockholm	0,92	0,88	0,91	0,85	0,95	0,89
Berlin	0,81	0,95	0,78	0,92	0,88	0,94
Budapest	0,15	0,22	0,31	0,28	0,65	0,08
London	0,96	0,98	0,72	0,94	0,91	0,97
Prága	0,12	0,25	0,58	0,61	0,72	0,08
...	...	...	...	...	...	...

11. táblázat: A vizsgált városok kalibrált halmaztagsági értékei (részlet)

Megjegyzés. A táblázat a folytonos változók életlen (fuzzy) halmazokká kalibrálását mutatja be. A modell azonban eredendően dichotóm változókat is kezelhet (például a város tagsága valamely szervezetben), amelyeket éles halmazokként (crisp sets) értelmez. Esetükben a kalibráció az 1-es (igen) és 0-s (nem) értékek egyszerű hozzárendelését jelenti.

A fenti kalibrációs eljárást minden feltételre és a kimenetelre is el kell végezni, de nem feltétlenül ugyanazokkal a paraméterekkel. Minden változóhoz egyedileg kell meghatározni a három horgonypontot az adott változó elméleti és empirikus sajátosságai alapján. Például a VC-nél más küszöbértékek lesznek relevánsak (mondjuk 500, 1500, 3000 millió €), mint az egyetemi rangsoroknál. A sigmoid függvény alkalmazását a QCA-szoftverek (például fsQCA, az R QCA-csomagja, TOSMANA) automatikusan elvégzik, miután megadtuk a három horgonypontot. Kézzel számolni nem szükséges és nem is célszerű. A kutató feladata csak a horgonypontok megalapozott meghatározása, a matematikai transzformációt a szoftver végzi.

A crisp QCA (csQCA) esetén továbbá nincs is szükség ilyen bonyolult kalibrációra. Ilyenkor egyszerűen küszöbértéket határozunk meg (például 60 pont felett erős egyetem = 1, alatta gyenge = 0), és dichotóm változókkal dolgozunk. Ha eleve bináris adataink vannak (például szakértői interjúkból kódolt igen/nem válaszok), akkor a kalibráció triviális, hiszen az „igen" értéke 1, a „nem" értéke 0.

A fuzzy kalibráció három esetben szükséges. Először, amikor folytonos változóink vannak (mint a példánkban). Másodszor, amikor nem akarunk információt veszíteni az éles 0/1 dichotomizálással. Harmadszor, amikor az elméleti fogalmaink maguk is fokozatiak (például az „innovatív város" nem fekete-fehér kategória).

A táblázatban az is látható, hogy a kalibráción nemcsak a feltételváltozóknak kell átesnie, hanem annak a tényezőnek is, amelyet magyarázni szeretnénk (példánkban az INNOVÁCIÓ).

3.3. A szükséges feltételek elemzése

Miután a nyers adatainkat sikeresen átalakítottuk halmaztagsági értékekké, a szükséges feltételeket kell azonosítani. A QCA-elemzés alapszabálya, hogy a szükséges és az elégséges feltételek elemzését külön, egymás után kell elvégezni, és mindig a szükségességvizsgálat az első. Ennek oka, hogy ha létezik olyan alapvető feltétel, amely nélkül az eredmény szinte soha nem fordul elő, azt érdemes előre azonosítani, mielőtt az összetett konfigurációkat vizsgálnánk. Emellett a szükségességvizsgálat kihagyása ahhoz vezethet, hogy egy valójában szükséges feltétel „elrejtőzik" az elégséges konfigurációk között, és ezáltal hibás következtetésekhez jut a kutató.

Emlékeztetőül, a feltétel (X) akkor szükséges a kimenetelhez (Y), ha a kimenetel nem következhet be a feltétel hiányában, vagyis nincs Y X nélkül. A példában most azt vizsgáljuk, hogy a sikeres vállalkozáshoz (INNOVÁCIÓ)

szükséges-e például az erős egyetemi háttér, a bőséges tőke vagy a támogató vállalkozói kultúra önmagában.

A szükségességvizsgálat során három mutatót számítunk ki minden lehetséges feltételhez, amelyek együtt adnak képet arról, hogy az adott feltétel valóban szükséges-e a kimenetelhez.

a) A konzisztencia azt méri, hogy a részhalmazkapcsolat mennyire tökéletesen teljesül, vagyis az esetek mekkora hányadánál van jelen a feltétel akkor, amikor a kimenetel is jelen van.
b) A lefedettség azt méri, hogy a feltétel mennyire fedi le a kimenetelt. Segít kiszűrni a triviális feltételeket, például azokat, amelyek szinte minden esetben jelen vannak.
c) A relevancia szintén a trivialitás elkerülését segítő mutató, amely a konzisztenciát és a lefedettséget együttesen veszi figyelembe.

Mindhárom mutatót a kalibrált halmaztagsági értékekből (11. táblázat) számítjuk, és a QCA-szoftverek ezt automatikusan elvégzik. A logika megértéséhez azonban érdemes részletesen bemutatni legalább a konzisztencia képletét, a másik kettő hasonló elven működik, csak bonyolultabb. A konzisztenciát a következőképpen számoljuk.

$$\text{konzisztencia} = \Sigma(\min(X_i, Y_i)) / \Sigma(Y_i)$$

A képletben X_i az adott feltétel halmaztagsági értéke az i-edik esetnél, Y_i a kimenetel (INNOVÁCIÓ) halmaztagsági értéke az i-edik esetnél, $\min(X_i, Y_i)$ pedig a kettő közül a kisebb érték. A szummázás az összes esetre (városra) vonatkozik.

Például a KULTÚRA szükségességi konzisztenciáját a következőképpen számítjuk ki. Minden városnál vesszük a KULTÚRA és az INNOVÁCIÓ értékei közül a kisebbet, ezeket összeadjuk, majd elosztjuk az INNOVÁCIÓ értékeinek összegével. Az eredmény 0,91, amely azt jelenti, hogy az innovatív városok 91%-ban egyben magas kultúraértékkel is rendelkeznek. A következő táblázat mutatja a három mutató eredményeit minden feltételre.

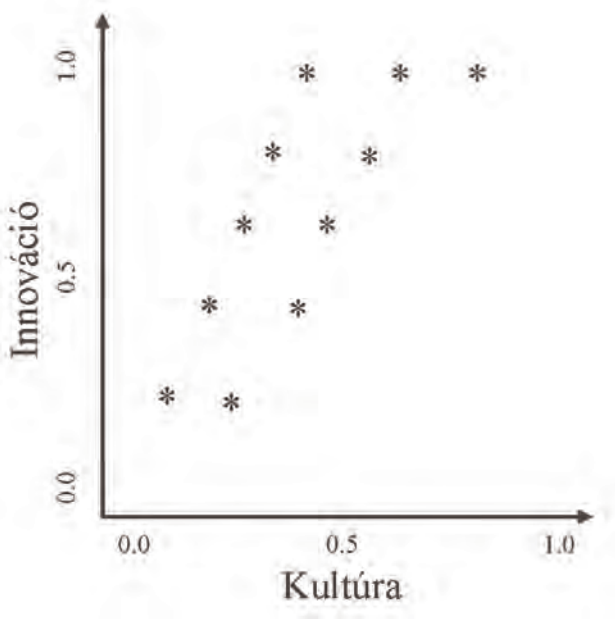
Feltétel	Konzisztencia	Lefedettség	Relevancia
EGYETEM	0,78	0,82	0,65
TŐKE	0,85	0,79	0,71
KORMÁNY	0,69	0,74	0,58
KULTÚRA	0,91	0,86	0,78
INFRA	0,74	0,71	0,62

12. táblázat: Az egyes feltételek szükségességvizsgálatának eredményei

A szakirodalmi gyakorlat szerint a feltételt akkor tekintjük potenciálisan szükségesnek, ha a konzisztencia értéke 0,9 felett van. Az 1,0 érték azt jelentené, hogy minden sikeres ökoszisztéma esetében jelen van az adott feltétel. A táblázatban egyértelműen látszik, hogy a KULTÚRA konzisztenciaértéke 0,91, amely meghaladja ezt a küszöbértéket. Ez azt jelenti, hogy a sikeres vállalkozásokkal rendelkező városok szinte kivétel nélkül erős tagnak számítanak a „fejlett vállalkozói kultúra" halmazában is. Ezzel szemben a többi feltétel (például a KORMÁNY 0,69-es értéke) messze elmarad ettől.

A konzisztencia azonban önmagában nem elegendő, a lefedettség és a relevancia értékei azt is megmutatják, hogy a talált szükséges feltétel nem triviális. Az olyan feltétel, amely szinte minden esetben jelen van (például „a város Európában van"), lehet szükséges, de nem informatív. A KULTÚRA 0,86-os lefedettségi értéke azt mutatja, hogy ez a feltétel releváns, nem triviális. A szakirodalom nem határoz meg szigorú küszöbértéket a lefedettségre, de általános iránymutatásként a 0,25 feletti értékeket már informatívnak tekinthetjük. Az 1,0-hez közeli értékek, különösen alacsony relevancia mellett, felvethetik a feltétel trivialitásának gyanúját.

A relevancia esetében a magasabb érték a kedvezőbb, mivel az jelzi, hogy a feltétel nem triviálisan szükséges a kimenetelhez. Erre sincs szigorú küszöb, de a 0,6 feletti értékeket általában már informatívnak tekintjük. A KULTÚRA 0,78-as relevanciaértéke itt is a feltétel érdemi fontosságát támasztja alá. Az eredményeket grafikonon is ellenőrizhetjük, hogy képszerűen is lássuk a részhalmazkapcsolatot.



8. ábra: XY-plot megjelenítése a szükségesség vizsgálatához

A grafikont a szükségesség logikája alapján kell értelmezni: „Nincs kimenetel (Y) a feltétel (X) nélkül." Ez azt jelenti, hogy nem létezhet olyan eset, ahol a kimenetel (INNOVÁCIÓ) mértéke magas, miközben a feltétel (KULTÚRA) mértéke alacsony. Vizuálisan ez azt jelenti, hogy a grafikon bal felső sarkának üresnek kell lennie. A fenti ábrán pontosan ezt látjuk, nincsenek pontok (városok) a bal felső területen. Az összes eset a fő átló (a bal alsó sarokból a jobb felső sarokba tartó képzeletbeli vonal) fölött vagy ahhoz közel helyezkedik el. Ez képszerűen is megerősíti, hogy az INNOVÁCIÓ halmaz a KULTÚRA halmaz részhalmaza.

Összegezve, az elemzés harmadik lépésében arra jutottunk, hogy a vállalkozói kultúra (KULTÚRA) kváziszükséges feltétele[3] a sikeres vállalkozás kialakulásának. Bár a kapcsolat nem tökéletes (hiszen a konzisztencia nem 1,0), nagyon erős. A KULTÚRA kváziszükségességének azonosítása, más elemzésekben

[3] A QCA módszertanában a feltételt akkor tekintjük kváziszükségesnek, ha annak konzisztenciamutatója eléri vagy meghaladja a szakirodalomban elvárt 0,90-es küszöbértéket. Ez azt jelzi, hogy a kimenetel (siker) szinte soha nem fordul elő az adott feltétel nélkül, de a modell megenged néhány kivételt (deviáns esetet). Lásd: Schneider, C. Q., & Wagemann, C. (2012). *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis.* Cambridge University Press.

természetesen más feltételé vagy feltételeké, három szempontból is lényeges a további elemzés számára.

a) Az igazságtábla értelmezése. Tudva, hogy a KULTÚRA kváziszükséges, nem meglepő, ha az igazságtáblában nem találunk majd olyan sikeres konfigurációt, ahol a KULTÚRA hiányzik (~KULTÚRA). Ha mégis találnánk ilyet, az deviáns esetnek számítana, amelyet külön meg kellene vizsgálni.
b) A megoldási formulák egyszerűsítése. Mivel a KULTÚRA szinte mindig jelen van a sikeres esetekben, minden sikeres útvonalban meg fog jelenni. Ez egyszerűsíti az értelmezést, hiszen nem kell olyan alternatív utakat keresnünk, ahol a kultúra hiányát más tényezők ellentételeznék.
c) Gyakorlati következtetések. A döntéshozók számára ez azt jelenti, hogy a vállalkozói kultúra fejlesztése nem tetszőleges, hanem alapfeltétel. Hiába van erős egyetem vagy sok tőke, kultúra nélkül a siker valószínűtlen.

Ha nem találtunk volna szükséges feltételt, akkor az összes konfiguráció egyformán esélyes lenne a sikerre. Most viszont tudjuk, hogy a KULTÚRA = 0 összetételeket már eleve kisebb eséllyel fogjuk sikeresnek találni. Így lépünk tovább a negyedik lépésre, ahol a bonyolult, többtényezős konfigurációkat vizsgáljuk.

3.4. Az igazságtábla létrehozása és elemzése

Miután azonosítottuk a lehetséges szükséges feltételeket, áttérünk a QCA-elemzés központi elemére, a feltétel-összetételek, azaz a konfigurációk vizsgálatára. Nem azt kérdezzük többé, hogy egyetlen tényező önmagában elég-e a sikerhez, hanem azt, hogy a feltételek milyen konfigurációi vezetnek el a kívánt eredményhez. Ennek az elemzésnek a legfontosabb eszköze az igazságtábla.

Az igazságtábla olyan táblázat, amely a kutatásunkban szereplő feltételek összes lehetséges kombinációját felsorolja. Mivel öt feltételünk van (EGYETEM, TŐKE, KORMÁNY, KULTÚRA, INFRASTRUKTÚRA), és minden feltételnek két állapota lehet (jelen van vagy hiányzik), a lehetséges összetételek száma $2^5 = 32$. Az igazságtábla minden sora ilyen egyedi összetételt képvisel.

Fontos megjegyezni, hogy a konzisztencia fogalma itt más logikát követ, mint a 3.3. fejezetben. Ott a szükségességi konzisztenciát számítottuk, amely

azt méri, hogy a feltétel halmaza tartalmazza-e a kimenetel halmazát ($Y \subseteq X$). Az igazságtáblában ezzel szemben az elégségességi konzisztenciát használjuk, amely fordított irányú részhalmazkapcsolatra kérdez rá. Mennyire igaz, hogy a feltételkonfiguráció (X) halmaza a kimenetel (Y) halmazának része, vagyis hogy az X-ek mindig Y-hoz vezetnek?

Az igazságtáblában a feltétel jelenlétét (ha a kalibrált értéke > 0,5) 1-gyel, a hiányát (ha a kalibrált értéke < 0,5) pedig 0-val jelöljük. A pontosan 0,5-ös értéket mutató, maximálisan kétértelmű eseteket a konzisztencia számításából kizárjuk. Ezután minden városunkat (esetünket) besoroljuk a táblázat megfelelő sorába a kalibrált értékeik alapján. Például az olyan város, ahol az egyetemek és a tőke erősek (1, 1), a kormányzati támogatás gyenge (0), a kultúra és az infrastruktúra pedig erős (1, 1), az 11011 sorba kerül. Minden sorhoz rögzítjük, hogy az oda tartozó városok jellemzően mutatják-e a kimenetelt (sikeres innováció).

Sor	EGYETEM	TŐKE	KORMÁNY	KULTÚRA	INFRA	Esetek száma	Kimenet	Nyers konzisztencia	PRI konzisztencia
1	1	1	1	1	1	3	1	0,95	0,93
2	1	1	0	1	1	2	1	0,92	0,89
3	0	1	1	1	1	1	1	0,88	0,85
4	1	0	1	1	1	1	0	0,72	0,65
5	0	0	0	1	1	2	0	0,45	0,38
6	0	0	0	0	1	3	0	0,31	0,25
...	...	...	...	...	...	...	...	...	...
32	0	0	0	0	0	0	?	-	-

13. táblázat: Az elemzés igazságtáblája a megfigyelt konfigurációkkal (részlet)

Megjegyzés. A 32. sorban azért van kérdőjel a kimenetnél és gondolatjel a konzisztenciáknál, mert ez a sor logikai maradék. A legfontosabb adat ebben a sorban az „esetek száma" oszlopban található. Az értéke 0. Ez azt jelenti, hogy a mintában (a vizsgált városok között) egyetlenegy olyan eset sincs, amely a 32. sor konfigurációját mutatná (azaz ahol mind az öt feltétel hiányzik: 0 0 0 0 0). A konzisztencia azt méri, hogy a sorhoz tartozó valós esetek milyen megbízhatóan vezetnek a kimenetelhez. Mivel itt nincs egyetlen valós eset sem, a konzisztenciát nem lehet kiszámítani. A kimenet oszlop értékét (1 vagy 0) a konzisztencia alapján határozzuk meg; mivel a konzisztenciát nem tudtuk kiszámítani, a kimenet értékét sem tudjuk empirikus adatok alapján eldönteni. A kérdőjel azt jelzi, hogy ez az összetétel empirikusan nem megfigyelt, ismeretlen kimenetelű útvonal.

14. doboz: A fuzzy értékek szerepe és kezelése az igazságtáblában

Felmerülhet a kérdés, hogy miért használunk fuzzy értékeket, ha az igazságtáblában 0 vagy 1 lesz. Bár az igazságtáblába való besoroláskor dichotomizálunk (>0,5 = 1, <0,5 = 0), ez nem ugyanaz, mint a crisp QCA igen/nem felosztása. A fuzzy kalibráció során már figyelembe vettük a kvalitatív küszöböket. Például a 0,51-es és a 0,95-ös érték egyaránt 1-es lesz az igazságtáblában, de amikor azt számoljuk, hogy az adott konfiguráció mennyire vezet megbízhatóan a kimenetelhez, nem 0 és 1 értékekkel dolgozunk, hanem az eredeti fuzzy értékekkel. Az elégségességi konzisztencia képlete a következő:

nyers konzisztencia = $\Sigma(\min(X_i, Y_i)) / \Sigma(X_i)$

A képletben X_i és Y_i a fuzzy halmaztagsági értékek. A nevezőben most $\Sigma(X_i)$ szerepel, szemben a 3.3. fejezet szükségességi konzisztenciájának $\Sigma(Y_i)$ nevezőjével – ez tükrözi az ellentétes irányú részhalmazkapcsolatot. Így a 0,95-ös egyetemi érték jobban hozzájárul a konzisztenciához, mint a 0,51-es. Ezenkívül a megoldási útvonalak lefedettségének számítása szintén a fuzzy értékeket használja, nem a dichotomizált változatot. A város lehet „kissé tag" sikeres konfigurációban (például 0,7-es értékkel), amely informatívabb, mint az egyszerű „tag" vagy „nem tag" besorolás. Tehát az igazságtábla 0/1 értékei csak a sorok azonosítására szolgálnak (melyik konfigurációról van szó), de az elemzés számításai továbbra is a finomabb fuzzy értékekkel dolgoznak. Ez a crisp és a fuzzy QCA közötti különbség, és talán a fuzzy QCA előnye is – megőrizzük a Boole-algebra egyszerűségét, miközben nem veszítjük el a fokozatosság információtartalmát.

Ahhoz, hogy eldöntsük, mely összetételek (sorok) vezetnek megbízhatóan a sikerhez, a következő lépésben két küszöbértéket kell meghatározni.

a) Frekvenciaküszöb. Meghatározzuk, hogy az összetételhez legalább hány valós esetnek kell tartoznia ahhoz, hogy érdemben foglalkozzunk vele. Mivel a mintánk kicsi, ezt a küszöböt alacsonyan tartjuk; a példánkban legalább 1 eset. Ez kiszűri azokat a sorokat, amelyekhez egyetlen város sem tartozik.
b) Konzisztenciaküszöb. A nyers konzisztencia azt méri, hogy az adott sorba tartozó esetek milyen megbízhatóan mutatják a kimenetelt. Ez az elégségesség mértéke. A gyakorlatban általában 0,80 feletti értéket várunk el. Ha a sor konzisztenciája eléri ezt a küszöböt, akkor a Kimenet oszlopban 1-est kap, jelezve, hogy ez a konfiguráció sikeres útvonal. Ha a konzisztencia alacsonyabb, 0-t kap. A „Proportional Reduction in Inconsistency"

(PRI) haladóbb mutató. Azt segít kiszűrni, amikor a konfiguráció ellentmondásos, azaz nemcsak a kimenetelhez (sikerhez), hanem annak tagadásához (kudarchoz) is erősen vezetne. Pontosabban a PRI a szimultán részhalmazkapcsolatok problémáját kezeli: előfordulhat, hogy ugyanaz a konfiguráció egyszerre részhalmaza mind az Y-nak (sikeres esetek), mind a ~Y-nak (sikertelen esetek), ami logikai ellentmondást jelent. A magas PRI-érték (általában 0,75 felett) azt jelzi, hogy a konfiguráció egyértelműen inkább az eredményhez, mint annak hiányához vezet.

A 13. táblázatban látható, hogy az első három sor konzisztenciája meghaladja a 0,80-as küszöböt, így ezek sikeres utakhoz vezető konfigurációk. A 4. és 5. sor nem. Az igazságtáblánkból az is kiderül, hogy a 32 elméletileg lehetséges összetételből csak 12-höz tartozik valós eset. A maradék 20 sor, amelyhez a mintánkban egyetlen város sem tartozik, logikai maradék. Ezek a valóságunkban (még) nem megfigyelt összetételek. Kezelésük a QCA egyik legfontosabb kihívása, mert a végső megoldás egyszerűsítése során felhasználhatjuk őket. Kezelésükre három stratégia létezik.

a) Komplex megoldás. A legkonzervatívabb megközelítés. Nem tesz semmilyen feltételezést a logikai maradékokról, kizárólag a valós, megfigyelt eseteken alapuló utakat veszi figyelembe. Az eredmény pontos, de gyakran túlságosan összetett.
b) Parszimónikus (takarékos) megoldás. A legengedékenyebb megközelítés. Bármilyen logikai maradékot bevonhat az elemzésbe, ha az segít a formula egyszerűsítésében. Az eredmény a lehető legegyszerűbb, de fennáll a veszélye, hogy olyan feltételezéseket tesz, amelyek elméletileg nem plauzibilisek (például azt sugallja, hogy a gyenge egyetemek ÉS a tőke hiánya vezet sikerhez).
c) Köztes megoldás. Az arany középút, a leggyakrabban használt stratégia. Itt a kutató az elméleti tudása alapján tesz észszerű feltételezéseket. Meghatározzuk, hogy az egyes feltételek (például erős egyetemek, tőke) jelenléte vagy hiánya járul-e hozzá a sikerhez. Ezeket nevezzük irányított elméleti elvárásoknak. Minden feltételről előzetesen eldöntjük, hogy elméletileg a jelenléte (1) vagy a hiánya (0) várható-e a sikeres kimenetelhez vezető konfigurációkban. Például az „erős egyetemek" esetén azt várjuk, hogy a jelenléte (nem pedig a hiánya) járul hozzá az innovációhoz.

A példában a köztes megoldást fogjuk követni, mivel ez biztosítja a legjobb egyensúlyt az empirikus megalapozottság és az elméleti koherencia között.

3.5. Logikai minimalizálás

Miután az igazságtáblában azonosítottuk a sikerhez vezető konfigurációkat (azokat a sorokat, ahol a kimenet értéke 1), a feladatunk az, hogy ezeket a konfigurációkat a lehető legegyszerűbb formájukra hozzuk. Ezt a folyamatot nevezzük logikai minimalizálásnak. A minimalizálás célja, hogy a hosszú, sajátos konfigurációkból kiszűrjük a felesleges, irreleváns feltételeket, és így megtaláljuk a sikerhez vezető utak lényegi magját. Hosszú lista helyett néhány tömör, általánosabb érvényű konfigurációt szeretnénk kapni.

A minimalizálás a Boole-algebra egy egyszerű elvén alapul, amit a Quine–McCluskey-algoritmus használ. Az elv a következő: ha találunk két sikerhez vezető konfigurációt, amelyek mindenben megegyeznek, kivéve egyetlen feltételt (az egyikben ez jelen van [1], a másikban hiányzik [0]), akkor ez a feltétel az adott kontextusban logikailag redundáns, és elhagyható a közös formulából. Például ha (EGYETEM * TŐKE * KORMÁNY * KULTÚRA * INFRA) is sikeres, és (EGYETEM * TŐKE * ~KORMÁNY * KULTÚRA * INFRA) is sikeres, akkor a KORMÁNY megléte vagy hiánya nem számít ebben a kontextusban. A két út egyszerűsített formája tehát EGYETEM * TŐKE * KULTÚRA * INFRA.

A minimalizálás folyamatát a példánkon keresztül mutatjuk be. Vegyük a három, sikeresnek azonosított konfigurációt az igazságtáblánkból, és végezzük el rajtuk a minimalizálást lépésről lépésre. (A * jel a logikai ÉS kapcsolatot jelöli, a ~ jel pedig a feltétel hiányát.)

a) Kiindulási konfigurációk (ahol a kimenet = 1):

Siker 1: EGYETEM * TŐKE * KORMÁNY * KULTÚRA * INFRA
Siker 2: EGYETEM * TŐKE * ~KORMÁNY * KULTÚRA * INFRA
Siker 3: ~EGYETEM * TŐKE * KORMÁNY * KULTÚRA * INFRA

b) Első egyszerűsítési kör: párok összehasonlítása (páronként összevetjük a fenti sorokat, és keressük azokat, amelyek csak egyetlen feltételben térnek el):

Siker 1 kontra Siker 2: EGYETEM * TŐKE * KORMÁNY * KULTÚRA * INFRA kontra EGYETEM * TŐKE * ~KORMÁNY * KULTÚRA * INFRA. Látható, hogy a két sor mindenben azonos, kivéve a KORMÁNY feltételt. Az alapelv szerint tehát a KORMÁNY jelenléte vagy hiánya itt irreleváns. A két sort összevonhatjuk, és a KORMÁNY feltételt elhagyhatjuk. Eredmény (E1): EGYETEM * TŐKE * KULTÚRA * INFRA

Siker 1 kontra Siker 3: EGYETEM * TŐKE * KORMÁNY * KULTÚRA * INFRA kontra ~EGYETEM * TŐKE * KORMÁNY * KULTÚRA * INFRA. Ez a két sor is mindenben azonos, kivéve az EGYETEM feltételt. Az alapelv szerint az EGYETEM itt redundáns. Eredmény (E2): TŐKE * KORMÁNY * KULTÚRA * INFRA

Siker 2 kontra Siker 3: EGYETEM * TŐKE * ~KORMÁNY * KULTÚRA * INFRA kontra ~EGYETEM * TŐKE * KORMÁNY * KULTÚRA * INFRA. Ez a két sor két feltételben eltér (EGYETEM és KORMÁNY), ezért őket nem lehet egyszerűsíteni egymással.

c) Második egyszerűsítési kör és a végső formula.

Megvizsgáljuk, hogy az újonnan kapott egyszerűsített formulákat (E1 és E2) lehet-e tovább egyszerűsíteni. E1: EGYETEM * TŐKE * KULTÚRA * INFRA kontra E2: TŐKE * KORMÁNY * KULTÚRA * INFRA. Ezek több feltételben is eltérnek, így további egyszerűsítés nem lehetséges.

A minimalizálás végeredménye a megmaradt tovább már nem egyszerűsíthető formulák összessége. A teljes megoldás ezeknek a „VAGY" (+) kapcsolattal való összekötése. A végső megoldási formula (ebben az esetben a köztes megoldás) a következők szerint alakul. A fenti lépések a megfigyelt esetek alapján történő egyszerűsítést mutatják be. A QCA-szoftverek a köztes megoldás során az elméletileg észszerű logikai maradékokat is bevonják a folyamatba a még tömörebb eredmény érdekében. Ezt a bonyolultabb, automatizált folyamatot követve a szoftver a következő végső formulát adja:

INNOVÁCIÓ ← EGYETEM * TŐKE * KULTÚRA * INFRA + TŐKE * KORMÁNY * KULTÚRA * INFRA

Ez a formula a mi végeredményünk. Azt jelenti, hogy két, egymástól független út vezet a sikeres vállalkozáshoz. A következő, utolsó lépésben ezt a tömör matematikai formulát fogjuk lefordítani és értelmezni.

3.6. Az eredmények értelmezése és validálása

A logikai minimalizálás során tömör, absztrakt Boole-formulát kaptunk. Ez azonban nem a végcél, hanem egy eszköz. A következő lépésben lefordítjuk a formulát hétköznapi nyelvre, számszerűsítjük az egyes megoldási utak

fontosságát, ellenőrizzük az eredmények megbízhatóságát, és visszatérünk az esetekhez, hogy mélyebben megértsük a feltárt mechanizmusokat. A minimalizálás végeredménye a következő formula volt: INNOVÁCIÓ ← (EGYETEM * TŐKE * KULTÚRA * INFRA) + (TŐKE * KORMÁNY * KULTÚRA * INFRA). Ez a formula azt mutatja, hogy a sikeres vállalkozás kialakulásához két, egymással egyenértékű, alternatív út (konfiguráció) vezet. Most megvizsgáljuk ezeket az utakat a szoftver által számított minőségi mutatók segítségével.

Megoldási útvonal	Nyers lefedettség	Egyedi lefedettség	Konzisztencia	Tipikus esetek
1. EGYETEM * TŐKE * KULTÚRA * INFRA	0,67	0,42	0,91	Stockholm, Berlin, London
2. TŐKE * KORMÁNY * KULTÚRA * INFRA	0,38	0,13	0,88	Dublin, Amsterdam
Teljes megoldás	0,80	-	0,89	-

14. táblázat: A sikeres innovációhoz vezető megoldási útvonalak és azok paraméterei

A táblázatban az 1. útvonal 0,91-es konzisztenciája azt jelenti, hogy azok a városok, amelyekben egyszerre van jelen erős egyetem, tőke, kultúra és infrastruktúra, 91%-os megbízhatósággal válnak sikeres innovációs központtá. A nyers lefedettség azt mutatja meg, hogy az adott útvonal a sikeres esetek mekkora hányadát képes megmagyarázni. Az 1. útvonal (0,67) a sikeres esetek 67%-át, míg a 2. útvonal (0,38) a 38%-át fedi le (az összeadott értékük nagyobb lehet 1-nél, mert lehetnek átfedések, azaz esetek, amelyeket mindkét út magyaráz). Az egyedi lefedettség azt a hányadot mutatja, amelyet az útvonal kizárólagosan magyaráz, és más út nem. Az 1. útvonal 0,42-es értéke azt jelenti, hogy a sikeres esetek 42%-a csak ezzel az úttal értelmezhető. Ez a mutató segít azonosítani az elméletileg leginkább megkülönböztethető stratégiákat. A táblázat utolsó sora (teljes megoldás) a teljes modellt értékeli. A 0,80-as lefedettség azt jelenti, hogy a két azonosított útvonalunk együttesen a sikeres esetek 80%-át képes megmagyarázni. A fennmaradó 20% más, általunk nem vizsgált tényezőkkel vagy egyedi esetekkel magyarázható. A 0,89-es konzisztencia pedig a modell egészének magas megbízhatóságát mutatja.

A következő lépés az útvonalak elnevezése és értelmezése. Az első útvonal neve lehetne „tudásalapú ökoszisztéma", mivel a megoldás lényege az erős technológiai egyetemek, a bőséges kockázati tőke, a támogató vállalkozói kultúra és

a fejlett digitális infrastruktúra szinergiája. Ez a négy feltétel együttesen elegendő a sikerhez. A második útvonal neve lehetne „kormányzati katalizátor" (TŐKE * KORMÁNY * KULTÚRA * INFRA). Ez alternatív út, amelyben a kormányzati támogatás kiemelkedő szerepet játszik, és képes ellentételezni a (relatíve) gyengébb egyetemi hátteret. Itt a siker receptje a tőke és a kultúra megléte mellett a proaktív kormányzati politika és a fejlett digitális infrastruktúra.

Ezt követően érdemes ellenőrizni, hogy a feltárt útvonalak stabilak-e, vagy csupán az elemzés során hozott döntéseink (például küszöbértékek) termékei. Ezt nevezzük robusztusságvizsgálatnak, melynek három módszere létezik.

a) Kalibrációérzékenység. Enyhén módosítjuk a kalibrációs horgonypontokat (például a világszínvonalú egyetem küszöbét 80 pont helyett 75-re vagy 85-re állítjuk), és újrafuttatjuk az elemzést.
b) Konzisztenciaküszöb-érzékenység.4 Megnézzük, mi történik, ha a konzisztenciaküszöböt 0,80 helyett 0,75-re vagy 0,85-re állítjuk.
c) Esetérzékenység. Egyesével kihagyunk eseteket (városokat) az elemzésből, hogy lássuk, egyetlen eset eltávolítása megváltoztatja-e alapvetően a végeredményt.

Ha a megoldási formulánk lényegi elemei (a két fő útvonal) ezen tesztek során is stabilan megmaradnak, az jelentősen megerősíti eredményeink megbízhatóságát.

Az elemzés utolsó mozzanataként visszatérünk az esetekhez. A QCA ereje abban rejlik, hogy összeköti a formális elemzést az esetalapú tudással, és végül „kvalitatív" leírást ad. Az utolsó lépés tehát, hogy a kapott absztrakt útvonalakat visszakapcsoljuk a konkrét városokhoz, hogy a „mi" után a „hogyan"-ra is választ kapjunk. Az alábbi táblázatban néhány esetet besorolunk az útvonalakhoz.

[4] A konzisztenciaküszöb az a határérték, amely alapján eldöntjük, hogy az igazságtábla egy adott sora (konfigurációja) elégségesnek tekinthető-e az eredmény bekövetkezéséhez. A módszertani standardok szerint ez az érték nem lehet alacsonyabb 0.75-nél. A kutatási gyakorlatban a 0.80-as küszöb használata megszokott, mivel ez már erős bizonyítékot szolgáltat az elégséges kapcsolatra, de még kezelni tudja a társadalomtudományi adatok természetes zaját. Lásd: Schneider, C. Q., & Wagemann, C. (2012). *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge University Press.

Város	Útvonal	Tagsági érték az útvonalban	Megjegyzés
Stockholm	1	0,85	Tipikus eset az 1. útvonalra
Dublin	2	0,72	Tipikus eset a 2. útvonalra
Lisszabon	1+2	0,61; 0,58	Mindkét útvonalat részben követi, vegyes stratégia
Budapest	Egyik sem	0,22; 0,18	Nem követi egyiket sem (deviáns eset a kudarc szempontjából)

15. táblázat: A vizsgált városok besorolása a feltárt megoldási útvonalak szerint

Ez a táblázat a kiindulópont a kvalitatív elemzés folytatásához. Most már célzottan vizsgálhatjuk például a tipikus eseteket. Egy mélyinterjú a stockholmi vállalkozói ökoszisztéma szereplőivel feltárhatja, hogyan működik együtt az egyetem, a tőke és a kultúra a gyakorlatban. Lisszabon esete megmutathatja, hogyan lehet a két stratégiát ötvözni (hibrid eset). Budapest vizsgálata (lehetett volna bármely más város is a példában) rávilágíthat a sikert gátló tényezőkre, és finomíthatja az elméletünket (deviáns eset). Ezzel a lépéssel a kör bezárul: strukturálatlan tudásból indultunk, formális elemzést végeztünk, és most a formális elemzés eredményeivel térünk vissza a mélyebb megértéshez.

15. doboz: QCA-szoftverek kiválasztása és az adatok előkészítése

A QCA gyakorlati alkalmazása során, különösen több feltétel és esetszám mellett, a számítások kézzel szinte elvégezhetetlenek, de legalábbis hibalehetőségekkel terheltek lennének. Bár a logikai alapok minden QCA-szoftverben azonosak, a felhasználói felületek és a kezelési módok jelentősen eltérhetnek egymástól. A rendelkezésre álló eszközök két fő kategóriába sorolhatók, amelyek különböző felhasználói igényeket szolgálnak ki: a grafikus felületű programok és a programozási nyelvekhez rendelt csomagok. A választás leginkább a kutató preferenciáin, technikai felkészültségén és a kutatási projekt sajátos igényein múlik. A doktorandusz, aki most tanulja a módszert, valószínűleg jobban jár egy grafikus felületű programmal, míg a tapasztalt kutató, aki több száz modellt szeretne összehasonlítani, inkább a programozható megoldást választja.

A QCA-szoftverek kiindulópontja mindig egyszerű táblázat, eset × változó mátrix. A legtöbb eszköz a vesszővel tagolt értékeket tartalmazó (CSV) formátumot várja bemenetként. A táblázat struktúrája egyszerű. Az első sor a változók neveit tartalmazza, beleértve a feltételeket és a kimenetet is. Ezeket az elnevezéseket érdemes egyszerűen tartani, kerülve a szóközöket és egyedi karaktereket, mivel egyes szoftverek érzékenyek lehetnek ezekre. Minden további sor egy-egy esetet képvisel. Az adatoknak már kalibráltnak kell lenniük, amikor a szoftverbe töltjük őket. Éles halmazok esetén ez 0 vagy 1 értékeket jelent, életlen halmazok esetén pedig 0 és 1 közötti értékeket. A kalibrálás maga általában a QCA-szoftveren kívül történik, bár egyes eszközök támogatást nyújtanak ehhez a lépéshez is.

A szükséges feltételek vizsgálata a legtöbb szoftverben külön funkcióként érhető el. Az igazságtáblát a szoftver automatikusan létrehozza. Ezt követően az első döntés az esetszámküszöb meghatározása. Kis mintánál ez gyakran egy vagy két eset, nagyobb mintánál lehet magasabb is. A konzisztencia értékét is itt kell megadni. A beállított igazságtábla alapján a szoftver elvégzi a Boole-algebrai egyszerűsítést. Ez a folyamat a Quine–McCluskey-algoritmusra vagy annak valamely változatára épül, amely szisztematikusan azonosítja azokat a feltételeket, amelyek bizonyos körülmények között lényegtelenek. A legtöbb korszerű QCA-eszköz három különböző megoldási stratégiát kínál fel, amelyek abban térnek el egymástól, hogy hogyan kezelik a logikai maradékokat. Az elemzés végeredménye egy vagy több Boole-formula, amely az eredményhez vezető oksági utakat írja le, valamint egy részletes táblázat a modell illeszkedési mutatóival.

A formula jelölésrendszere egyetemes a QCA-ban. A csillag jel (*) a logikai ÉS műveletet jelöli, a plusz jel (+) a logikai VAGY műveletet, a tilde (~) pedig a feltétel hiányát jelzi. Így például a FELTÉTEL1 * ~FELTÉTEL2 + FELTÉTEL3 formula azt jelenti, hogy vagy az első feltétel jelenléte ÉS a második feltétel hiánya együttesen, VAGY a harmadik feltétel önmagában elegendő a kimenethez.

A numerikus eredményeken túl a legtöbb korszerű QCA-szoftver vizuális eszközöket is kínál az eredmények jobb megértéséhez és validálásához. A legfontosabb ilyen eszköz az XY-grafikon, amely a feltétel vagy konfiguráció halmaztagsági értékeit ábrázolja a kimenet halmaztagsági értékeivel szemben kétdimenziós koordinátarendszerben.

A QCA-kutatási jelentésben legalább a következő táblázatokat kell közölni: a kalibrált adattáblát, a szükségességvizsgálat eredményeit, a megoldási útvonalak táblázatát és az igazságtáblát (vagy annak részét). Érdemes továbbá közölni az esetek besorolását és a robusztussági tesztek eredményeit. Függelékbe tehető a nyers adatok táblázata, a kalibrációs döntések részletes indoklása és a teljes igazságtábla (ha a terjedelme miatt a főszövegben csak részletet közöltünk).

4. A minőség biztosítása

A kalibráció – vagyis a nyers adatok halmaztagsági értékekké alakítása – a QCA egyik legnagyobb kutatói szakértelmet igénylő szakasza. Mivel a halmaztagság meghatározása alapvetően befolyásolja az elemzés végeredményét, a minőség biztosításának legfontosabb eleme e folyamat átláthatósága és elméleti megalapozottsága. A legsúlyosabb hiba, ha a kutató pusztán az adatok belső eloszlása alapján határozza meg a halmaztagság határait. Amikor a kutató az átlag vagy a medián alapján kalibrálja az adatait, lényegében relativizálja a halmaztagságot az adott mintához, amely ellentmond a QCA alapvető logikájának. A kutatónak világosan és meggyőzően érvelnie kell amellett, hogy miért éppen az adott értékek jelölik a töréspontokat. A kutatási jelentésben vagy annak függelékében részletesen dokumentálni kell a kalibrációs folyamatot minden feltétel és a kimenet esetében. Ennek tartalmaznia kell a felhasznált adatforrásokat, az elméleti indoklást a horgonypontok elhelyezésére és magát a kalibrációs eljárást vagy képletet.

A társadalomtudományi kutatásokban ritkán fordul elő, hogy a kutatónak minden logikailag lehetséges konfigurációra van tapasztalati esete. Már öt feltétel esetén is 32 lehetséges összetétel létezik, és ezek jelentős részéhez gyakran nem tartozik egyetlen megfigyelt eset sem. Ezeknek a hiányzó eseteknek – a logikai maradékoknak – a kezelése alapvetően befolyásolja a végső megoldást, és itt különül el a gondos, minőségi QCA-kutatás a mechanikus elemzéstől.

A végső megoldási formulát világosan és érthetően kell bemutatni, használva a QCA szabványos jelölésrendszerét, ahol a * jelöli a logikai ÉS műveletet, a + a logikai VAGY műveletet, a ~ pedig a feltétel hiányát. A formula önmagában azonban nem elég. Minden oksági útvonalhoz közölni kell a konzisztencia- és lefedettségértékeket, valamint a teljes megoldás összesített mutatóit is. A képi

megjelenítés segíti az absztrakt logikai formulák megértését. Nagyobb számú feltétel esetén a konfigurációs diagramok vagy a megoldási útvonalak táblázatos bemutatása lehet megfelelőbb. Lényeges, hogy az eredményeket visszakapcsoljuk az egyes esetekhez. Táblázatos formában érdemes bemutatni, hogy melyik eset melyik oksági útvonal tipikus példája, melyek azok az esetek, amelyek több útvonalat is követnek, és melyek a deviáns esetek, amelyek nem illeszkednek a megoldási formulához. A deviáns esetek különösen fontosak, mivel ezek mélyebb vizsgálata gyakran vezet új elméleti belátásokhoz vagy a modell finomításához.

5. Mikor használjuk és mikor ne használjuk a módszert?

5.1. A kutatási kérdés

A QCA akkor a legalkalmasabb az oksági összetettségre összpontosító kérdések megválaszolására, amikor többféle út vezethet ugyanahhoz az eredményhez, vagy amikor a feltételek összetétele számít, nem pedig egyetlen tényező önmagában.

Jó QCA-kutatási kérdések

a) Milyen feltételek vezetnek a hátrányos helyzetű tanulók sikeres egyetemi felvételijéhez? (Például iskolai támogatás * családi háttér * tanácsadói program * ösztöndíj.)
b) Melyek a feltételei a korai iskolaelhagyás sikeres megelőzésének? (Például több út is lehetséges, mint az erős tanár-diák kapcsolat, a családi támogatás vagy az iskolai megelőzés különböző kombinációi.)
c) Milyen feltételek magyarázzák az online oktatás sikerességét különböző iskolatípusokban? (technológiai felszereltség * tanári képzettség * szülői támogatás * diák önszabályozása)
d) Milyen tényezők vezetnek a tanári kiégéshez? (magas terhelés * alacsony szintű önrendelkezés * vezetői támogatás hiánya * munka-magánélet egyensúly hiánya)

e) Milyen különböző útvonalak vezetnek a reziliencia kialakulásához traumát átélt gyerekeknél? (védőfaktorok különböző összetételei)
f) Milyen tényezőkonfigurációk magyarázzák a sikeres terápiás kimenetelt szorongásos zavaroknál? (terápiás kapcsolat * motiváció * társas támogatás * kezelési típus)

Nem megfelelő QCA-kutatási kérdések

a) Hogyan hat a családi háttér a tanulmányi eredményre? – Ez lineáris kapcsolatot feltételez, a regresszió megfelelőbb.
b) Melyik a legfontosabb tényező a motiváció növelésében? – A QCA nem rangsorol, hanem konfigurációkat vizsgál.
c) Mi a tanári kiégés átlagos szintje? – Ez leíró statisztika, nem oksági elemzés.
d) Növeli-e a csoportterápia a gyógyulási arányt? – Igen/nem kérdés, RCT vagy t-teszt való hozzá.

5.2. A minta

Az esetszám szempontjából a QCA jellemzően közepes méretű mintával, tipikusan 10-50 fővel dolgozik. Ez az a tartomány, ahol az esetek száma már túl nagy ahhoz, hogy a hagyományos esettanulmány minden részletre kiterjedjen, de még túl kicsi ahhoz, hogy a statisztikai modellek megbízhatóan működjenek. Kisebb esetszám mellett, körülbelül 4-10 eset között a QCA még mindig hasznos lehet, bár itt a logikai maradékok magas száma kihívást jelenthet. Ilyen esetekben a módszer értéke nem feltétlenül az empirikus általánosításban, hanem a kutatói gondolkodás formalizálásában és a lehetséges oksági utak szisztematikus feltérképezésében mutatkozik meg. A strukturált megközelítés segít azonosítani azokat a konfigurációkat, amelyeket érdemes lenne további esetekkel tesztelni. Megjegyezzük, hogy érdekes módon a QCA egyre gyakrabban jelenik meg nagy mintás kutatásokban is, tipikusan 100 vagy több eset vizsgálatánál.

5.3. A kutatás jellege

A QCA sikeres alkalmazásának strukturális feltételei vannak. Az egyik legfontosabb korlát a feltételek számának észszerű szinten tartása. Ideális esetben 4-7 feltétellel dolgozunk, mivel a lehetséges konfigurációk száma exponenciálisan növekszik. Az időbeli dimenzió kezelése szintén sajátos kihívást jelent. A QCA alapvetően keresztmetszeti elemzésre készült, ahol az eseteket egy adott időpontban vagy időszakban vizsgáljuk. Ha időbeli folyamatokat, változásokat vagy szekvenciákat szeretnénk elemezni, egyedi megközelítésekre van szükség. A QCA időbeli sorozatokban való alkalmazása lehetséges (temporal QCA), de további módszertani megfontolásokat igényel. Alternatívaként a process tracing vagy a szekvenciaelemzés lehet megfelelőbb választás dinamikus folyamatok vizsgálatára. Az esetek kiválasztása a QCA-ban nem véletlenszerű mintavételezésen alapul, hanem célzott kiválasztáson. Az eseteket aszerint kell kiválasztani, hogy relevánsak-e a vizsgált jelenség szempontjából, és fontos, hogy a mintánkban legyenek pozitív (ahol a vizsgált eredmény bekövetkezett) és negatív esetek (ahol nem) is. Ez a megközelítés különbözik a statisztikai reprezentativitás logikájától, és inkább az elméleti relevancia elvét követi.

6. A módszer változatai és haladó alkalmazásai

A kvalitatív összehasonlító elemzés alapváltozatai – a crisp-set és a fuzzy-set QCA – mellett számos specializált változat fejlődött ki az elmúlt évtizedekben. Ezek részben a módszer technikai korlátainak meghaladására, részben új kutatási területeken való alkalmazására irányulnak.

A többértékű QCA (mvQCA) akkor hasznos, amikor a változók természetüknél fogva kategoriálisak, de kettőnél több értéket vesznek fel. Gondoljunk egy ország kormányzati rendszerére (parlamentáris, prezidenciális, félprezidenciális) vagy egy személy végzettségére (alapfokú, középfokú, felsőfokú). Az mvQCA lehetővé teszi, hogy ezeket a nominális változókat természetes formájukban kezeljük anélkül, hogy mesterségesen dichotóm változókká bontanánk őket. A módszer minden többértékű feltételt úgy kezel, mintha több, egymást kölcsönösen kizáró dichotóm feltétel lenne, megőrizve azt a logikai kapcsolatot, hogy

közülük mindig pontosan egy teljesül. Ez különösen hasznos politikai rendszerek összehasonlításában vagy szervezeti kutatásokban, ahol a különböző formák kvalitatíve különböznek egymástól. Hátránya, hogy már néhány többértékű változó esetén is drasztikusan megnő a vizsgálandó konfigurációk száma.

Az időbeli QCA (tQCA) a szabványos módszer statikus megközelítését haladja meg azáltal, hogy formálisan bevonja az idődimenziót. Két módon kezeli az időbeliséget: figyelembe veszi a feltételek bekövetkezési sorrendjét (például demokratizációs folyamatban kritikus lehet, hogy először a gazdasági liberalizáció vagy a politikai nyitás történik-e), valamint modellezi az események közötti időintervallumokat is. Ez a még fejlesztés alatt álló módszer különösen történeti elemzésekben vagy longitudinális kutatásokban hasznos, ahol az időbeli dinamika központi szerepet játszik.

A QCA egyedülálló helyet foglal el a kevert módszertani megközelítések körében, mivel párhuzamos elrendezésben mindkét adattípust közös analitikai keretben kezeli – a kvantitatív mutatókat és a kvalitatív értékeléseket egyaránt halmaztagsági értékekké kalibrálja, majd egyetlen integrált modellben elemzi. Ez különösen értékes az összetett beavatkozások értékelésénél. Szekvenciális feltáró elrendezésben a kezdeti kvalitatív szakasz segít azonosítani a releváns feltételeket, majd a QCA-elemzés feltárja a lehetséges oksági konfigurációkat. Ezek strukturálják a második, kvantitatív szakaszt, amelyben célzottan vizsgálható az azonosított konfigurációk elterjedtsége nagyobb populációban. Fordított esetben, szekvenciális magyarázó elrendezésben a kvantitatív QCA-elemzés után célzottan választunk eseteket a kvalitatív vizsgálathoz, például minden azonosított útvonalból tipikus eseteket, valamint deviáns eseteket, amelyek nem illeszkednek a modellbe.

A QCA a hermeneutikai jellegű kutatáson belül is szolgálhat strukturáló eszközként. A tematikus elemzésben a kódolás után következő absztrakciós lépést segíti, ahol a kategóriákat QCA-feltételekként kezelve feltárja, hogy milyen konfigurációik kapcsolódnak a központi jelenséghez. Az így születő témák nem egyszerű kategóriacsoportok, hanem bonyolult összefüggéseket tükröző konstrukciók.

7. Összefoglalás

Ebben a fejezetben a kvalitatív összehasonlító elemzést (QCA) mutattuk be mint olyan kutatási módszert és gondolkodásmódot, amely hidat képez az esetalapú mélyértelmezés és a szabályszerűségeket kereső elemzés között. A QCA alapvető célja az oksági komplexitás megértése. Olyan jelenségeket vizsgál, ahol az eredményhez több, alternatív út (ekvifinalitás) vezethet, és ahol a feltételek nem önmagukban, hanem sajátos összetételekben (konjunkturális okság) fejtik ki hatásukat. Megismertük a QCA-elemzés lépéseit a feltételek kiválasztásától kezdve a legkritikusabb szakaszon, az elméletileg megalapozott kalibráción át egészen a végső Boole-formula értelmezéséig. Láthattuk, hogy az igazságtábla segítségével hogyan azonosíthatjuk a sikerhez vezető konfigurációkat, a logikai minimalizálás révén pedig hogyan egyszerűsíthetjük ezeket a konfigurációkat a lényegi magjukra. Figyelmet fordítottunk a kutatói döntések fontosságára, mint például a konzisztenciaküszöb beállítására vagy a logikai maradékok tudatos kezelésére, mert mindezek alapvetően meghatározzák a végeredményt. Hangsúlyoztuk, hogy bár a szoftverek elvégzik az összetett számításokat, a minőségi kutatás a kutató és a szoftver folyamatos párbeszédén alapul, amelyben a kutató hozza az elméleti tudást, az esetek ismeretét és a kontextuális megértést. A módszer ereje abban rejlik, hogy a formális elemzés után lehetővé teszi a visszatérést az esetekhez, segítve a tipikus és deviáns esetek mélyebb vizsgálatát. A fejezet végén áttekintettük a minőségbiztosítás lényeges szempontjait, a módszer alkalmazási korlátait, valamint a haladóbb változatokat, amelyek a QCA-t a korszerű, integratív és kevert módszertanú kutatások erőteljes eszközévé teszik.

8. Önellenőrző kérdések

1. Mi a legsúlyosabb hiba, amelyet egy kutató a kalibráció során elkövethet, és miért az?

Válasz: A legsúlyosabb hiba, ha a kutató a horgonypontokat (a teljes tagság, az átmenet és a teljes nem tagság küszöbértékeit) nem külső, elméleti tudás alapján, hanem a vizsgált minta belső eloszlása (például átlag, medián, kvartilisek) alapján határozza meg. Ez azért hiba, mert relativizálja a halmaztagságot az adott mintához, ami ellentmond a QCA alapvető logikájának, miszerint a halmaztagságnak külső, tárgyilagos mércén kell alapulnia. Az ilyen eljárás aláássa az eredmények érvényességét és összehasonlíthatóságát.

2. Mik azok a logikai maradékok, és miért fontos a kezelésük?

Válasz: A logikai maradékok azok az elméletileg lehetséges konfigurációk, amelyekre a vizsgált mintában nincs egyetlen tapasztalati eset sem. Kezelésük azért lényeges, mert a logikai minimalizálás során a szoftver felhasználhatja őket a végső formula egyszerűsítésére. A kutató döntése arról, hogy mely maradékok bevonását engedélyezi (a komplex, parszimónikus és köztes megoldás stratégiáin keresztül), alapvetően befolyásolja a végeredmény egyszerűségét és elméleti plauzibilitását. A tudatos kezelés biztosítja az egyensúlyt az empirikus megalapozottság és az értelmes általánosítás között.

3. Mi a legfontosabb különbség a konzisztencia és a lefedettség mutatói között egy megoldási útvonal értékelésénél?

Válasz: A konzisztencia a megbízhatóságot vagy a bizonyosságot méri, tehát azt mutatja meg, hogy ha egy eset követ egy adott utat, milyen biztosan vezet az a kívánt kimenetelhez. Ez az elégségesség mértéke. Ezzel szemben a lefedettség a magyarázóerőt vagy a fontosságot méri, és azt mutatja meg, hogy az adott út az összes sikeres eset mekkora hányadát képes megmagyarázni. Lehetséges egy nagyon konzisztens (megbízható), de alacsony lefedettségű (ritka, kevés esetet magyarázó) út, és fordítva.

4. Milyen típusú kutatási kérdés esetén választaná a QCA-t egy hagyományos regressziós elemzéssel szemben?

Válasz: A QCA-t akkor választanám, ha a kutatási kérdésem az oksági komplexitásra, a hogyanra és a milyen feltételek együttesen kérdésre összpontosít.

Például milyen különböző konfigurációk vezetnek a sikeres társadalmi integrációhoz? Ezzel szemben a regressziós elemzést akkor választanám, ha a változók nettó, elkülönült hatására lennék kíváncsi. Például minden más modellbe beemelt tényezőre kontrollálva, egy évvel több iskolázottság mennyivel növeli a jövedelmet? A QCA a konfigurációkat, a regresszió a hozzávalók egyedi hatását vizsgálja.

5. Mit jelent a gyakorlatban, hogy a QCA az ekvifinalitás elvét kezeli? Mondjon egy példát!

Válasz: Az ekvifinalitás azt az elvet jelenti, hogy ugyanazt a kimenetelt több, egymástól különböző oksági úton, eltérő konfigurációk révén is el lehet érni. Nincs királyi út a sikerhez. Például a fejezetben bemutatott kutatásban a sikeres vállalkozáshoz két eltérő út is vezethetett: az egyik a „tudásalapú ökoszisztéma" (erős egyetem, tőke, kultúra és infrastruktúra), a másik a „kormányzati katalizátor" út. Mindkét feltételkombináció sikerhez vezetett, de különböző tényezőkön keresztül.

Ajánlott irodalom

Di Paola, N., Chari, S., Iannacci, F., & Kraus, S. (2025). Configurational theory in business and management research: Status quo and guidelines for the application of qualitative comparative analysis (QCA). *Technological Forecasting and Social Change, 211*, 123907. https://doi.org/10.1016/j.techfore.2024.123907

Duşa, A. (2019). *QCA with R*. Springer. https://doi.org/10.1007/978-3-319-75668-4

Fiss, P. C., Cambré, B., & Marx, A. (Eds.). (2013). *Configurational Theory and Methods in Organizational Research*. Emerald Group Publishing Limited. https://doi.org/10.1108/rso

Geremew, Y. M., Huang, W.-J., & Hung, K. (2024). Fuzzy-Set Qualitative Comparative Analysis as a Mixed-Method and Analysis Technique: A Comprehensive Systematic Review. *Journal of Travel Research, 63*(1), 3–26. https://doi.org/10.1177/00472875231168619

Greckhamer, T., Furnari, S., Fiss, P. C., & Aguilera, R. V. (2018). Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. *Strategic Organization, 16*(4), 482–495. https://doi.org/10.1177/1476127018786487

Greckhamer, T., Misangyi, V. F., & Fiss, P. C. (2013). The Two QCAs: From a Small-N to a Large-N Set Theoretic Approach. In P. C. Fiss, B. Cambré, & A. Marx (Eds.), *Research in the Sociology of Organizations. Configurational Theory and Methods in Organizational Research* (Vol. 38, pp. 49–75). Emerald Group Publishing Limited. https://doi.org/10.1108/S0733-558X(2013)0000038007

Hak, T., Jaspers, F., & Dul, J. (2013). The Analysis of Temporally Ordered Configurations: Challenges and Solutions. In P. C. Fiss, B. Cambré, & A. Marx (Eds.), *Research in the Sociology of Organizations. Configurational Theory and Methods in Organizational Research* (Vol. 38, pp. 107–127). Emerald Group Publishing Limited. https://doi.org/10.1108/S0733-558X(2013)0000038009

Marx, A., & Dusa, A. (2011). Crisp-Set Qualitative Comparative Analysis (csQCA), Contradictions and Consistency Benchmarks for Model Specification. *Methodological Innovations Online, 6*(2), 103–148. https://doi.org/10.4256/mio.2010.0037

Oana, I.-E., Schneider, C. Q., & Thomann, E. (2021). *Qualitative Comparative Analysis Using R*. Cambridge University Press. https://doi.org/10.1017/9781009006781

Ragin, C. C. (2000). *Fuzzy-set social science*. University of Chicago Press.

Ragin, C. C. (2014). *The comparative method : moving beyond qualitative and quantitative strategies*. University of California Press.

Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge University Press.

Skaaning, S.-E. (2011). Assessing the Robustness of Crisp-set and Fuzzy-set QCA Results. *Sociological Methods & Research, 40*(2), 391–408. https://doi.org/10.1177/0049124111404818

Sun, Y., Hsieh, J., & Gao, X. (2025). Integrating Qualitative Comparative Analysis With Reflexive Thematic Analysis in Theme Development. *International Journal of Qualitative Methods, 24*, Article 16094069251318749. https://doi.org/10.1177/16094069251318749

VI. EPISZTEMIKUS HÁLÓZATELEMZÉS (ENA)

TURÓS MÁTYÁS

A fejezet a következő kérdéseket válaszolja meg:

a) Mik az episztemikus hálózatelemzés (ENA) elméleti alapjai?
b) Hogyan zajlik egy ENA-kutatás gyakorlati folyamata az adatgyűjtéstől az interpretációig?
c) Milyen minőségbiztosítási elveket kell követni az ENA alkalmazása során?
d) Mikor érdemes és mikor nem tanácsos ezt a módszert választani?
e) Milyen egyedi változatai és fejlett alkalmazásai léteznek a módszernek?
f) Hogyan értelmezhetők és validálhatók az ENA eredményei?

1. Bevezetés

Legyen szó egy politikai vita érvelési logikájáról, egy szervezet belső kommunikációs mintázatairól vagy egy szakértői vélemény mögött rejlő tudásstruktúráról, a kutatásnak célja lehet az is, hogy a jelenség elemeinek azonosításán túl megértse az azok között lévő kapcsolatrendszert is. Ugyanis a megértés kulcsa gyakran a kapcsolatokban rejlik. Például képzeljük el, hogy egy tapasztalt sebész és egy rezidens ugyanazt az összetett műtéti esetet elemzi. Mindketten ugyanazokat a szakmai fogalmakat használják – anatómiai struktúrák, kockázati tényezők, műtéti technikák –, de a fogalmakat is ugyanúgy kapcsolják össze a gondolkodásukban? Lehet, hogy a szakértő gondolkodásában az egyes elemek egy jól szervezett, összetett hálózatot alkotnak, ahol az anatómiai tudás szorosan összefonódik a kockázatértékeléssel. A kezdő esetében ugyanezek az elemek talán elszigetelten jelennek meg. De hogyan tehetjük láthatóvá és mérhetővé e különbséget? Az episztemikus hálózatelemzés (Epistemic Network Analysis, ENA) erre ad választ. Ez a viszonylag új, a kvantitatív etnográfia keretein belül kifejlesztett módszer azt mutatja meg, hogy a kutatásban azonosított elemek hogyan kapcsolódnak össze, milyen hálózatot alkotnak – és mindezt meg is jeleníti egy könnyen értelmezhető ábrán. A módszer három fogalom köré épül. A kódok azok az elemek (fogalmak, cselekvések, értékek), amelyek egy diskurzus építőköveit alkotják. A sztanza határozza meg azt a kontextust, amelyen belül az elemek együttes megjelenését kapcsolatként értelmezzük. Végül a hálózati modell az a képi és matematikai megjelenítés, amely megmutatja, hogyan szerveződnek ezek a kapcsolatok egy adott személy vagy csoport gondolkodásában.

Bár az ENA a szociális hálózatelemzésből (SNA) kölcsönzi képi nyelvét, és alapvetően SNA típusú kérdéseket válaszol meg, míg az SNA jellemzően emberek közötti kapcsolatokat vizsgál, az ENA a fogalmak, gondolatok és cselekvések közötti kognitív vagy szemantikai kapcsolatokra összpontosít. Sokoldalúságát mutatja, hogy nemcsak a tanulás vagy a szakértelem, hanem például sebészeti beavatkozások alatti kommunikáció vagy akár multimodális (szóbeli és nem szóbeli) kölcsönhatások elemzésére is sikeresen alkalmazták. Az ENA a kvalitatív kódolás gazdagságát megőrizve matematikailag megalapozott, következtető statisztikával tesztelhető modelleket hoz létre. Ez lehetővé teszi például, hogy tárgyilagosan hasonlítsuk össze különböző csoportok (például politikai pártok, szakértői csoportok) gondolkodási struktúráit, vagy nyomon kövessük, hogyan változik egy gondolkodási séma az idő múlásával.

2. Elméleti alapok

Az episztemikus hálózatelemzés egy tágabb kutatásmódszertani keret, a kvantitatív etnográfia (Quantitative Ethnography, QE) központi módszere. A QE célja, hogy megőrizze a sűrű, értelmező leírások gazdagságát, de lehetővé tegye nagy mennyiségű adat rendszerező és statisztikai elemzését is. Az ENA ennek a célkitűzésnek a gyakorlati megtestesülése. A célkitűzésnek megfelelően az ENA elméleti alapjait több, egymást kiegészítő tudományos hagyomány alakította: a) az elosztott megismerés (distributed cognition) szerint a tudás megtestesül az ember által használt eszközökben és a társas kölcsönhatásokban is; b) a szituatív tanulás elgondolása (situated learning) szerint a szakértővé válás nemcsak információk felhalmozását jelenti, hanem egy közösség gondolkodási és cselekvési mintáinak fokozatos elsajátítását is; c) az episztemikus keret (epistemic frame) fogalma, amely szerint minden szakmai közösség egy jellegzetes gondolkodási és cselekvési struktúrán belül működik.

Az episztemikus keret egy olyan összetett kognitív struktúra, amely meghatározza, hogyan gondolkodnak és cselekednek az adott terület művelői. Öt szorosan összefonódó elemből áll: a) készségek (mit kell tudni csinálni?); b) tudás (milyen tényeket és fogalmakat kell ismerni?); c) értékek (mi számít fontosnak vagy helyesnek az adott területen?); d) identitás (hogyan látják magukat a terület művelői?); e) episztemológia (honnan tudják, hogy amit tudnak, az érvényes?).

Például egy építészmérnök episztemikus keretében a tudás magában foglalja a statikai számításokat, a készségek közé tartozik a tervezőszoftverek használata, az értékek között szerepel a biztonság és a fenntarthatóság. Az identitás azt jelenti, hogy alkotó problémamegoldónak tekinti magát, az episztemológia pedig meghatározza, hogy a szimulációk és számítások hogyan igazolják egy terv helyességét.

Az ENA alapvető felismerése és központi elméleti állítása, hogy a szakértelem lényege nem az, hogy valaki az ötből hány elemet sajátított el, hanem az, hogy az elemeket hogyan kapcsolja össze. Az ENA ezt az elvet egy egyszerű, de erőteljes feltételezéssel operacionalizálja, amely szerint, ha két fogalom vagy cselekvés gyakran jelenik meg együtt valakinek a diskurzusában, akkor ezek az elemek nagy valószínűséggel összekapcsolódnak az illető gondolkodásában is. Ez a pszicholingvisztikából és a kognitív tudományból származó felismerés – mely szerint az emberi elme a fogalmak jelentését és azok kapcsolatait a

környezetében észlelt statisztikai mintázatokon, azaz az elemek együttes előfordulásának gyakoriságán keresztül rögzíti – teszi lehetővé, hogy a megfigyelhető nyelvi és cselekvési adatokból a mögöttes, rejtett kognitív struktúrákra következtessünk. Így válik az ENA képessé arra, hogy ne csak azt mutassa meg, mit tud valaki, hanem azt is, hogyan szervezi a tudását.

3. A kutatás folyamata

3.1. Adatgyűjtés és előkészítés

Az ENA használható interjúk és fókuszcsoportok átirataival, online fórum-interakciókkal, hivatalos dokumentumok szövegeivel, írásbeli válaszokkal. A lényeg, hogy az adatoknak legyen egy belső logikája vagy időrendje, amely mentén a kapcsolatok feltárhatók. A kutatóknak érdemes szem előtt tartani, hogy az átírás rendkívül időigényes, egy óra hanganyag jellemzően négy-hat órát vesz igénybe. Az átirat részletessége kritikus, mert a szünetek, hezitálások jelölése fontos információkat hordoz a gondolkodás folyamatáról. Már a folyamat elején érdemes a későbbi sztanzaelemzésre gondolni, mert a túl hosszú, tagolatlan szövegblokkok megnehezítik az elemzési ablakok későbbi meghatározását. Kitalált példánkban stimulált felidézéses interjúkat készítünk, ahol a tanárok saját órájuk videófelvételét nézve szólnak hozzá döntéseikhez, és az így kapott szövegeket elemezzük ENA segítségével.

3.2. Kódolás

A kódok lesznek a hálózat csomópontjai, ezért közvetlenül kapcsolódniuk kell a kutatási kérdéshez. A kódolási séma (vagy kódkönyv) létrehozásának első lépéseként példánkban a szakirodalom alapján, deduktívan határozunk meg néhány kiinduló kódot, mint a PED_ELV (pedagógiai elvek) vagy a BEAV_STRAT (beavatkozási stratégia). Ezt követi az induktív bővítés, amikor az első néhány interjú elemzése során az adatokból kiindulva új, releváns kódokat azonosítunk, mint például a REFLEX (saját gyakorlat reflexiója). A kódok kedvező száma általában öt és tizenöt között van. Ha túl kevés van, elveszítjük a részleteket, ha túl sok, a modell áttekinthetetlenné válik. A minőség biztosítása érdekében két

kódoló dolgozik, a kódolók közötti megbízhatóságot Cohen-féle kappával mérjük. A kódolók először az adatok egy részhalmazát (10%) kódolják egymástól függetlenül. Ha a kappa értéke alacsony (például 0,62), a nézeteltéréseket a kódolók megvitatják, és a kódkönyvet finomítják. Ezt a folyamatot addig ismétlik egy újabb adatrészleten, amíg el nem érik a megfelelő, 0,8 feletti egyezést. Ekkor az egész szöveget kódolják, a kódokon újrafuttatjuk a megbízhatósági elemzést, és a kappa értékét a kutatási jelentésben közöljük (vagy a 0,8 feletti egyezés elérése után az egyik kódoló a finomított kódkönyv segítségével kódolja a teljes adathalmazt). Mivel a konzisztencia fenntartását nagyban segíti egy határeseti példatár készítése, a nehezen eldönthető eseteket rögzítjük, és a dokumentumot a kódkönyvvel együtt a kutatási jelentés függelékeként közöljük.

3.3. Sztanzaelemzés

A sztanzaelemzés az ENA legkritikusabb és legkevésbé intuitív lépése. A logikájának megértésében egy szemléletes analógia segíthet, amelyet a módszer kidolgozói is használtak. Képzeljük el az elemzést egy koktélpartiként. Ebben a partiban a vendégek maguk a kódok (fogalmak, cselekvések), a beszélgetőcsoportok pedig a sztanzák. A kutató időközönként lefotózza a partit. Minden egyes fotó (vagyis minden egyes csúszó ablak) megmutatja, hogy éppen mely kódok beszélgetnek egymással egy csoportban, azaz melyek jelennek meg együtt az adott sztanzán belül. Az ENA alapvető feltételezése az, hogy azok a vendégek (kódok), amelyek a parti során gyakran jelennek meg ugyanabban a beszélgetőcsoportban (sztanzában), szorosabb kapcsolatban állnak egymással a vizsgált személy gondolkodásában is. A sok fotó elemzése után kirajzolódik a parti teljes társasági hálózata, és kiderül, mely kódok alkotnak szoros közösségeket, és melyek állnak távol egymástól. Ez a hálózat felel meg a végső ENA-modellnek. Elméleti alapja az a feltételezés, hogy a diskurzusban egymáshoz közel elhangzó fogalmak a gondolkodásban is összekapcsolódnak.

A sztanzadöntéssel tehát azt határozzuk meg (operacionalizáljuk), hogy mit jelent a gyakorlatban az „egymáshoz közeli". A folyamat során egy „csúszó ablakot" mozgatunk végig a szövegen, amely meghatározza, hogy mi számít együttes előfordulásnak. Ehhez két paramétert kell megadnunk: a szövegegységet (például egy beszédforduló) és az ablak méretét (például négy egymást követő beszédforduló). A megfelelő ablakméret beállítása azért fontos, mert a túl kicsi ablak töredezett hálózatot eredményezhet, míg a túl nagy ablak elmoshatja a szerkezetet. Érdemes több ablakmérettel is kísérletezni, és a döntést az adatok

természetére alapozva (például „a tanárok átlagosan három-négy beszédfordulóban fejtenek ki egy gondolatot") meghozni.

Ez a kutatói döntés felvet egy fontos módszertani kérdést: ha a sztanza paramétereit a kutató állítja be, hol van ebben a tudományos objektivitás? Míg a kódolás megbízhatóságát egy statisztikai mutatóval (például Cohen-kappa) is ellenőrizhetjük, itt nincs ilyen egzakt mérőszám. A sztanzaelemzés tudományosságát azonban nem a szubjektivitás teljes kiküszöbölése, hanem a kutatói döntések elméleti megalapozottsága, empirikus indoklása és módszertani átláthatósága biztosítja. A döntés nem önkényes. A kutatónak a vizsgált jelenség természete és a kutatás elméleti kerete alapján kell meghatároznia, hogy mekkora az a gondolati egység, amelyen belül a fogalmak összekapcsolódása jelentéssel rendelkezik. Néhány példa:

a) Ha két ember vitáját elemezzük, ahol a felek közvetlenül egymás mondataira válaszolnak. Itt egy rövid ablak (például két-három beszédforduló) lehet indokolt, mert az elméletünk szerint a szoros összefüggés számít.
b) Ha egy kreatív csoport ötletelését vizsgáljuk, ahol a gondolatok szabadabban kapcsolódnak, egy tágabb ablak (például öt-hét beszédforduló vagy egyperces időablak) lehet a jó választás, mert az elméletünk szerint a tágabb kontextus is megfelelő.
c) Egy gyors, aszinkron felületen, ahol a hozzászólások egymásra csúsznak, lehet, hogy nem a beszédfordulók száma, hanem egy időbélyeg-alapú ablak (például minden 30 másodpercen belül írt üzenet) a leginkább indokolt.

Továbbá a tudományos szigor legfontosabb eszköze a robusztussági vizsgálat. A helyes kutatói gyakorlat az, hogy a véglegesnek szánt modell mellett az elemzést lefuttatjuk kissé eltérő sztanzaparaméterekkel is. Célunk annak tesztelése, hogy az eredményeink mennyire érzékenyek a beállítások megváltoztatására. Néhány példa:

a) Ablakméret változtatása: Ha a kutató az empirikus és elméleti indokai alapján a négyes ablakméret mellett döntött, érdemes az elemzést elvégeznie egy szűkebb (például hármas) és egy tágabb (például ötös) ablakkal is. Ezt követően összeveti a kapott hálózati modelleket és a statisztikai eredményeket. Ha a kutatás fő következtetései – például a szakértői és kezdő hálózatok közötti alapvető strukturális különbség – stabilak maradnak a különböző beállítások mellett is, akkor az eredmények robusztusak, ami megerősíti az értelmezés érvényességét.

b) Szövegegység újradefiniálása: Ha az adatok lehetővé teszik, kipróbálható, hogy mi történik, ha nem beszédfordulókban, hanem például mondatokban határozzuk meg az alapegységet. Ha a főbb mintázatok ekkor is hasonlóak maradnak, az szintén az eredmények stabilitását jelzi. Amennyiben az eredmények drasztikusan megváltoznak a paraméterek módosításával, az nem a kutatás kudarcát jelenti, hanem önmagában is egy fontos tudományos eredmény, mert azt jelzi, hogy a vizsgált jelenség rendkívül érzékeny a kontextus méretére. A lényeg minden esetben a folyamat átláthatósága, és a kutatási jelentésben a választott paramétereket, azok indoklását és az elvégzett robusztussági vizsgálat eredményét is dokumentálni kell.

Miután a kódolás és a sztanzadefiníciók elkészültek, a szoftver elvégzi a számításokat. A program a sztanzák alapján szomszédsági mátrixokat (találkozhatunk az „adjacenciamátrix" kifejezéssel is) hoz létre, ezeket összegzi minden elemzési egységre (például minden tanárra), majd a kapott sokdimenziós vektorokat egy dimenziócsökkentő eljárás (SVD) segítségével egy közös, általában kétdimenziós vizuális térben ábrázolja. Ennek eredményeképpen a kutató többféle kimenetet is kap: minden résztvevőhöz egyéni hálózati modellt, csoportonkénti átlagolt hálózatokat és egy közös ENA-teret, ahol a hálózatok a centroidjaik (súlypontjaik) alapján összehasonlíthatóvá válnak.

16. doboz: Az episztemikus hálózatelemzés matematikai alapjai és transzformációs lépései

Bár az ENA használatához nem szükséges a mögöttes matematika mélyreható ismerete, a folyamat koncepcionális megértése segít elkerülni, hogy az eszközt fekete dobozként kezeljük. Az ENA lényegében egy négy lépésből álló transzformációt hajt végre: a strukturálatlan szövegtől eljut egy értelmezhető, kvantitatív vizuális térig.
Az elemzés első lépése, hogy minden egyes elemzési egységre (például minden tanárra) létrehoz egy hálózati modellt. Ez matematikailag egy hosszú vektort (számsort) jelent, amely azt rögzíti, hogy az adott egység diskurzusában a kódok (például „PED_ELV", „TAN_SZÜK") milyen gyakran jelentek meg egymással egy sztanzán belül. Egy 10 kódot tartalmazó modellben ez a vektor az 1-es és 2-es, az 1-es és 3-as, ..., a 9-es és 10-es kódok közötti kapcsolatok erősségét tartalmazza. Ez a nyers hálózatreprezentáció.

Ahogy a futók teljesítményének összevetéséhez sem elég a lépéseik számát nézni (a rövidebb távot futó kevesebbet lép), úgy a diskurzusok összehasonlításánál sem elég a nyers kapcsolati gyakoriságokat nézni. Lehet, hogy egy tanár egyszerűen csak többet beszélt, ezért nála minden kapcsolat erősebbnek tűnik. A normalizálás ezt a bőbeszédűséget statisztikailag kontrollálja, biztosítva, hogy a hálózatok szerkezetét és ne a puszta megszólalások számát hasonlítsuk össze.

A normalizált hálózati vektorok egy rendkívül sokdimenziós térben léteznek (annyi dimenzió, ahány lehetséges kapcsolat van a kódok között). Ezt emberi szemmel képtelenség értelmezni. Az SVD egy olyan statisztikai eljárás – a faktorelemzéshez és a főkomponens-elemzéshez hasonlóan –, amely megkeresi a legfontosabb mintázatokat, a legjelentősebb variációt hordozó fő tengelyeket ebben a sokdimenziós térben. Lényegében leegyszerűsíti a komplex teret egy kezelhetőbb, tipikusan kétdimenziós térképpé (az ENA-ábrává), miközben igyekszik a lehető legtöbb lényeges információt megőrizni. Az ábra X és Y tengelye ennek a dimenziócsökkentésnek az eredménye, azaz azok a legfontosabb mögöttes mintázatok, amelyek mentén a vizsgált hálózatok a leginkább különböznek.

Végül a szoftver az SVD által létrehozott kétdimenziós térben helyezi el minden egyes elemzési egység hálózatának súlypontját (centroidját). Azok az egységek, amelyeknek a gondolkodási hálózata hasonló, egymáshoz közel kerülnek ezen a térképen, míg a különböző struktúrákat használók távolabb lesznek. A csoportok átlagos hálózatait és a köztük lévő különbségeket is ebben a térben tudjuk vizsgálni és statisztikailag tesztelni.

3.4. Értelmezés

Ebben a szakaszban a matematikai modelleket és a képi megjelenítéseket értelmezzük, hogy érdemben válaszoljunk a kutatási kérdésünkre. Az értelmezés során a vizuális mintázatok mellett két metrikára támaszkodunk: a hálózat egészének kohézióját mérő súlyozott sűrűségre (weighted density) és az egyes kódok központi szerepét jelző relatív centralitásra (relative centrality). Az ENA lehetővé teszi a csoportok hálózatainak többszintű összehasonlítását. Először elemezzük a csoportok átlagolt hálózatait, amelyek megmutatják az adott csoportra jellemző tipikus kapcsolati struktúrát. A különbségek feltárására egy másik hatékony eszköz a kivonási hálózat, amely két hálózat különbségét jeleníti meg, kiemelve, hogy mely kapcsolatok erősebbek az egyik vagy a másik csoportban. A vizuális különbségek statisztikai szignifikanciáját a hálózatokat sűrítő centroidok pozíciójának összehasonlításával (például t-próbával) ellenőrizzük. A képzeletbeli ábra alapján a kezdő tanárok hálózata egy reaktív mintázatot mutat: az erős TAN_SZÜK ↔ BEAV_STRAT kapcsolat arra utal, hogy a tanulói viselkedésre azonnali

beavatkozással reagálnak, de ezt ritkán kapcsolják össze átfogóbb pedagógiai elvekkel (PED_ELV). Ezzel szemben a szakértők hálózata egy mérhetően integráltabb struktúrát tár fel, amelynek szignifikánsan magasabb a súlyozott sűrűsége. Ebben a hálózatban a PED_ELV (pedagógiai elvek) kód magas relatív centralitása és a PED_ELV ↔ TAN_SZÜK ↔ OSZT_DIN háromszög együttesen jelzi, hogy a tanulói viselkedést az osztály kontextusában, a pedagógiai elveik központi szervezőerejével értelmezik, és a beavatkozásuk ebből az összetett mérlegelésből fakad. A két csoport centroidjának statisztikai összehasonlítása ($t(28) = 4{,}32$ $p < 0{,}001$) megerősíti, hogy ez a strukturális különbség szignifikáns.

3.5. Az eredmények értelmezése

A kutatásban a számszerű és ábrákon megjelenített eredményeket visszavezetjük az eredeti adatokhoz, és ezzel bezárjuk az értelmezési kört. Például a modellben azonosított mintázatokat idézetek segítségével töltjük meg jelentéssel. A kezdő tanár rövid, lényegre törő reakciója – „Péter nem figyel, rászólok. Ha még mindig nem, akkor kiküldöm." – tökéletesen szemlélteti a reaktív, elvektől kevéssé vezérelt hálózatot. Ezzel szemben a szakértő tanár idézete – „Péter figyelmetlensége [TAN_SZÜK] jelzés, hogy az osztály fárad [OSZT_DIN]. A differenciálás elvét követve [PED_ELV] páros munkát adok [BEAV_STRAT]..." – életre kelti az integrált, kontextualizált hálózati modellt.

Természetesen a kutatási folyamat ritkán ennyire vonalszerű. Az eredmények értelmezése során gyakran merülnek fel új kérdések, amelyek visszavezethetnek a korábbi szakaszokhoz, például a kódkönyv finomításához. A reflexív, iteratív folyamat a minőségi ENA-kutatás része lehet.

17. doboz: ENA-szoftverek (Web Tool, R csomagok) használata és beállításai

Az episztemikus hálózatelemzés a gyakorlatban összetett számításokat igénylő módszer, amely szoftveres támogatás nélkül szinte elvégezhetetlen. A legismertebb és legszélesebb körben használt ENA-szoftver az ENA Web Tool, amely egy grafikus felületű webalkalmazás. Mellette léteznek programozási környezetben futó csomagok is (például az R nyelvű rENA és ona csomagok), amelyek nagyobb rugalmasságot kínálnak a jártasabb felhasználóknak. A kötet alapelveihez híven nem célunk konkrét szoftvertermékek részletes bemutatása. Ehelyett azokat az alapvető funkcionális elveket és elemzési lépéseket ismertetjük, amelyek az ENA-szoftverek közös jellemzői.

Az elemzés első lépése minden esetben egy strukturált, táblázatos adatállomány (jellemzően CSV- vagy Excel-fájl) előkészítése. A szoftverek logikája erre a struktúrára épül, és lehetővé teszi a kutató számára, hogy az elemzés alapvető paramétereit az adatállomány oszlopainak kijelölésével határozza meg. A legfontosabb paraméterek, amelyeket minden ENA-szoftver bekér, az elemzési egységek (Units), amelyek az összehasonlítandó entitásokat azonosítják (például egyének, csoportok), a kódok (Codes), amelyek a hálózat csomópontjait fogják alkotni, valamint a kontextus és a sztanzadefiníció, amellyel a kutató azt határozza meg, hogy a szoftver milyen logikai egységeken belül és mekkora ablakban keresse az együtt-előfordulásokat. A szoftverek gyakran lehetőséget adnak a paraméterek egymásba ágyazására is (például egy adott felszólaló hálózatának vizsgálata egy adott színdarab adott felvonásán belül), ami rendkívül finomhangolt összehasonlításokat tesz lehetővé.

A paraméterek megadása után a szoftver a háttérben automatikusan elvégzi a számításokat. A felhasználó az összetett matematikai lépésekből (adjacenciamátrixok létrehozása, vektorok normalizálása, dimenziócsökkentés) keveset lát, a folyamat végeredménye egy interaktív képi felület. Ezen a felületen jelenik meg a központi ENA-tér, a csomópontok és a köztük lévő kapcsolatok erősségét jelző élek, valamint az egyes elemzési egységek hálózatait sűrítő centroidok. A kutatónak lehetősége van a megjelenítés testre szabására, például a csoportok átlagolt hálózatainak megjelenítésére, a kontrasztív elemzést segítő kivonási hálózatok létrehozására vagy a tengelyek és címkék módosítására.

A korszerű ENA-szoftverek grafikus felületén a kutató általában rá tud kattintani egy sajátos kapcsolatra a hálózati ábrán, és a szoftver megjeleníti azokat az eredeti szövegrészleteket, amelyek alapján az adott kapcsolat létrejött. Ez a funkció elengedhetetlen az eredmények mélyebb, kontextualizált értelmezéséhez. A szoftverek lehetőséget biztosítanak a különböző kimenetek kimentésére. Ide tartoznak a publikációs minőségű hálózati ábrák, valamint azok a számszerűsített adatok – mint a centroidok koordinátái vagy az élek súlyai –, amelyek más statisztikai szoftverekben további elemzések (például regresszió, klaszteranalízis) bemeneteként szolgálhatnak. Ez a funkció teszi lehetővé az ENA integrálását más matematikai módszerekkel, összekötve a strukturálatlan adatokból nyert strukturális modelleket a statisztikai elemzések szélesebb világával.

4. A minőség biztosítása

A minőséget tekintve az első kritikus terület az alapok szilárdsága. Ennek sarokköve a kódolási séma érvényessége. A kódok a hálózat csomópontjai, így ha ezek rosszul meghatározottak, az egész modell torzulni fog. A konstruktumvaliditás azt a kutatásmódszertani alapelvet takarja, hogy a kódjaink valóban azt a mögöttes jelenséget ragadják-e meg, amelyet mérni szándékozunk. Egy túl tág kód, mint például a „POZITÍV" minden pozitív megnyilvánulásra, összemoshat olyan különböző jelenségeket, mint a dicséret, az egyetértés vagy egy pozitív érzelem kifejezése, ezzel használhatatlanná téve az elemzést. Ezzel szemben egy jól kialakított kód, mint a „FORMATÍV_ÉRT" (Formatív értékelés), pontos meghatározással, valamint befoglalási és kizárási feltételekkel rendelkezik. Egy jó kódkönyv konkrét pozitív („Látom, hogy jó úton jársz…") és negatív („Ez négyes munka") példákkal segíti a konzisztens alkalmazást.

A megfelelőség egy másik pillére a kódolói megbízhatóság. Ez biztosítja, hogy a kódolás folyamata megismételhető és konzisztens, nem pedig a kutató szubjektív benyomásaitól függ. A gyakorlatban ez azt jelenti, hogy legalább két, egymástól függetlenül dolgozó kódoló egyezését mérjük az adatok egy részhalmazán (jellemzően 10-20%-án). Az egyezést leggyakrabban a Cohen-kappa vagy a Krippendorff-alfa mutatóval számszerűsítjük, ahol a 0,8 feletti érték tekinthető a közös, megbízható értelmezés jelének. Ha az első körös megbízhatóság 0,5 alatt van, az általában nem a kódolók figyelmetlenségét jelzi, hanem a kódkönyv alapvető problémáira utal. A bevált gyakorlati protokoll iteratív, tehát a kódolók a nézeteltérések megbeszélése után finomítják a kódkönyvet, majd egy új adagrészleten ismét ellenőrzik az egyezést, egészen a kívánt szint eléréséig.

A következő minőségi biztosíték a modellezési folyamat átláthatósága és érvényessége. Itt a legfontosabb a sztanzadöntés megalapozottsága. Mivel a sztanza mérete alapvetően meghatározza az eredményeket, ennek a döntésnek elméletileg és empirikusan is indokoltnak kell lennie. Nem elfogadható az a gyakorlat, hogy a kutató a szoftver alapbeállítását használja indoklás nélkül. A helyes eljárás az, ha a döntést az adatok természetére alapozzuk (például „Előzetes elemzésünk kimutatta, hogy a résztvevők átlagosan három-négy beszédfordulóban fejtik ki egy gondolatmenetüket, ezért négyes mozgó ablakot alkalmaztunk."), és ideális esetben érzékenységi elemzéssel is teszteljük, hogy a különböző ablakméretek mennyire változtatják meg a végső modellt. A modellépítés során intő jel lehet, ha a hálózatban minden mindennel kapcsolódik

(valószínűleg túl nagy az ablak), vagy éppen szinte semmi sem kapcsolódik (túl kicsi az ablak). Szintén probléma, ha egyetlen, túl általános kód dominálja a hálózatot, vagy ha a megmagyarázott variancia túl alacsony (<60%), ami azt jelzi, hogy a kétdimenziós ábrázolás túlságosan egyszerűsíti az adatainkban rejlő bonyolultságot.

A minőségi ENA kutatási jelentésben az interpretáció többszintű validálása szükséges. Nem elég vizuálisan különbözőnek látszó hálózatokat bemutatni, hanem a különbségeket statisztikailag is alá kell támasztani. A minimális elvárás a csoportok centroidjainak összehasonlítása (például t-próbával) és a hatásméret (például Cohen-féle *d* segítségével) közlése. Az interpretáció erejét jelentősen növeli a trianguláció, azaz az eredmények megerősítése több oldalról. A módszertani trianguláció során például összevethetjük, hogy az ENA-ban talált erős kapcsolatok megfelelnek-e egy hagyományos tematikus elemzés főbb témáinak. Az adatforrás-trianguláció pedig azt vizsgálja, hogy egy interjúk alapján felállított modell mintázatai megjelennek-e például a résztvevők által készített dokumentumokban is.

Végül, de nem utolsósorban a minőségbiztosítás része az átláthatóság, a reprodukálhatóság és az etikai megfontolások kezelése. A kutatás reprodukálhatóságát egy részletes dokumentáció segíti, amely tartalmazza a teljes kódkönyvet, a megbízhatósági mutatókat, a sztanzaparaméterek indoklását és a statisztikai tesztek részletes eredményeit.

5. Mikor használjuk és mikor ne használjuk a módszert?

5.1. A kutatási kérdés

A módszer azokra a kérdésekre a legalkalmasabb, amelyek a kapcsolatokra, struktúrákra és a gondolkodás vagy a diskurzus belső mintázataira összpontosítanak, nem pedig gyakoriságokra vagy direkt ok-okozati viszonyokra. Az ideális ENA-kérdések a „hogyan" típusú kérdések, amelyek a tudás szerveződését, a diskurzus szerkezetét vagy a gondolkodásbeli fejlődést vizsgálják. Ahelyett, hogy azt kérdeznénk, „Melyek a leggyakoribb témák a vitában?", egy ENA-kérdés inkább így hangzik: „Hogyan kapcsolódnak össze a központi érvek és

ellenérvek a vita során a különböző felszólalóknál?" Ahelyett, hogy „X tényező okozza-e Y eredményt?", azt kérdezzük: „Miben különbözik a sikeres és a sikertelen csapatok kommunikációs hálózata egy probléma megoldása során?" Példánkban a kiinduló kérdésünk nem egyszerűen az, hogy egy tanár milyen pedagógiai eszközöket használ, hanem az, hogy ezek hogyan alkotnak rendszert. A kutatási kérdésünk tehát: „Hogyan különbözik a szakértő és kezdő tanárok gondolkodási struktúrája egy problémás osztálytermi helyzet kezelése során?" Természetesen egy valós kutatásban előre tisztázni kellene, hogy mit értünk gondolkodási struktúra alatt (esetünkben ez azt jelentette, a tanárok hogyan kapcsolják össze a pedagógiai elveket, a tanulók egyéni szükségleteit, az osztálytermi dinamikát és a konkrét beavatkozási stratégiákat).

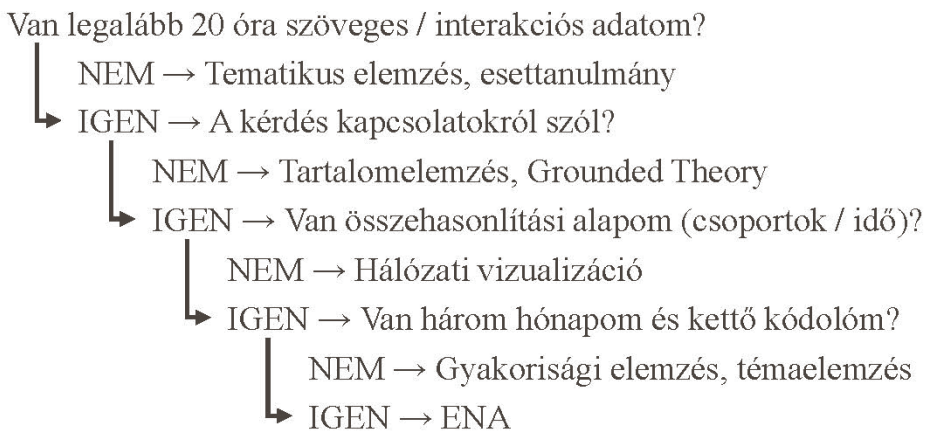
5.2. A minta

Az ideális ENA-adatok gazdag, természetes diskurzusból származnak, és szekvenciális természetűek. Legyen szó interjúról, spontán beszélgetésről vagy online fórum-interakcióról, fontos, hogy az adatok időben vagy logikailag rendezettek legyenek, összetett, többrétegű tartalmat hordozzanak, és elegendő mennyiségben álljanak rendelkezésre – a gyakorlat azt mutatja, hogy bár egyszerűbb modellek már néhány órányi átiratból (körülbelül 50+ oldal) is készíthetők, a robusztusabb, összetettebb elemzésekhez jelentősebb adatmennyiség (akár 20+ óra) szükséges. Alkalmatlanok az ENA-ra a strukturált, előre meghatározott válaszlehetőségeket tartalmazó adatok, mint a Likert-skálás kérdőívek vagy az igen/nem válaszok. Ugyanígy problémásak az elszigetelt, kontextus nélküli adatpontok, a túl rövid, fragmentált szövegek vagy az olyan statikus szövegek, mint egy tankönyv, amelyekből hiányzik a kölcsönhatás és az időbeli dinamika, ami a kapcsolatok értelmezésének alapfeltétele.

5.3. A kutatás jellege

Az ENA egy idő- és munkaigényes módszer. Reális választás, ha a kutatócsoport rendelkezik a szükséges feltételekkel. Jellemzően két-három hónapra van szükség egy teljes elemzési ciklushoz az adat-előkészítéstől az interpretációig. A minőségbiztosításhoz elengedhetetlen legalább két képzett kódoló, akik biztosítják a kódolás megbízhatóságát. A kutatócsoportnak jártasnak kell lennie a strukturálatlan adatok elemzési módszereiben. Nem praktikus az ENA választása sürgős, néhány hetes projektek esetén, ha a kutató egyedül dolgozik,

vagy nincs tapasztalata a kvalitatív kódolásban. Bár a módszer jelenleg időigényes, érdemes megjegyezni, hogy a mesterséges intelligencia fejlődése változást hozhat. A mesterséges intelligencia terjedése a hermeneutikai elemzésben a jövőben enyhítheti ezeket a korlátokat, és szélesebb körben is elérhetővé teheti a módszert.



9. ábra: Döntési fa ENA-elemzéshez

A módszer gyakorlati alkalmazhatóságát legjobban talán a kutatási tapasztalatok szemléltetik.

a) Egy kutatócsoport 12 sebészeti csapat közlésfolyamatait vizsgálta négyórás, összetett műtétek során. A 48 órányi átiratot nyolc kóddal elemezték, amelyek az anatómiai referenciáktól a technikai megoldásokon át a kockázatértékelésig terjedtek. Az ENA révén kimutatták, hogy a sikeres csapatoknál az anatómiai referenciák szorosan kapcsolódnak a kockázatértékeléshez, míg a kevésbé sikeres csapatoknál ezek az elemek elszigetelten jelentek meg. A módszer azért működött kiválóan, mert gazdag interakciós adatokkal dolgoztak, világos összehasonlítási alapjuk volt (sikeres kontra problémás műtétek), és a kutatási kérdés pontosan a kommunikáció struktúrájára irányult.
b) Egy kutatócsoport 20 középiskolai fizikatankönyv szerkezetét próbálta összehasonlítani ENA-val. A projekt kudarcba fulladt, mert a tankönyvszövegek statikusak, nincs bennük valódi kölcsönhatás vagy időbeli szekvencia. Az ENA hálózatai értelmezhetetlenek voltak, mivel a módszer

alapfeltevése – hogy az együttes megjelenés kognitív kapcsolatot jelez – nem teljesült. A strukturális tartalomelemzés sokkal megfelelőbb lett volna erre a célra.

c) Egy harmadik példában egy vállalat 200 egyoldalas teljesítményértékelését próbálták elemezni ENA-val. A szövegek azonban túl rövidek és sablonszerűek voltak, kevés variációval. A kódok között alig mutatkozott értelmezhető kapcsolat, a hálózatok gyakorlatilag megkülönböztethetetlenek voltak. A kutatók később szentimentelemzéssel és egyszerű kategorizálással sokkal használhatóbb eredményeket értek el.

A gyakorlatban gyakran találkozunk olyan helyzetekkel, ahol nem egyértelmű, hogy az ENA megfelelő választás-e. Ilyenkor érdemes kombinált vagy alternatív megközelítésekben gondolkodni. Vegyük például azt az esetet, amikor azt vizsgáljuk, hogyan tanulnak a diákok programozni. Itt van diskurzus (a diákok megbeszélik a problémákat), de a kód-output minősége is fontos. A megoldás egy vegyes megközelítés lehet, ahol az ENA feltárja a gondolkodási folyamat struktúráját, ezt kiegészítjük a programkód minőségének kvantitatív értékelésével, majd regresszióval vizsgáljuk, hogy az ENA-térben elfoglalt pozíció bejósolja-e a programozási teljesítményt. Hasonló dilemma merül fel longitudinális vizsgálatoknál. Ha egy év alatt háromszor-négyszer mérjük a tanárok reflektív gyakorlatát, ez kevés egy valódi idősorelemzéshez, de elég lehet egy longitudinális ENA-hoz, feltéve, hogy minden időpontban gazdag adatunk van. Ha viszont az adatok szegényesek, jobb lehet a hagyományos hermeneutikai elemzés kombinálva egyszerű változáselemzéssel.

A skálázás is gyakori kihívás. Képzeljük el, hogy ezer orvos-beteg konzultációt kellene elemeznünk. Ez túl sok a kézi kódoláshoz, de túl összetett a teljes automatizáláshoz. Vegyes megoldás lehet, hogy először gépi előkódolást végzünk természetes nyelvfeldolgozással a főbb témákra, majd egy százas almintán emberi finomkódolást alkalmazunk, ezen futtatjuk az ENA-t, és az eredményeket óvatosan általánosítjuk a teljes mintára. A következő táblázatban különböző módszereket hasonlítunk össze a fókusz, az adatigény és egyéb szempontok mentén.

Szempont	ENA	TA	SNA	TM	QCA
Fókusz	Kognitív kapcsolatok	Témák, jelentések	Társas kapcsolatok	Témák eloszlása	Oksági konfigurációk
Adatigény	Közepes-nagy	Kicsi-közepes	Hálózati adatok	Nagy	Közepes
Időigény	2-3 hónap	1-2 hónap	1 hónap	1-2 hét	2-3 hónap
Kódolás	Szükséges	Szükséges	Nem kell	Nem kell	Szükséges
Vizualizáció	Kiváló	Korlátozott	Kiváló	Jó	Közepes
Statisztika	Van	Nincs	Van	Van	Van

16. táblázat: Módszerek összehasonlítása

TM: témaelemzés, SNA: szociális hálózatelemzés, TA: tematikus elemzés (itt az átfogó értelemben, ernyőfogalomként)

A módszerválasztás végső soron három alapkérdésen múlik. Először is, elég gazdag-e az adatunk ahhoz, hogy értelmes kapcsolatokat találjunk benne? Másodszor, van-e olyan összehasonlítási alapunk, amely értelmet ad a struktúrák különbségeinek? Harmadszor, rendelkezünk-e a szükséges idővel és erőforrásokkal a módszer igényeinek kielégítésére? Ha mindhárom kérdésre igennel tudunk válaszolni, az ENA valószínűleg jó választás. Ha egy vagy több kérdésre nemleges a válasz, érdemes megfontolni az alternatívákat vagy a kombinált megközelítéseket. Végső soron ne feledjük: a legjobb módszer mindig az, ami a kutatási kérdésünkre választ tud adni, nem pedig az, amelyik a legbonyolultabb vagy a legújabb. Az ENA egy eszköz a sok közül – hatékony és elegáns eszköz, de nem egyetemes megoldás minden kutatási problémára.

6. A módszer változatai és haladó alkalmazásai

A longitudinális ENA talán a legintuitívabb kiterjesztése az alapmodellnek. Ahelyett, hogy különböző csoportokat hasonlítanánk össze egy időpontban, ugyanazt a csoportot vagy egyént követjük nyomon az időben, feltárva a gondolkodási struktúra evolúcióját. A megközelítés lényege, hogy az adathalmazt

értelmes időszakokra bontjuk – ezek lehetnek egy szemeszter hetei, egy projekt szakaszai vagy akár egy problémamegoldási folyamat percei. Minden időszakot úgy kezelünk, mintha külön csoport lenne, de mivel ugyanabban az ENA-térben ábrázoljuk őket, a centroidok elmozdulása egy „pályavonalat" rajzol ki. A pályavonal nem egyszerűen változást mutat, hanem a változás irányát és természetét is feltárja. A longitudinális ENA technikai és értelmezési kihívásokat is felvet. A legfőbb probléma a pályavonalak olvashatósága, mivel a kutatónak egyszerre három változót kell értelmeznie: a hálózat pozícióját az x és y tengelyen, valamint ezek változását az időben. Ez a háromdimenziós értelmezési feladat teszi a pályavonalakat nehezen olvashatóvá. A legfontosabb döntés az időablakok mérete, mert a túl rövid ablakok zajosak, a túl hosszúak elmossák a finomabb változásokat. A legújabb megjelenítési megoldások, mint a koregisztrált subplotok használata éppen ezt a problémát igyekeznek megoldani azzal, hogy a dimenziónkénti változásokat külön-külön, de a fő térképpel összehangoltan ábrázolják.

A multimodális ENA felismeri, hogy a gondolkodás és a tanulás nem csak szóban megnyilvánuló vagy leképeződő jelenségek. Egy sebészeti beavatkozás során nem elég elemezni, mit mond az orvosi csoport – azt is látni kell, hová néznek, hogyan mozognak, milyen eszközöket használnak. A multimodális ENA ezeket a különböző adatfolyamokat integrálja egyetlen modellbe. A technikai megvalósítás bonyolult folyamat. Először is szinkronizálni kell a különböző modalitásokat, például a hangfelvételt, a szemkövető adatokat és a mozgásrögzítést. Majd mindegyiket kódolni kell – de míg a beszéd kódolása viszonylag egyértelmű, a nem szóbeli jelek kódolása új kihívásokat vet fel. Mikor számít jelentőségteljesnek egy pillantás? Hogyan különböztetjük meg a funkcionális és a kommunikatív gesztusokat? A multimodális ENA etikai és gyakorlati kérdéseket is felvet. A résztvevők folyamatos, többcsatornás megfigyelése invazív lehet. A hatalmas adatmennyiség feldolgozása számítási kihívás. És talán a legfontosabb kérdés az, hogy hogyan értelmezzük a crossmodális kapcsolatokat? Ha valaki másra néz, miközben egy ötletet magyaráz, az figyelmetlenség vagy éppen a társas validáció keresése?

Módszertani szempontból fontos a különbségtétel a longitudinális megközelítés és az ONA között. Míg a longitudinális ENA az időt makroszinten kezeli – azt vizsgálva, hogyan változik meg a teljes hálózati struktúra és annak súlypontja (centroidja) különböző időpontok (például hetek vagy hónapok) között –, addig az ONA az időt mikroszinten, a diskurzus belső szekvenciájaként értelmezi. Míg előbbi a „mi változott?" kérdésre fókuszál a fejlődési pályák ábrázolásával, utóbbi a „milyen sorrendben történt?" problémáját válaszolja meg azáltal, hogy irányított kapcsolatokat hoz létre a kódok között.

Az Ordered Network Analysis (ONA) tehát az ENA egy olyan változata, amely megőrzi az események időbeli sorrendjét. Míg az alapmodell az adjacenciamátrix alapján szimmetrikus, irányítatlan hálózatokat hoz létre (az A-B kapcsolat ugyanaz, mint a B-A), az ONA aszimmetrikus adjacenciamátrixot használ, és így irányított, súlyozott hálózatokat generál, megkülönböztetve, hogy melyik elem előzi meg a másikat. Ez fontos olyan kontextusokban, ahol a sorrend lényeges. Egy diagnosztikus folyamatban például kritikus különbség, hogy a tünetek értékelése vezet-e a hipotézisalkotáshoz, vagy fordítva, a hipotézis befolyásolja a tünetek értelmezését. Az ONA ezt a különbséget láthatóvá teszi a hálózatban. Az irányítottságot nem a hagyományos nyilakkal, hanem egy „broadcast" modellel jelöli, ahol a kapcsolatot jelző él egy háromszöget ad ki. A háromszög csúcsa a kiinduló, a talpa pedig az érkező kód felé mutat, a vonal vastagsága pedig a temporális kapcsolat erősségét jelzi.

A directed ENA és a means rotation olyan technikai fejlesztések, amelyek sajátos kutatási kérdések megválaszolását segítik. A directed ENA lehetővé teszi, hogy a kutató előre meghatározza, mely kapcsolatok a legfontosabbak, és az ENA-tér tengelyeit ezekhez igazítsa. Ez hasznos hipotézisvezérelt kutatásokban, ahol konkrét elméleti várakozásaink vannak arról, mely kapcsolatok különböztetik meg a csoportokat. A means rotation ezzel szemben matematikai optimalizációt használ, és úgy forgatja az ENA-teret, hogy a csoportok közötti különbség maximális legyen az első dimenzióban.

A módszer legizgalmasabb fejleményei azok, ahol az ENA iteratív ciklusban működik más módszerekkel. Egy háromfázisú modell első szakaszában grounded theoryval azonosítják a releváns kódokat. A második szakaszban ENA-val modellezik a struktúrákat. A harmadik szakaszban a deviáns eseteket – amelyek kilógnak az ENA-modellből – kvalitatívan elemzik, ami új kódok felfedezéséhez vezet. Ez az iteratív ciklus addig folytatódik, amíg telítődés nem következik be. Egy másik úttörő megközelítés az ENA kombinálása gépi tanulással. Az ENA-térben elfoglalt pozíció prediktív változóként szolgálhat: előre jelzi-e a tanulói sikerességet? Ha igen, mely kapcsolatok a meghatározók?

A legújabb fejlesztések közé tartozik a valós idejű ENA, amely élő kölcsönhatások azonnali elemzését teszi lehetővé. Képzeljünk el egy online kiscsoportos oktatási helyzetet, ahol a segítségül hívott rendszer valós időben elemzi a diák gondolkodási struktúráját, és ennek alapján határozza meg a segítségnyújtás módját, tartalmát és formáit. A személyre szabott tanulási útvonalak tervezésében az ENA prediktív modelljei kiemelkedő szerepet játszhatnak. Ha tudjuk, hogy egy adott gondolkodási struktúra vezet a sikerhez, akkor azonosíthatjuk, mely kapcsolatok hiányoznak egy diák gondolkodási hálózatából, és célzott beavatkozásokat tervezhetünk a hiányzó elemek kiépítésére.

A nagy adathalmazok kezelése is áttörés előtt áll. A hagyományos ENA néhány száz résztvevővel működik jól, de mi történik, ha a minta mérete több mint tízezer? A hierarchikus ENA lehetővé teszi a többszintű modellezést, azaz az egyéni, csoportos és populációs szintű struktúrák egyidejű elemzését. Végül, a mesterséges intelligenciával támogatott kódolás forradalmasíthatja a módszert. Képzeljük el, hogy egy nyelvi modell előkódolja az adatokat, amelyeket aztán emberi szakértők finomítanak. Ez nagyságrendekkel növelhetné az elemezhető adatmennyiséget.

7. Összefoglalás

Az episztemikus hálózatelemzés (ENA) a kvantitatív etnográfia egy központi módszere, amely a strukturálatlan adatok gazdagságát megőrizve teszi lehetővé kognitív és kommunikációs struktúrák matematikai modellezését és képi megjelenítését. A módszer három pillérre épül: az elosztott megismerés elméletére, a szituatív tanulás eszméjére és az episztemikus keretek fogalmára. Az ENA alapvető újítása, hogy a hogyan kérdésre összpontosít, és nemcsak azt mutatja meg, hogy milyen elemek jelennek meg egy diskurzusban, hanem azt is, hogy ezek hogyan kapcsolódnak össze a résztvevők gondolkodásában. Ez a megközelítés értékes a tanulás és a bonyolult kognitív folyamatok vizsgálatában. A módszer gyakorlati alkalmazása öt fő lépésből áll: adatgyűjtés és előkészítés, kódolás, sztanzaelemzés, értelmezés és eredmények validálása. A lépések közül talán a sztanzaelemzés a legösszetettebb, mivel itt dől el, hogy milyen kapcsolatokat tekintünk érvényesnek. A minőségbiztosítás központi elemei a kódolási megbízhatóság, a modellezési folyamat átláthatósága és az interpretáció többszintű validálása. Az ENA nem egyetemes módszer. Ideális alkalmazási területe a gazdag, szekvenciális diskurzusok elemzése, ahol a kutatási kérdés a struktúrákra és kapcsolatokra irányul. Alkalmatlan strukturált adatokra, ok-okozati viszonyok feltárására vagy gyors, exploratív elemzésekre. A módszer időigényes voltát ellensúlyozza az, hogy olyan mélységű megértést tesz lehetővé, amely más módszerekkel nehezen elérhető. Fejlődése a következő irányokban zajlik: a longitudinális ENA a hálózatok globális, időbeli elmozdulását és fejlődési pályáit követi nyomon (makroszint), a multimodális ENA a szóbeli és nem szóbeli adatok együttes kezelését teszi lehetővé, az ONA pedig a

hálózaton belüli kapcsolatok mikroszintű sorrendiségét és irányát – vagyis azt, hogy melyik elem követi a másikat a folyamatban – modellezi. A mesterséges intelligencia térnyerése várhatóan forradalmasítani fogja a kódolási folyamatot, jelentősen csökkentve az idő- és munkaigényt. Az ENA lehetővé teszi, hogy a gazdag, kontextualizált megértést statisztikailag is alátámasztott, reprodukálható formában jelenítsük meg. Ez értékes olyan területeken, mint az oktatás, az egészségügy vagy a szervezetfejlesztés, ahol az egyedi esetek mélyreható megértése és az általánosítható minták azonosítása egyaránt szükséges.

8. Önellenőrző kérdések

1. Mi a legsúlyosabb hiba, amelyet egy kutató a sztanzaelemzés során elkövethet, és miért az?

Válasz: A legsúlyosabb hiba, ha a kutató a sztanzaparamétereket (az ablak méretét és a szövegegységek meghatározását) nem elméleti megalapozottsággal, hanem önkényesen vagy kényelmi szempontok alapján határozza meg (például a szoftver alapbeállítását használja magyarázat nélkül). Ez azért hiba, mert a sztanzadöntés alapvetően meghatározza, hogy milyen kapcsolatokat tekintünk érvényesnek a gondolkodási szerkezetben. Egy rosszul választott ablakméret torzíthatja az egész hálózatot: túl kicsi ablak töredezett, túl nagy ablak pedig elmosott kapcsolatokat eredményez. A helyes eljárás az elméleti indoklás (például „a résztvevők átlagosan három-négy beszédfordulóban fejtenek ki egy gondolatot") és a robusztussági vizsgálat.

2. Mi a legfontosabb különbség a kódolási megbízhatóság és a statisztikai szignifikancia között az ENA-eredmények értékelésekor?

Válasz: A kódolási megbízhatóság (például Cohen-kappa) a kvalitatív alapok szilárdságát méri: azt mutatja meg, hogy a kódolók konzisztensen ugyanazokat a fogalmakat azonosítják-e a szövegben. Ez a hálózat érvényességének alapfeltétele. Ezzel szemben a statisztikai szignifikancia (például t-próba a centroidok között) azt méri, hogy a már kódolt adatok alapján feltárt strukturális különbségek véletlennek tekinthetők-e. Egy magas kódolási megbízhatóság nélkül

a statisztikai szignifikancia értelmetlen, mivel rossz alapokra épül. Fordítva: jó kódolási megbízhatóság mellett sem biztos, hogy szignifikáns különbségeket találunk a csoportok között.

3. Milyen típusú kutatási kérdés esetén választaná az ENA-t egy hagyományos tematikus elemzéssel szemben?

Válasz: Az ENA-t akkor választanám, ha a kutatási kérdésem a kapcsolatokra, a struktúrákra és a gondolkodás szerveződésére összpontosít. Például: „Hogyan különbözik a szakértő és kezdő orvosok diagnosztikai gondolkodásának szerkezete?" Ezzel szemben a tematikus elemzést akkor választanám, ha a tartalmakra, a jelentésekre és a témák azonosítására lennék kíváncsi. Például: „Milyen főbb témák jelennek meg az orvos-beteg konzultációkban?" Az ENA a hogyan szerveződik kérdést válaszolja meg, a tematikus elemzés a miről beszélnek kérdést. Az ENA matematikai modellt és képi megjelenítést ad, a tematikus elemzés értelmező leírást és kategóriákat.

4. Mit jelent a gyakorlatban, ha egy ENA-vizualizációban egy kapcsolat vastag vonallal van jelölve, de a két csomópont távol helyezkedik el egymástól a térben?

Válasz: Ez ellentmondást jelez az adatokban vagy a modellben. A vastag vonal azt mutatja, hogy a két kód gyakran fordul elő együtt az egyes sztanzákban (erős lokális kapcsolat), de a térben való távolság azt jelzi, hogy általában különböző típusú hálózatokban/kontextusokban jelennek meg (gyenge globális asszociáció). Ez gyakran a heterogén minta jele: lehet, hogy az egyik alcsoport számára ez a kapcsolat természetes, a másik alcsoport számára szokatlan. A kutató számára ez egy fontos jel, hogy mélyebben kell elemezni az adatokat, esetleg alcsoportokat kell azonosítani, vagy újra kell gondolni a kódolási sémát.

5. Milyen előnyeit és korlátait ismeri fel a longitudinális ENA-nak egy hagyományos longitudinális kvalitatív elemzéssel összehasonlítva?

Válasz: A longitudinális ENA fő előnye, hogy matematikailag pontos képet ad a gondolkodási struktúra időbeli változásairól: a centroidok pályavonala számszerűsíthetően mutatja a változás irányát és mértékét. A képi megjelenítés azonnal láthatóvá teszi, hogy mely kapcsolatok erősödnek vagy gyengülnek az időben. Korlátja azonban, hogy a háromdimenziós értelmezés (x tengely, y tengely, idő) összetett lehet, és könnyen elveszhetnek a kontextuális árnyalatok.

A hagyományos longitudinális kvalitatív elemzés gazdagabb értelmezéseket ad a változás okairól és jelentéséről, de nehéz belőle konkrét, összehasonlítható mintázatokat azonosítani. A legjobb megközelítés gyakran a kettő vegyítése: ENA a struktúrák változásának feltárására, hermeneutikai elemzés a jelentések megértésére.

Ajánlott irodalom

Brohinsky, J., Marquart, C., Wang, J., Ruis, A. R., & Shaffer, D. W. (2021). Trajectories in Epistemic Network Analysis. In A. R. Ruis & S. B. Lee (Eds.), *Communications in Computer and Information Science. Advances in Quantitative Ethnography* (Vol. 38, pp. 106–121). Springer International Publishing. https://doi.org/10.1007/978-3-030-67788-6_8

Chang, H.-Y., & Tsai, C.-C. (2023). Epistemic Network Analysis of Students' Drawings to Investigate Their Conceptions of Science Learning with Technology. *Journal of Science Education and Technology*. Advance online publication. https://doi.org/10.1007/s10956-022-10026-9

Charmaz, K. (2025). *Constructing grounded theory. Introducing qualitative methods*. SAGE.

McLevey, J., Scott, J., & Carrington, P. J. (Eds.). (2024). *The Sage handbook of social network analysis*. SAGE.

Ruis, A. R., & Lee, S. B. (Eds.). (2021). *Communications in Computer and Information Science. Advances in Quantitative Ethnography*. Springer International Publishing. https://doi.org/10.1007/978-3-030-67788-6

Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics, 3*(3), 9–45. https://doi.org/10.18608/jla.2016.33.3

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., & Mislevy, R. (2009). Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning. *International Journal of Learning and Media, 1*(2), 33–53. https://doi.org/10.1162/ijlm.2009.0013

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic Network Analysis: A Worked Example of Theory-Based Learning Analytics. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), *Handbook of Learning Analytics* (pp. 175–187). Society for Learning Analytics Research. https://doi.org/10.18608/hla17.015

Siebert-Evenstone, A. L., Arastoopour Irgens, G., Collier, W., Swiecki, Z., Ruis, A. R., & Williamson Shaffer, D. (2017). In Search of Conversational Grain Size: Modeling Semantic Structure using Moving Stanza Windows. *Journal of Learning Analytics, 4*(3). https://doi.org/10.18608/jla.2017.43.7

Swiecki, Z., Ruis, A. R., Farrell, C., & Shaffer, D. W. (2020). Assessing individual contributions to Collaborative Problem Solving: A network analysis approach. *Computers in Human Behavior, 104*, 105876. https://doi.org/10.1016/j.chb.2019.01.009

Tan, Y., Swiecki, Z., Ruis, A. R., & Shaffer, D. (2024). Epistemic Network Analysis and Ordered Network Analysis in Learning Analytics. In M. Saqr & S. López-Pernas (Eds.), *Learning Analytics Methods and Tutorials* (pp. 569–636). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54464-4_18

Viberg, O., Baars, M., Mello, R. F., Weerheim, N., Spikol, D., Bogdan, C., Gasevic, D., & Paas, F. (2024). Exploring the nature of peer feedback: An epistemic network analysis approach. *Journal of Computer Assisted Learning, 40*(6), 2809–2821. https://doi.org/10.1111/jcal.13035

Zhao, L., Echeverria, V., Swiecki, Z., Yan, L., Alfredo, R., Li, X., Gasevic, D., & Martinez-Maldonado, R. (2024). Epistemic Network Analysis for End-users: Closing the Loop in the Context of Multimodal Analytics for Collaborative Team Learning. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 90–100). ACM. https://doi.org/10.1145/3636555.3636855

VII. A TERMÉSZETES-NYELV-FELDOLGOZÁS NÉHÁNY MÓDSZERE (NLP)

TURÓS MÁTYÁS

A fejezet a következő kérdéseket válaszolja meg:

a) Milyen elméleti alapokon nyugszik a természetesnyelv-feldolgozás (NLP), és hogyan alakítja át a szöveget elemezhető adattá?
b) Melyek a legfontosabb adat-előkészítési lépések, és miért jelentenek ezek kritikus módszertani döntéseket?
c) Hogyan működik a gyakorlatban a névelem-felismerés (NER), a szentimentelemzés és a témaelemzés?
d) Milyen minőségbiztosítási elveket kell követni a felügyelt és felügyeletlen NLP-modellek értékelésekor?
e) Mikor érdemes és mikor nem tanácsos NLP-módszereket választani egy társadalomkutatási kérdés megválaszolására?
f) Milyen haladó NLP-technikák (például embeddingek, transzformerek) forradalmasítják a szövegelemzést?

1. Bevezetés

A természetesnyelv-feldolgozás (Natural Language Processing, NLP) a mesterséges intelligencia, az informatika és a nyelvészet határterületén helyezkedik el, és ma már az adattudomány (data science) egyik legfontosabb és legdinamikusabban fejlődő ágát képezi. Szorosan épít a gépi tanulás statisztikai módszereire, hogy a szövegekben rejlő mintázatokat feltárja, ugyanakkor gyakran merít a számítógépes nyelvészet eredményeiből is, amely a nyelvi struktúrák formális leírására összpontosít. Ez a fejezet kifejezetten a társadalomkutatók számára lényeges, gyakorlatias NLP-eszközökre összpontosít. Míg a korábbi fejezetekben bemutatott részben vagy egészében hermeneutikai módszerek, mint a reflexív tematikus elemzés (RTA) vagy a szemantikus tartalomelemzés (SCA), kiváló eszközöket nyújtanak a mély, kontextuális megértéshez néhány interjú vagy dokumentum esetében, addig a kutatók egyre gyakrabban szembesülnek olyan korpuszokkal, amelyek méretüknél fogva meghaladják a kézi elemzés kapacitását. Több száz interjú, több ezer online hozzászólás vagy egy teljes parlamenti ciklus felszólalásainak feldolgozása hermeneutikai eszközökkel szinte lehetetlen. Erre a kihívásra ad választ a természetesnyelv-feldolgozás, amely a számítógépes nyelvészet, az informatika és a mesterséges intelligencia eszköztárát kínálja a szöveges adatok szisztematikus és nagy léptékű elemzéséhez. Az NLP tehát egy olyan módszercsalád, amely a társadalomkutató számára eredendően értelmezhetetlen, strukturálatlan szöveges adatokat olyan strukturált formátumba rendezi, amely már alkalmas a statisztikai elemzésre. A társadalomtudományi gyakorlatban ezeket az eljárásokat gyakran kvantitatív szövegelemzésnek (Quantitative Text Analysis, QTA) vagy szövegbányászatnak (text mining) is nevezik, és ezen a néven a magyar szakirodalomban is meghonosodtak.

A természetesnyelv-feldolgozás helyének és szerepének megértéséhez még azt is fontos látni, hogy a szövegelemzés módszerei egy kontinuumon helyezhetők el, amelynek egyik végpontján a tisztán kézi kódolás, középen a géppel támogatott kvalitatív adatelemzés (Computer-Assisted Qualitative Data Analysis, CAQDAS), a másik végponton pedig a teljesen automatizált gépi elemzés áll. Míg a korábbi fejezetek a kézzel végzett és a kvalitatív adatelemző (QDA) szoftverekkel – mint például az ATLAS.ti vagy a MAXQDA – támogatott elemzések logikáját mutatták be, jelen fejezet ezen a kontinuumon tesz egy lépést, és a harmadik kategóriára, a teljesen gépi elemzésre összpontosít.

A gépi elemzés felé tett lépés egy kontinuum mentén értelmezhető. A skála egyik végén a legegyszerűbb, gyakoriságalapú vizsgálatok állnak. Ezek alapja a „szózsákmodell" (Bag-of-Words, BoW), amely – ahogy azt a következőkben részletesen is látni fogjuk – a szöveget úgy kezeli, mint egy zsák szót, és a nyelvtani szerkezetet és a szórendet figyelmen kívül hagyva pusztán azt számolja, hogy mely szavak milyen gyakran fordulnak elő. Bár ez a megközelítés hatékony lehet a dokumentumok fő témáinak azonosítására, alapvető korlátja, hogy minden szót elszigetelt egységként kezel. Nem képes megragadni a szavak közötti jelentésbeli viszonyokat, és számára a „király", az „uralkodó" és a „káposzta" csupán három, egymástól teljesen független elem.

A modern NLP forradalma éppen ebben az elmozdulásban rejlik, mert a puszta gyakoriságszámlálástól eljutott azokig a módszerekig, amelyek a szavak kontextusbeli jelentését is képesek megragadni. Ezt a technológiai ugrást a szóbeágyazás (word embedding) tette lehetővé. Ennek lényege, hogy a szavakat már nem egyszerűen megszámláljuk, hanem sűrű, numerikus vektorokként (számsorokként) ábrázoljuk egy többdimenziós térben. A modell a szavak jelentését a felhasználási környezetükből tanulja meg, így a hasonló kontextusban használt szavak (például „király" és „uralkodó") vektorai matematikailag is „közel" kerülnek egymáshoz a szemantikai térben.

Ahogy látni fogjuk, a modern NLP ereje nagyrészt ezen a kifinomultabb szemantikai reprezentáción múlik. A vektoriális ábrázolás teszi ugyanis lehetővé, hogy a gép olyan, korábban elképzelhetetlen feladatokat is végrehajtson, mint a szinonimák azonosítása vagy akár analógiás kapcsolatok feltárása. Az NLP célja tehát nem a hermeneutikai munka segítése, hanem a szövegekben rejlő mintázatok nagy léptékű, matematikai elemzése.

E fejezetben az NLP-t a társadalomkutató számára elérhető, gyakorlatias eszköztárként mutatjuk be. Az NLP központi szerepe a könyvünk logikájában, hogy hídként funkcionál: egy olyan eljáráscsalád, amely képes a hermeneutikai kutatások nyers, strukturálatlan adatait (például interjúleiratokat, nyílt végű kérdőíves válaszokat) olyan strukturált adattá alakítani (például szemantikai hálózatokká, számszerűsített véleménymutatókká, a szózsákmodell segítségével pedig dokumentum-term mátrixokká), amely már további, akár matematikai elemzések számára is hozzáférhetővé válik. Az előállított strukturált adatok ugyanis közvetlen bemenetként szolgálhatnak a könyv későbbi, sőt korábbi fejezeteiben tárgyalt eljárásokhoz: egy témaelemzés (Topic Modeling, TM) eredményeiből származó kategóriák képezhetik egy kvalitatív összehasonlító elemzés (QCA) feltételeit, a fogalmak együttes előfordulásai pedig egy episztemikus hálózatelemzés (ENA) alapját adhatják.

Végül megjegyezzük, hogy míg a számítógépes szövegelemzés korábban szinte kizárólag programozási ismeretekkel rendelkező szakemberek kiváltsága volt, a terület fejlődése – a grafikus felületű szoftverek (például Orange, Voyant Tools) megjelenése – demokratizálta ezen eszközök használatát. Ma már azok a társadalomkutatók is képesek bonyolult szövegelemzési feladatokat elvégezni, akik egyetlen sor kódot sem szeretnének vagy tudnak megírni, így a figyelmük a technikai kivitelezésről a módszertani tudatosságra és az eredmények kritikus értelmezésére és közlésére helyeződhet át.

2. Elméleti alapok

A természetesnyelv-feldolgozás lényege, hogy a számítógépek számára eredendően értelmezhetetlen, strukturálatlan emberi nyelvet egy olyan strukturált, numerikus formátumba rendezzük, amelyen már statisztikai elemzések hajthatók végre. Az NLP tehát egy módszertani híd, amely lehetővé teszi a statisztikai és gépi tanulási módszerek alkalmazását számszerűsített szövegekre, hogy azokból rejtett mintázatokat és struktúrákat tárjunk fel. Annak megértéséhez, hogy ez a transzformáció mit is jelent valójában, érdemes túllépni a nyelvészet határain. Az NLP-ben használt eljárások ereje abban rejlik, hogy sokszor olyan általános statisztikai elveken alapulnak, amelyek más tudományterületeken is rendkívül hasznosnak bizonyultak. A legjobb példa erre a később részletesen is tárgyalt témaelemzés és annak legnépszerűbb algoritmusa, az LDA (Latent Dirichlet Allocation). Ezt a módszert a szövegelemzésben arra használják, hogy egy nagy dokumentumhalmaz rejtett tematikus struktúráját feltárjuk. Azonban az LDA-hoz hasonló bayesi modelleket néhány évvel korábban, a nyelvészettől teljesen függetlenül, a populációgenetikában is kifejlesztették, ahol arra használták, hogy genetikai adatokból (genotípusokból) azonosítsák a rejtett populációs alcsoportokat. Ez a párhuzamos felfedezés rávilágít az NLP lényegére: a módszer szempontjából mindegy, hogy az adat egy dokumentumban lévő szavak gyakorisága vagy egy egyén genetikai markereinek összessége. Ha az adatot egy dokumentum-term mátrixként (vagy annak genetikai megfelelőjeként) tudjuk reprezentálni, akkor hasonló elvű statisztikai modellek képesek mindkettőből kiolvasni a rejtett struktúrát – legyen az egy politikai diskurzus témája vagy egy ősi európai

genetikai alcsoport. Az NLP tehát a szöveg számszerűsítésének és az így kapott numerikus adatok statisztikai modellezésének tudománya.

A folyamat első és talán legfontosabb lépése, hogy a folyó szöveget elemzési alapegységekre, úgynevezett tokenekre bontsuk. Ez azért szükséges, mert míg a kutató számára egyértelmű, hogy mi egy szó, a számítógép számára ez korántsem az. A tokenizáció folyamata – vagyis a folyó szöveg tokenekre bontása – az NLP első és megkerülhetetlen lépése, ugyanis nem mindegy, hogy a kutya-macska barátság kifejezést három különálló szóként vagy egyetlen egységként kezeljük, vagy hogy az írásjeleket és a ragozott szóalakokat hogyan dolgozzuk fel. Ezért a korszerű NLP-eljárások gyakran nem is szavakkal, hanem annál kisebb, statisztikailag gyakori szórészekkel (szubszavakkal) dolgoznak, hogy hatékonyabban kezeljék az összetett szavakat, a ragozást és az ismeretlen kifejezéseket. A számítógép nem érti a szavak mögötti jelentést vagy a kontextust, számára a szavak csupán különböző karaktersorozatok, amelyek között nincsen inherens jelentésbeli kapcsolat. Az NLP alapvető célja, hogy ezt a szemantikai szakadékot hidalja át, és a szavak viszonyait is megragadhatóvá tegye a gép számára. Az NLP elméleti alapjai ezért arra a transzformációs folyamatra épülnek, amely a szöveget strukturált, numerikus adattá alakítja. E folyamat megértése elengedhetetlen ahhoz, hogy a kutató ne fekete dobozként tekintsen az NLP-eszközökre. Bár a modellek kapják a legnagyobb figyelmet, a társadalomtudományi NLP-kutatásokban a munka több mint 90%-át gyakran maga az adat-előkészítés teszi ki: a források gyűjtése, a szövegek tisztítása, normalizálása és a megfelelő numerikus formátumba helyezése.

Ezek a látszólag technikai lépések olyan kritikus módszertani döntések, amelyek alapvetően meghatározzák, hogy a modell milyen mintázatokat lesz képes felismerni.

A transzformációs folyamatok elméleti alapját a disztribúciós hipotézis adja, amely szerint egy szó jelentését nagymértékben meghatározza az a környezet, amelyben előfordul: „Egy szót a társaságáról ismersz meg." („You shall know a word by the company it keeps." – John Rupert Firth) Ez az alapelv teszi lehetővé, hogy a szavak jelentésbeli hasonlóságait statisztikai alapon, a szavak együttes előfordulásainak vizsgálatával ragadjuk meg. A disztribúciós hipotézist a modern NLP-modellek úgy használják ki, hogy a kontextusból próbálják megjósolni a hiányzó szavakat. Sok erőteljes NLP-módszer tehát egy meglepően egyszerű alapelvre épül: a predikcióra. A modellek hatalmas korpuszokból tanulnak meg nyelvi mintázatokat azáltal, hogy újra és újra megpróbálják megjósolni a következő szót egy adott szövegkörnyezetben. Ez az előrejelző képesség teszi őket alkalmassá arra, hogy a társadalomkutató számára is értékes, rejtett szemantikai és tematikus struktúrákat tárjanak fel.

Ezen a ponton fontos különbséget tenni a modellek két alapvető típusa között. A generatív modellek (mint például a később tárgyalt LDA) azt próbálják leírni, hogyan jön létre egy szöveg egy adott kategóriában. Ezzel szemben a diszkriminatív modellek (mint a legtöbb korszerű szentimentelemző vagy a neurális hálózatok) nem a szöveg keletkezésével foglalkoznak, hanem közvetlenül a kategóriák közötti határvonalak, a különbségek megtanulására összpontosítanak. Míg a generatív modellek mélyebb betekintést engedhetnek az adatok struktúrájába, a diszkriminatív modellek gyakran pontosabbak a konkrét osztályozási feladatokban. A különbség megértéséhez képzeljünk el két szakértőt, akiknek az a feladatuk, hogy felismerjék a macskákat.

a) Az első szakértő egy biológus (generatív modell). Ő mindent megtanul arról, hogy mitől macska egy macska: ismeri a csontozatát, a génjeit, a jellegzetes hangját, a bajszának funkcióját. A tudása olyan mély, hogy ha kell, képes lenne lerajzolni vagy akár szobrot készíteni egy macskáról pusztán a tudására hagyatkozva. A társadalomkutatásban ilyen a generatív modell (LDA): azt próbálja megérteni és leírni, hogy egy adott kategória (például egy téma) milyen összetevőkből, milyen belső recept alapján jön létre.
b) A második szakértő egy biztonsági őr (diszkriminatív modell). Neki egyetlen feladata van: megakadályozni, hogy kutyák jussanak be egy macskakiállításra. Nem kell tudnia, mi az a mitokondriális DNS, csak a legjellemzőbb különbségeket kell megtanulnia, amelyek elválasztják a macskát a kutyától (például a nyávogás és az ugatás közti különbség). Neki nem kell tudnia macskát rajzolni, csak a határvonalat kell ismernie a két kategória között. Ilyen a diszkriminatív modell (mint egy szentimentelemző): nem a szöveg keletkezésével foglalkozik, hanem kizárólag a kategóriák közötti különbségek megtanulására összpontosít, hogy a lehető legpontosabban el tudja dönteni, egy szöveg ebbe vagy abba a csoportba tartozik-e.

De miért fontos ez a megkülönböztetés a kutató számára? Azért, mert a modellválasztás közvetlenül a kutatási céltól függ. Ha a cél a feltárás és a megértés – például szeretnénk feltárni a parlamenti viták rejtett tematikus struktúráit, vagy megérteni, hogy az online hozzászólók milyen szavakkal írnak le egy politikai eseményt –, akkor a generatív modell (a biológus) a megfelelő választás, mert az betekintést nyújt a témák belső szerkezetébe, és segít értelmezni a jelenséget. Ha azonban a cél egy konkrét, előre meghatározott osztályozási feladat hatékony elvégzése – például több ezer újságcikket akarunk automatikusan „sport” vagy „politika” kategóriába sorolni, vagy egy termékről írt

véleményről akarjuk eldönteni, hogy „pozitív" vagy „negatív" –, akkor a diszkriminatív modell (a biztonsági őr) lesz a hatékonyabb, mert ezzel érhetünk el magasabb pontosságot. A döntés tehát a kutatói célon múlik, azon, hogy egy gazdag, értelmezhető, de talán kevésbé pontos leíró modellt szeretnénk, vagy egy rendkívül pontos, de nehezebben értelmezhető prediktív modellt? Ez a választás alapvetően meghatározza, hogy milyen típusú kérdésekre kaphatunk választ a szöveges adatainkból.

Az első és legelterjedtebb koncepcionális egyszerűsítés, amely lehetővé teszi a szövegek gépi feldolgozását, a szózsákmodell. E modell lényege, hogy egy dokumentumot (legyen az interjú, újságcikk vagy online hozzászólás) úgy kezelünk, mint egy zsákot, amelybe az összes szavát beletettük, figyelmen kívül hagyva a nyelvtani szerkezetet, a szórendet és a köztük lévő kapcsolatokat. Az egyetlen dolog, ami ebben a modellben számít, az, hogy mely szavak és milyen gyakorisággal fordulnak elő a dokumentumban. Ez a modell tehát egy fontos kompromisszumot köt azáltal, hogy a nyelvi információk jelentős részét, különösen a szórendet, feláldozza egy egyszerűbb, de matematikailag könnyen kezelhető, úgynevezett parszimónikus reprezentációért cserébe. Ezek a nyers gyakorisági adatok képezik a legegyszerűbb, valószínűségi alapú nyelvi modellek alapját. Például a „A gazdasági növekedés a célunk, nem a szociális biztonság." mondat a BoW-modellben a következő szógyakoriságok halmazává egyszerűsödik: {a: 2, gazdasági: 1, növekedés: 1, célunk: 1, nem: 1, szociális: 1, biztonság: 1} (ebben a példában a számok az adott szó előfordulásának gyakoriságát, vagyis a darabszámát jelentik az eredeti mondatban). A modell ereje az egyszerűségében és hatékonyságában rejlik, hiszen az összetett nyelvi struktúrákat könnyen kezelhető, számszerűsíthető egységekre bontja. Ugyanakkor a korlátai is ebből fakadnak. A modell teljesen figyelmen kívül hagyja a környezetet. A modell mögött az a nagyfokú egyszerűsítés, az úgynevezett „szózsák-feltételezés" húzódik meg, hogy egy dokumentum témájának vagy hangulatának megragadásához a szavak jelenléte és gyakorisága elegendő, a köztük lévő nyelvtani és sorrendi kapcsolatok pedig lényegtelenek. Az „A kutya kergeti a macskát" és az „A macska kergeti a kutyát" mondatok a BoW-modellben tökéletesen megegyeznek, holott a jelentésük ellentétes. Ez a példa rávilágít, hogy a pusztán szógyakoriságokra épülő BoW-alapú módszerek elsősorban a témák azonosítására, nem pedig a finomabb nyelvi-logikai kapcsolatok feltárására alkalmasak.

Eddig egyetlen dokumentum reprezentációját vizsgáltuk. A társadalomkutatásban azonban ritkán dolgozunk egyetlen szöveggel, hanem általában egy egész szöveggyűjteményt, vagyis korpuszt elemzünk. Hogyan tudjuk a sok különálló szózsákot egyetlen, elemzésre alkalmas struktúrába rendezni? Erre szolgál az úgynevezett dokumentum-term mátrix (Document-Term Matrix,

DTM). Érdemes megjegyezni, hogy a szakirodalomban egyre gyakrabban találkozhatunk az ennél általánosabb dokumentum-jellemző mátrix (Document-Feature Matrix, DFM) kifejezéssel is. Míg a term (kifejezés) klasszikusan egyetlen szóra utal, a feature (jellemző) egy tágabb fogalom, amely lehet egy szó („kormány"), egy többszavas kifejezés (bigram: „magyar kormány", trigram: „magyar kormány döntése") vagy akár egy komplexebb nyelvi mintázat is. Bár ebben a bevezető fejezetben a könnyebb érthetőség kedvéért a DTM elnevezést használjuk, fontos tudni, hogy a mögöttes logikai struktúra mindkét esetben azonos.

Képzeljünk el tehát egy egyszerű táblázatot, ahol a mátrix sorai a dokumentumokat (például interjúkat), oszlopai pedig a teljes korpuszban előforduló összes egyedi szót (a „term"-eket vagy „feature"-öket) képviselik. A táblázat celláiban pedig az adott szó előfordulásának gyakorisága szerepel az adott dokumentumban.

Dokumentum	a	célunk	gazdasági	növekedés	szociális	biztonság
Dokum 1	2	1	1	1	0	0
Dokum 2	2	1	0	0	1	1

17. táblázat: Egy egyszerűsített dokumentum-term mátrix (DTM) példája

A nyers szógyakoriságok használata azonban torzító lehet. A nagyon gyakori szavak (még a töltelékszavak, az úgynevezett stopszavak kiszűrése után is) aránytalanul nagy súlyt kapnak, függetlenül attól, hogy mennyire informatívak a dokumentum megkülönböztetése szempontjából. A TF-IDF (Term Frequency-Inverse Document Frequency) egy súlyozási eljárás, amely ezt a problémát orvosolja azzal, hogy megbecsüli egy szó fontosságát egy adott dokumentumban a teljes korpuszt figyelembe véve. A módszer logikája kettős: azoknak a szavaknak adja a legnagyobb súlyt, amelyek egy adott dokumentumban gyakoriak (ez a Term Frequency, TF), de a teljes korpuszban összességében ritkák (ez az Inverse Document Frequency, IDF). Azok a szavak, amelyek szinte minden dokumentumban előfordulnak (például „kutatás", „adat" egy módszertani könyvben), kevésbé segítenek a dokumentumok megkülönböztetésében, így alacsony súlyt kapnak. Ezzel szemben a csak néhány dokumentumban megjelenő, specifikus szavak (például „etnográfia", „diskurzuselemzés") magas súlyt kapnak, mert ezek hordozzák a legnagyobb megkülönböztető erőt. A TF-IDF tehát segít a dokumentumokat leginkább jellemző, valóban informatív kulcsszavak azonosításában.

Miután megvizsgáltuk, hogyan kezelhető egy teljes dokumentumgyűjtemény, térjünk vissza a szózsákmodell egy másik alapvető korlátjához, a környezet és

a szórend teljes figyelmen kívül hagyásához. A kontextusvakságnak az enyhítésére a legegyszerűbb kísérletet az n-gram modellek jelentik. Ahelyett, hogy csak önálló szavakat (unigramokat) számolnánk, ezek a modellek rövid, n szóból álló szekvenciákat – szópárokat (bigram), szóhármasokat (trigram) stb. – is figyelembe vesznek. A bigram modell például a „gazdasági növekedés" vagy a „szociális biztonság" kifejezéseket egyetlen egységként képes kezelni.

A gyakorlatban azonban ritkán cseréljük le a teljes szókincset n-gramokra. A leghatékonyabb megközelítés egy hibrid reprezentáció, ahol a legtöbb szó önállóan marad, de a kutató bizonyos, jelentéssel bíró szókapcsolatokat egyetlen egységként (tokenként) kezel. De hogyan döntjük el, hogy mely szókapcsolatokat érdemes összevonni?

a) Statisztikai megközelítés: A leggyakoribb módszer az, hogy statisztikai mutatók (például kölcsönös információ, pontonkénti kölcsönös információ) segítségével azonosítjuk azokat a szókapcsolatokat, amelyek szignifikánsan gyakrabban fordulnak elő együtt, mint az a szavak egyéni gyakorisága alapján várható lenne. Az olyan kifejezések, mint a „Fehér Ház" vagy a „Központi Bank" magas pontszámot kapnának, míg a véletlenszerűen egymás mellé kerülő szavak (például „zöld és") alacsonyat. Az algoritmusok ezek alapján automatikusan képesek a releváns n-gramokat azonosítani.
b) Kutatói, elméletvezérelt döntés: A másik megközelítés, amikor a kutató a saját szakterületi tudása alapján előre meghatároz egy listát azokról a kulcsfontosságú kifejezésekről, amelyeket mindenképpen egyben szeretne tartani (például egy politikatudományi elemzésben a „parlamenti választás" vagy a „civil szervezet").

Az n-gramok használata tehát egy újabb fontos módszertani döntési pont. Bár ez még mindig messze van a valódi nyelvi megértéstől, már a szópárok gyakoriságának vizsgálata is sokkal finomabb elemzést tesz lehetővé, mint a szavak puszta megszámlálása, hiszen olyan állandósult szókapcsolatokat és kollokációkat is azonosíthat, amelyek önmagukban jelentést hordoznak. Fontos azonban megkülönböztetni az n-gramok kétféle használatát. Amit eddig tárgyaltunk, az az n-gramok használata jellemzőként (feature) a dokumentumok reprezentálására, a szórend figyelmen kívül hagyásával. Létezik azonban az n-gram nyelvi modellek világa is, amelyeket a következő szó valószínűségének a megjóslására használnak. Ezek a klasszikus nyelvi modellek az úgynevezett Markov-feltételezésen alapulnak, amely szerint a következő szó valószínűsége csak egy szűk, fix méretű

előzménytől (az előző n–1 szótól) függ. Bár jelenlegi célunk nem a szöveggenerálás, hanem a meglévő szövegek elemzése, ez a mögöttes statisztikai logika mutatja, hogy a modern NLP hogyan kezdte el a kontextust modellezni.

Miután a dokumentumokat számszerűsítettük, minden dokumentumot egy-egy numerikus vektorként kezelhetünk egy többdimenziós térben. Ebben a térben a dokumentumok közötti távolság a tematikus hasonlóságukat tükrözi, ahol azok a dokumentumok, amelyek hasonló szavakat használnak, „közel" lesznek egymáshoz. A hasonlóság mérésére a legelterjedtebb módszer a koszinuszhasonlóság, amely a két dokumentumot reprezentáló vektorok által bezárt szöget méri. Ha a szög kicsi (a koszinusz értéke 1-hez közelít), a két dokumentum tematikailag nagyon hasonló. Ez az alapelv teszi lehetővé a dokumentumok automatizált csoportosítását (klaszterezését), a releváns irodalom keresését vagy akár egy vitában elhangzó hozzászólások politikai hovatartozás szerinti skálázását. Fontos megkülönböztetni a hasonlóság és a távolság mérését. Míg a koszinuszhasonlóság a vektorok iránya közti szöget méri, addig más metrikák, mint például az euklideszi távolság, a vektorok végpontjai közötti légvonalbeli távolságot számszerűsítik. Szövegelemzésnél gyakran a koszinuszhasonlóságot részesítik előnyben, mert az figyelmen kívül hagyja a dokumentumok hosszát (a vektorok nagyságát), és csak a szavak arányaira, vagyis a dokumentum tematikus fókuszára összpontosít. Egy rövid és egy hosszú dokumentum is lehet tematikailag nagyon hasonló (magas koszinuszhasonlóság), miközben a vektoraik végpontjai a térben távol esnek egymástól (nagy euklideszi távolság). Ezen elméleti alapok megértése biztosítja a kutató számára, hogy az NLP-eszközök által generált eredményeket ne csupán elfogadja, hanem kritikusan értelmezni és értékelni is tudja.

Bár a dokumentum-term mátrix egy erőteljes lépés a szövegek strukturálása felé, két komoly kihívással küzd. Az egyik a ritkaság, mert a mátrix celláinak túlnyomó többsége nullát fog tartalmazni, mivel a legtöbb szó nem fordul elő a legtöbb dokumentumban. Ez a mátrixot nehezen kezelhetővé és számításigényessé teszi. A másik, mélyebb probléma, hogy a DTM oszlopai (a szavak) egymástól függetlenek. A modell nem tudja azt, ami egy emberi olvasó számára nyilvánvaló: a „király" és az „uralkodó" szavak jelentésükben nagyon közel állnak egymáshoz, míg a „király" és a „káposzta" szavak jelentése távol esik egymástól. A DTM-ben ezek a szavak csupán különböző, egymással ekvivalens dimenziók. A következő fejezetekben bemutatott neurális hálózati modellek és szóbeágyzások ezekre a problémákra adnak választ azáltal, hogy a szavakat nem elszigetelt egységekként, hanem sűrű, többdimenziós jelentéstérben elhelyezkedő pontokként reprezentálják.

3. A kutatás folyamata

Mielőtt a bemutatott elemzési módszerek bármelyikét alkalmaznánk, a nyers szöveget fel kell bontanunk a számítógép számára értelmezhető alapegységekre, ún. tokenekre. Ezt a folyamatot nevezzük tokenizációnak. Bár egyszerűnek tűnhet, ez egy kritikus módszertani döntés. Döntést kell hoznunk például arról, hogy az írásjeleket külön tokenként kezeljük-e (például a „vége." az „vége" és „." tokenekre bomlik), vagy hogy az összetett szavakat (például „természetesnyelv-feldolgozás") egyben hagyjuk-e. A korszerű NLP-eszközök gyakran nem is szavakkal, hanem szubszavakkal dolgoznak, amelyek statisztikailag gyakori szórészek (például a „feldolgozhatatlan" szót felbonthatják a „feldolgoz", „hatat" és „lan" tokenekre). Ez lehetővé teszi a modell számára, hogy a ragozott vagy ritka szavakat is hatékonyan kezelje. Társadalomkutatóként nem kell feltétlenül mélyen ismernünk ezeknek az algoritmusoknak a működését, de tudatában kell lennünk, hogy az elemzésünk alapját képező „szavak" listáját ez a kezdeti lépés határozza meg. A legtöbb szoftver (például Orange, Python könyvtárak) ezt a lépést automatikusan elvégzi, de a paramétereik gyakran testre szabhatók.

A tokenizáció azonban csupán az első lépése egy tágabb folyamatnak, amelyet adat-előkészítésnek nevezünk. Célja, hogy a nyers szöveges adatokat megtisztítsuk a zajtól, és olyan formátumra hozzuk, amely az elemzési céljainknak a leginkább megfelel. Nincs egyetlen, minden helyzetben helyes recept, hanem a lépések kiválasztása mindig olyan módszertani döntés, amelyet a kutatási kérdés vezérel. A leggyakoribb előkészítési lépések a következők:

a) Kisbetűssé alakítás: A legegyszerűbb lépés, amely során a teljes szöveget egységesen kisbetűssé alakítjuk. Ennek célja, hogy a modell ne tegyen különbséget például a mondat elején álló „Kormány" és a mondat közben szereplő „kormány" szó között, így egységesítve a szókincset.
b) Írásjelek és számok eltávolítása: A legtöbb NLP-alapú, statisztikai mintázatkeresésre épülő elemzésnél az írásjelek és a számok felesleges zajt jelentenek, ezért gyakran eltávolítják őket. Azonban ez a lépés is körültekintést igényel. A kutatási cél itt is döntő, mert ha például a kutatási kérdés arra irányul, hogy hogyan használják a politikusok a felkiáltójeleket a szónoklataikban, az írásjelek eltávolítása tönkretenné az elemzést. Egy még gyakoribb hibaforrás az évszámok meggondolatlan törlése. Egy

történelmi vagy politikai szövegelemzésben az évszámok lényeges kontextuális információt hordoznak, és automatikus eltávolításukkal éppen a legfontosabb időbeli hivatkozásokat veszíthetjük el.

c) Stopszavak eltávolítása: A stopszavak a nyelv leggyakoribb, de tipikusan legkevesebb önálló jelentéssel bíró szavai (például a, az, és, hogy, is). Mivel ezek rendkívül gyakoriak, egy témaelemzésben elnyomnák a ritkább, de tartalmasabb kulcsszavakat. Eltávolításukkal a modell a valóban informatív szavakra tud fókuszálni. A klasszikus ellenpélda Mosteller és Wallace esete a Federalist Papersszel: az esszék szerzőségének azonosításához a kutatók pont a stopszavakat használták, mert ezek a szerző egyedi, tudattalan stílusát tükrözték, nem pedig a szöveg témáját.

d) Szótövesítés és lemmatizálás: E két eljárás célja, hogy a különböző ragozott szóalakokat visszavezessék egy közös alapalakra, csökkentve ezzel a szókincs méretét és segítve a fogalmak azonosítását. A szótövesítés egy gyors, de heurisztikus módszer, amely egyszerűen levágja a szavak végi toldalékokat (például a kutatások, kutatási, kutatja szavakból a kutat szótőt hozza létre). Előnye a sebessége, hátránya, hogy nyelvtanilag helytelen vagy értelmetlen töveket eredményezhet (például a generális szóból a generál tövet hozhatja létre, ami teljesen megváltoztatja a szó jelentését és szófaját). A lemmatizálás egy fejlettebb, szótáralapú módszer, amely a szó szótári alapalakját, a lemmát keresi meg (például a „kutatások" esetében a „kutatás", a „láttam" esetében a „lát" lenne a lemma). Ez nyelvtanilag sokkal pontosabb, de számításigényesebb, és egy jó minőségű nyelvi szótárat igényel. A jó minőségű szótárra azért van szükség, mert a lemmatizálónak fel kell ismernie a rendhagyó szóalakokat is, tehát a szótárnak tartalmaznia kell például azt az információt (magyar nyelv esetén), hogy a „vagyok" szó szótári alapalakja a „van" ige. Erre azért van szükség, mert a „vagyok" és a „van" szavak között nincs olyan egyszerű, szabályalapú, levezethető kapcsolat, amelyet egy számítógép magától kitalálhatna. Gondoljunk bele, hogyan működik egy gép. Csak azt tudja, amit lépésről lépésre megmondanak neki. A „vagyok" és a „van" betűsorozatok nem azonosak. A gép számára ezek két teljesen különböző dolgot jelentenek. A szótövesítés pedig egyszerű szabályokat követ: „vágd le a szó végéről a magánhangzó + mássalhangzó párokat". Ez a „vagyok"-ból valószínűleg a „vagy" tövet hozná létre.

A kutatási folyamat tehát nem egy lineáris recept, hanem döntések sorozata. Egy névelem-felismerő (NER) modellhez szinte nyers, változatlan szövegre van szükség, míg egy témaelemzéshez egy megtisztított adatbázisra. Az előkészítési

lépések megválasztása az első és egyik legfontosabb módszertani döntés, amelyet a kutatónak meg kell hoznia. Az adatok előkészítése után a kutatói eszköztárból ki kell választani azt a módszert, amely a leginkább megfelel a kutatási kérdésnek. Az alábbiakban három alapvető, de a társadalomkutatási gyakorlatban rendkívül hasznos és grafikus felületű szoftverekkel is elérhető módszert mutatunk be.

3.1. Névelem-felismerés (Named Entity Recognition – NER)

Míg az elméleti alapoknál bemutatott módszerek a szöveget szavak zsákjaként kezelik, a kutatót gyakran nemcsak a szavak általános tematikája, hanem a szövegben szereplő egyes szereplők is érdeklik. A névelem-felismerés a szekvenciacímkézés elnevezésű feladatkörbe tartozik, ahol a cél egy sorozat (például egy mondat szavai) minden eleméhez egy-egy címkét (például SZEMÉLY, SZERVEZET) rendelni. Ki mondta? Kiről van szó? Mely szervezeteket vagy helyszíneket említették? A névelem-felismerés erre a kutatási igényre ad választ. A módszer alapvető váltást jelent a szózsák-megközelítéshez képest, mert a célja nem a szavak megszámlálása, hanem a szövegben előforduló tulajdonnevek és más, sajátos entitások önműködő azonosítása és osztályokba sorolása. Ezzel a módszerrel a NER egy alapvető információkinyerési feladatot valósít meg, azaz a strukturálatlan szövegfolyamból diszkrét, strukturált adatokat (például személyek, szervezetek, helyszínek listáját) hoz létre. A végeredmény egy táblázat, amelyben a szövegben említett szereplők és kapcsolataik válnak elemezhetővé. Egy politikai vitákból álló korpusz elemzésekor például ahelyett, hogy csak a viták témáit ismernénk meg, egy NER-elemzés után azt is meg lehet mondani, hogy mely politikusok említik a leggyakrabban egymást, vagy mely szervezetek kerülnek legtöbbször szóba egy adott téma kapcsán.

A névelem-felismerésnek alapvetően két technikai megközelítése létezik. A korábbi, szabályalapú rendszerek szótárakra és nyelvtani mintázatokra (például a nagybetűvel kezdődő szó, amelyet a „kormánya” szó követ, valószínűleg egy ország) támaszkodtak. Ezek a rendszerek merevek, és nehezen tudnak kezelni új, a szótárban nem szereplő neveket. A korszerű, gépi tanuláson alapuló módszerek ezzel szemben ember által előre felcímkézett példaszövegeken tanulják meg az entitások kontextuális mintázatait. Megtanulják például, hogy egy név után gyakran áll egy ige, vagy hogy egy szervezet nevében gyakran szerepel a „Kft.” vagy az „Egyetem” szó. A tanulás folyamán a modell a szavak numerikus reprezentációit használja, vagyis azok a szavak, amelyek hasonló környezetben fordulnak elő (például a „Kft.” vagy „Zrt.” szavak előtt),

hasonló vektort kapnak. Így a modell – ha az egy ismert mintázatba illeszkedik – egy eddig sosem látott cégnevet is nagy valószínűséggel SZERVEZET-ként ismer fel. Az olyan elterjedt eszközök, mint a Stanford Named Entity Recognizer (NER) vagy a spaCy könyvtár, ilyen modelleket használnak. A módszer alkalmazása azonban nem triviális. A legnagyobb kihívást a kétértelműség jelenti: a „Washington" utalhat személyre, városra vagy államra is. Egy jó NER-modell a szövegkörnyezet alapján képes eldönteni a helyes kategóriát, de fontos a nyelvi megfelelőség, például egy angol nyelvre tanított modell magyar szövegen használhatatlan lesz a nyelvtani különbségek és a ragozás miatt.

A névelem-felismerés tehát a szövegekből kinyert szereplők hálózati kapcsolatainak, láthatóságának és diskurzusbeli szerepének további, mélyebb elemzését teszi lehetővé. Az alábbiakban egy kitalált példán keresztül mutatjuk be, hogyan használható a NER egy politikai kommunikációs kutatásban.

1. lépés: A kutatási kérdés és az adatok

A kutatási kérdés, hogy a 2024-es magyarországi önkormányzati választási kampányban mely politikai szereplők (pártok és politikusok) kapták a legnagyobb médiafigyelmet a vezető online hírportálokon, és milyen szervezetekkel említették őket a leggyakrabban együtt. A kutatáshoz összegyűjtjük a kampányidőszak (például 2024. április–június) alatt megjelent, választással kapcsolatos cikkeket négy vezető magyar hírportálról (például Index, Telex, 24.hu, Origo). Az összegyűjtött cikkek szövegét egyetlen táblázatba (például CSV- vagy Excel-fájlba) rendezzük, ahol minden sor egy-egy cikket tartalmaz.

2. lépés: A szoftvereszköz kiválasztása

Az elemzéshez az Orange Data Mining nevű ingyenes, nyílt forráskódú szoftvert választjuk. Előnye, hogy grafikus felületen, programozási tudás nélkül teszi lehetővé bonyolult adatelemzési folyamatok felépítését.

3. lépés: Az adatok betöltése és előkészítése

Az Orange munkafelületén elhelyezünk egy „Corpus" modult, és betöltjük vele a cikkeket tartalmazó táblázatunkat. A modult összekötjük a „Preprocess Text" modullal. A NER-elemzéshez nincs szükség a szöveg mélyebb tisztítására. A program a tokenizációt automatikusan elvégzi. Fontos, hogy ennél a lépésnél ne alakítsuk a szöveget kisbetűssé, ne végezzünk szótövesítést és ne távolítsuk el a stopszavakat. A korszerű NER-modellek gyakran támaszkodnak

a nagybetűs írásmódra mint fontos jelzésre a tulajdonnevek azonosításában, ezért az eredeti formátum megőrzése lényeges.

4. lépés: A névelem-felismerés futtatása

A „Preprocess Text" modul (gyakori kifejezéssel: widget) kimenetét összeköt-jük a „Named Entity Recognition" modullal, amely a művelet központi eleme. A modul beállításaiban ki kell választani a nyelvi modellt. Mivel magyar nyelvű szövegeket elemzünk, itt egy magyar nyelvre kialakított modellt kell kiválasz-tanunk, egy angol modell nem ismerné fel a magyar neveket és szervezeteket. A futtatás után az adathalmazunk kiegészül új megjegyzésekkel vagy metaada-tokkal, mert a program minden felismert entitást (például SZEMÉLY, SZERVE-ZET, HELYSZÍN) megjelöl a szövegben, és ezek a címkék elérhetővé válnak a további elemzés számára.

5. lépés: Az eredmények feltárása és elemzése

A NER által generált címkézett adatok többféleképpen is elemezhetők. Az első lehetőség a gyakorisági elemzés: Ki szerepelt a legtöbbet? Kössük a NER kime-netét egy „Word Cloud" alkalmazásra. A szófelhő beállításaiban adjuk meg, hogy csak a „PERSON" (személy) vagy „ORG" (szervezet) címkével ellátott toke-neket jelenítse meg. Az eredmény egy kép lesz arról, hogy mely politikusok és pártok nevei fordultak elő leggyakrabban. A pontos darabszámok megjelenítésére a „Concordance" alkalmazás használható.

A második lehetőség a hálózatelemzés: Kik és mik szerepeltek együtt? A kutatási kérdés második felének megválaszolásához az együttes előfordulásokat vizsgáljuk. A NER kimenetéből egy hálózatot hozunk létre, ahol a csomópontok a politikusok és szervezetek, az élek pedig az együttes előfordulásukat jelzik. Azonban meg kell határozni, hogy mit jelent az együttes előfordulás. A legegy-szerűbb, ha azt tekintjük kapcsolatnak, ha két entitás ugyanabban a cikkben szerepel. Az így kapott hálózatot egy „Network Explorer" alkalmazással jele-nítjük meg. Itt láthatóvá válnak a keresett összefüggések, például a politikusok klaszteresedése a saját pártjuk köré vagy bizonyos civil szervezetek és politikai oldalak közötti gyakori együttes említések.

6. lépés: Az eredmények értelmezése és a korlátok

Az elemzést követően válaszokat lehet adni a kutatási kérdésre (például: „Az X párt és annak vezetője kapta a legtöbb említést a vizsgált időszakban, és legygyakrabban az Y intézménnyel hozták összefüggésbe a híradások."), de a NER-modellek nem tévedhetetlenek. Előfordulhatnak hibás címkézések (például egy köznevet tulajdonnévként ismer fel) és kihagyások (nem ismer fel egy valós nevet). Ezért érdemes az eredmények egy kisebb, véletlenszerű mintáját személyesen, kézzel is ellenőrizni, hogy felmérjük a modell pontosságát. Az eredményeket mindig a modell lehetséges hibáinak tudatában kell értelmezni.

3.2. Szentimentelemzés (Sentiment Analysis)

Míg a szógyakoriság-elemzés hatékonyan tárja fel, hogy miről szólnak a szövegek, addig arról keveset árul el, hogy a szerzők hogyan viszonyulnak az adott témához. A szentimentelemzés vagy más néven véleménybányászat ezt a hiányosságot hivatott pótolni. A módszer a vélemények, attitűdök és érzelmek azonosítására és osztályozására összpontosít, és nagy mennyiségű szöveges adatban képes számszerűsíteni a vélekedések jellegét és erősségét.

A módszernek alapvetően két fő megközelítése terjedt el. A legkönnyebben érthető a szótáralapú megközelítés. Ennek logikája, hogy a szoftver egy előre összeállított szótárat használ, amelyben a szótár alkotói a szavakat pozitív vagy negatív címkével és gyakran egy polaritási pontszámmal (például „csodálatos": +0,9, „szörnyű": -0,8) látták el. Ez a megközelítés lényegében a súlyozott szózsákmodell egyedi esete, ahol a dokumentumot a benne lévő szavak összességeként kezeljük, de itt nem minden szó számít, csak azok, amelyek szerepelnek a szentimentszótárban, és ezeket is a hozzájuk rendelt polaritási pontszámmal vesszük figyelembe. Az elemzés során a program megvizsgálja a szöveget, azonosítja a szótárban szereplő szavakat, majd összesíti azok pontszámait, így állítva elő a teljes dokumentumra vonatkozó összesített szentimentértéket. A kifinomultabb eszközök, mint például a pszichológiai kutatásokban régóta használt LIWC (Linguistic Inquiry and Word Count), ennél azonban tovább mennek, és nemcsak egyszerű pozitív-negatív dimenziókat mérnek, hanem azonosítják többek között a szorongásra, a magabiztosságra vagy a jövőorientáltságra utaló nyelvi mintázatokat is. Sőt, a szakterület-specifikus szótárak gyakran még ennél is tovább mennek, és egy pénzügyi elemzésekhez készült szótár például külön kategóriákat használ a bizonytalanságot, a korlátozást

vagy a jogi kockázatot jelző kifejezésekre, így téve lehetővé a szövegek sokkal árnyaltabb, többdimenziós elemzését.

A szótáralapú logikának egy újabb változata a gépi tanuláson alapuló megközelítés. Itt a rendszer nem előre meghatározott szabályokból dolgozik, hanem ember által felcímkézett példákból tanulja meg a különböző vélemények nyelvi mintázatait. A kutató először manuálisan osztályoz egy több száz vagy ezer elemű mintát (például online hozzászólásokat címkéz pozitívnak, negatívnak), majd ezt a tanító adatbázist adja át a modellnek. (A gyakorlatban ez sokszor egy egyszerű táblázatot jelent, amelynek egyik oszlopában a szövegek – például hozzászólások –, a másikban pedig a kutató által hozzárendelt címkék – például „pozitív", „negatív" – szerepelnek.) Az algoritmus pedig feltárja azokat a szóhasználati és strukturális mintázatokat, amelyek a leginkább jellemzőek az egyes kategóriákra, és e tudás birtokában képes lesz új, számára ismeretlen szövegeket is nagy pontossággal besorolni.

Mindkét megközelítésnek azonban szembe kell néznie a természetes nyelv összetettségéből fakadó kihívásokkal. A legnehezebben kezelhető problémák az irónia és a szarkazmus, ahol a szavak felszíni jelentése éppen ellentétes a hordozott véleménnyel (például „Ez a reform fantasztikus, gratulálok!"), valamint a nyelvspecifikusság (esetünkben a magyar nyelv ragozó természete). Az alábbiakban egy kitalált példán keresztül mutatjuk be, hogyan használható a szentimentelemzés a fogyasztói visszajelzések kutatásában.

1. lépés: A kutatási kérdés és az adatok

Kutatási kérdésünk, hogy milyen volt a fogyasztók reakciója egy újonnan piacra dobott okostelefonra, és melyek a leggyakrabban említett pozitív és negatív tulajdonságai a terméknek a felhasználói értékelések alapján. A kutatáshoz egy népszerű e-kereskedelmi oldalról összegyűjtjük a termékhez tartozó összes szöveges felhasználói értékelést a megjelenést követő első hónapban. Az adatokat egy táblázatba (például CSV-fájlba) rendezzük, amely legalább egy oszlopot tartalmaz, az értékelés szövegét.

2. lépés: Az eszköztár kiválasztása

Az elemzéshez az R programozási környezetet és annak szövegelemző csomagjait (például a tidytext csomagot) választjuk. Az R előnye a rugalmassága és a tudományos közösségben való elterjedtsége, amely hozzáférést biztosít a legújabb módszerekhez és szentimentszótárakhoz. Bár a használata több előkészületet igényel, mint egy grafikus felületű szoftveré, a folyamat logikája jól követhető.

3. lépés: Az adatok betöltése és előkészítése

Az R környezetben beolvassuk az értékeléseket tartalmazó CSV-fájlt, majd elvégezzük az adat-előkészítés lépéseit. A tidytext csomag logikájával ez egy lépéssorozat, ahol az eredeti adattáblát alakítjuk át:

```
# Eredeti tábla: [értékelés_szövege]
# 1. lépés: A szöveget szavakra bontjuk (tokenizáció).
# Eredmény: Egy új tábla, ahol minden sor egy-egy szót tartalmaz.
# 2. lépés: Eltávolítjuk a stopszavakat.
```

A szótáralapú szentimentelemzés első lépése a szöveg tokenizálása. Gyakori, bár nem minden esetben elengedhetetlen további lépés a stopszavak eltávolítása. Bár ezek a szavak (például „és", „hogy") jellemzően érzelmileg semlegesek, és a szentimentszótárakban nem szerepelnek, eltávolításuk befolyásolhatja az eredményt, ha a szentimentet a dokumentum teljes hosszához viszonyítjuk. E lépés célja, hogy a modell a valóban érzelmi töltetet hordozó szavakra fókuszálhasson, de fontos tudni, hogy ez a normalizációs mutatók értelmezését is megváltoztatja, így alkalmazása tudatos kutatói döntést igényel. Ennél az elemzésnél a szótövesítést vagy lemmatizálást gyakran kihagyjuk, mert a szentimentszótárak konkrét szóalakokat tartalmaznak.

4. lépés: A szentimentelemzés futtatása

A szótáralapú elemzés egy egyszerű táblázat-összekapcsolási művelet. A hozzárendelés (vagy címkézés) logikája a következő:

```
# 1. Töltsünk be egy szentimentszótárat (például Bing lexikont vagy magyar
nyelvű szöveg esetén egy magyar nyelvű szentimentszótárat).
# Ez egy kéthasábos tábla: [szó | szentiment (pozitív/negatív)]
# 2. Kössük össze a szavakra bontott szöveges adatainkat a szótárral.
# A program minden szónál megnézi, szerepel-e a szótárban, és ha igen, hozzárendeli a címkét.
# Eredmény: Egy tábla, ahol a szavak mellett ott a szentimentcímke is.
```

Ezt követően minden egyes értékelésre összesítjük a talált pozitív és negatív szavak számát, és általában a pozitív szavak számából kivonva a negatívak számát egy összesített szentimentpontszámot képezünk. Az eredeti adattáblánk így kiegészül egy numerikus szentimentpontszámmal.

5. lépés: Az eredmények feltárása és elemzése

A szentimentpontszámok alapján az értékeléseket egy egyszerű szabály mentén három kategóriába sorolhatjuk. Ha az összesített pontszám nagyobb, mint nulla, az értékelést *pozitív*nak címkézzük. Ha kisebb, mint nulla, *negatív*nak. Ha a pontszám pontosan nulla (mert vagy nem tartalmazott szentimentszótárban szereplő szót, vagy a pozitív és negatív szavak kioltották egymást), akkor *semleges*nek minősítjük. Ezt követően egy oszlopdiagramon ábrázolhatjuk a vélemények általános megoszlását. A kutatási kérdés második felének megválaszolásához a leggyakoribb pozitív és negatív szavakat azonosítjuk. Ennek logikája:

1. Számoljuk meg a szavak gyakoriságát külön a pozitív, és külön a negatív kategóriában.
2. Hozzunk létre egy ábrát, amely a két kategória leggyakoribb szavait mutatja be.

Ez megmutathatja, hogy a pozitív értékelésekben például a kamera, kijelző, akkuidő szavak vannak túlsúlyban, míg a negatívakban a drága, lassú, melegszik kifejezések, így feltárva a fogyasztói vélemények mögött húzódó konkrét indokokat.

6. lépés: Az eredmények értelmezése és a korlátok

Az elemzés eredményeként konkrét következtetéseket vonhatunk le, például: „A termék fogadtatása összességében pozitív volt, a felhasználók leginkább a kiváló kamerát és a hosszú akkuidőt dicsérték. A kritikák elsősorban a magas árra és a készülék melegedésére összpontosítottak." Azonban a szótáralapú szentimentelemzés egy egyszerűsítő modell, amelynek komoly korlátai vannak. Az eredményeket mindig ezeknek a korlátoknak a tudatában kell értelmezni. A leggyakoribb kihívások a következők:

a) Doménspecifikusság: A módszer pontossága nagymértékben függ attól, hogy a használt szentimentszótár mennyire fedi le a vizsgált szövegek (a domén) egyedi nyelvezetét. Egy általános célú szótár, amelyet például filmkritikákra fejlesztettek, nem fogja tudni, hogy a pénzügyi hírekben a „volatilitás" vagy a „likviditás" szó negatív vagy pozitív értelemben jelenik-e meg. A mi példánkban lehet, hogy a „bloatware" (felesleges előtelepített szoftver) egyértelműen negatív kifejezés a technológiai közösségben, de ha ez a szó nem szerepel az általános szótárunkban, a modell

figyelmen kívül fogja hagyni. A megbízható elemzéshez gyakran az adott szakterületre finomhangolt vagy kifejezetten arra készített szótárakra van szükség.

b) A dokumentumhossz torzítása: Az egyszerű pontszámösszesítés (pozitív szavak száma - negatív szavak száma) nem veszi figyelembe a dokumentum hosszát. Egy hosszú, de alapvetően semleges szöveg, amelyben véletlenül eggyel több pozitív szó van, mint negatív, ugyanúgy +1 pontot kap, mint egy rövid, de egyértelműen pozitív vélemény. Ennek a torzításnak a kiküszöbölésére a kutatók gyakran normalizálják a pontszámot, például elosztják azt a dokumentum teljes szószámával, így egy arányszámot kapva, amely jobban tükrözi a szentiment intenzitását.

c) A környezet figyelmen kívül hagyása (tagadás, irónia, szarkazmus): A szótáralapú megközelítés a szavakat elszigetelten vizsgálja, ezért rendkívül sebezhető a kontextusfüggő jelentésmódosulásokkal szemben. A tagadás kezelése alapvető kihívás. A legegyszerűbb modell az „Ez a telefon egyáltalán nem lassú." mondatot a „lassú" szó miatt negatívnak fogja értékelni, holott a jelentése éppen az ellenkezője. Az iróniára és szarkazmusra való teljes érzéketlenség talán a legnehezebben megoldható probléma. A „Persze, remek ez a kétórás akkuidő..." mondatot egy egyszerű modell a „remek" szó miatt pozitívnak címkézheti. Bár léteznek kifinomultabb, úgynevezett „valence shifter" (értékmódosító) algoritmusok, amelyek megpróbálják kezelni a tagadószavakat (például „nem", „soha") vagy az erősítő kifejezéseket (például „nagyon", „rendkívül"), ezeknek a pontos alkalmazása már mélyebb programozói ismereteket igényel, és még ezek sem tévedhetetlenek.

A szentimentelemzés tehát egy hatékony első lépés a közhangulat feltérképezésére, de az eredményeit mindig kritikusan kell kezelni, és ideális esetben a félreértékeltnek tűnő esetek személyes, kézi elemzésével kell kiegészíteni.

3.3. Témaelemzés (Topic Modeling)

Míg a korábban bemutatott módszerek a szövegek felszíni, nyílt jellemzőire összpontosítanak – például a szavak gyakorisága –, a kutatót gyakran a mélyben meghúzódó, rejtett tematikus struktúrák érdeklik. Amikor egy több ezer dokumentumból álló korpuszt elemzünk, a célunk gyakran nem kevesebb, mint feltárni azokat a rejtett diskurzusokat vagy témaklasztereket, amelyek az adatainkat szervezik. Erre a kihívásra kínál megoldást a témaelemzés, amely a korszerű

NLP-eszköztár egyik legfejlettebb, mégis egyre könnyebben elérhető módszere. A témaelemzés elsősorban egy felfedező módszer. Nem arra való, hogy előre megfogalmazott hipotéziseket teszteljünk (például „a szövegek X témáról szólnake?"), hanem arra, hogy egy nagy, ismeretlen korpusz rejtett tematikus szerkezetét feltérképezzük, és ezáltal új kutatási kérdéseket és hipotéziseket generáljunk.

A témaelemzés legelterjedtebb algoritmusa a Latent Dirichlet Allocation (LDA). A módszer logikája leginkább egy statisztikai nyomozó munkájához hasonlítható. Az LDA egy generatív modell, ami azt jelenti, hogy feltételez egy mögöttes folyamatot, egy történetet arról, hogyan jöttek létre a dokumentumok. Eltérően a hagyományos klaszterezéstől, ami minden egyes dokumentumot egyetlen, diszkrét kategóriába sorol, az LDA egy kevert tagsági modell. Azt feltételezi, hogy minden dokumentum több téma keverékeként jön létre, és minden téma szavak egyedi eloszlásával jellemezhető. Fontos kiemelni, hogy az LDA – a szózsákmodellhez hasonlóan – teljesen figyelmen kívül hagyja a szavak sorrendjét. A modell számára egy dokumentum csupán a benne előforduló szavak gyakoriságainak halmaza. A témákat kizárólag a szavak együttes előfordulásai alapján azonosítja, a mondatok szerkezetét nem elemzi. Az LDA végignézi a teljes korpuszt, és azt vizsgálja, hogy mely szavak hajlamosak gyakran együtt előfordulni ugyanazokban a dokumentumokban. Ezen együttes előfordulási mintázatok alapján statisztikailag valószínűsíti, hogy mely szavak tartoznak közös témakosarakba. Az eredmény nem egyetlen lista, hanem két, egymást kiegészítő információ: egyrészt megkapjuk a témákat, amelyeket a hozzájuk leginkább kapcsolódó kulcsszavak listája határoz meg (például Téma 1: oktatás, iskola, diák, tanár; Téma 2: gazdaság, piac, adó, költségvetés), másrészt pedig minden egyes dokumentumról megtudjuk, hogy milyen arányban épül fel ezekből a témákból (például Dokumentum A: 60% Téma 1, 30% Téma 2, 10% egyéb).

A modell azonban a témákat nem a szó emberi értelmében fedezi fel: nem érti a szöveget, csupán a szavak valószínűségi eloszlásait elemzi. A kutató szerepe ebben a folyamatban megkerülhetetlen. A legfontosabb kutatói döntés a témák számának (k) meghatározása, amelyet a kutatónak kell megadnia a modell számára. Túl kevés téma esetén túlságosan általános, semmitmondó kategóriákat kapunk, míg túl sok téma esetén az eredmények széttöredeznek, és a klaszterek értelmezhetetlenné válnak.

A cél az, hogy megtaláljuk azt az optimális k értéket, amely a legjobb egyensúlyt teremti meg a modell részletessége és értelmezhetősége között. A gyakorlatban ez azt jelenti, hogy több modellt is lefuttatunk különböző k értékekkel (például k = 5, 10, 15, 20), és azt a pontot keressük, amely után a témák számának további növelése már nem eredményez szignifikáns javulást a modell minőségében. Hasonlóan a klaszteranalízis könyökmódszeréhez, itt is azt a töréspontot

figyeljük, ahol a modell további bonyolítása már csak minimális új információt ad hozzá a témák értelmezhetőségéhez vagy szétválaszthatóságához.

Bár léteznek ezt a folyamatot segítő statisztikai mutatók (mint a *perplexity* vagy a témakoherencia [coherence score]), amelyek a modell illeszkedését és a témák minőségét próbálják számszerűsíteni, a végső döntés mindig elméleti és interpretatív alapokon nyugszik. A kutatónak kell mérlegelnie, hogy melyik témaszám eredményez a kutatási kérdés szempontjából leginkább koherens, értelmes és hasznos kategóriákat.

Ezt követően szintén a kutató feladata, hogy a kapott szólistákat értelmezze, és emberi nyelven is érthető címkékkel lássa el (mint például a főkomponenselemzést követően a kapott látens dimenzió elnevezése).

Az eljárás ma már nem igényel mély programozási ismereteket. Az olyan szoftverek, mint az Orange Data Mining vagy a MALLET programhoz készült grafikus felületek lehetővé teszik a kutatók számára, hogy a teljes folyamatot – az adatok betöltésétől a témaszám beállításán át az eredmények képi megjelenítéséig – egy interaktív, könnyen kezelhető környezetben végezzék el. Egy tudományterület több ezer publikációjának absztraktjait elemezve például egy ilyen modellel feltárhatók a terület fő kutatási irányai, azok időbeli népszerűségének változása vagy akár az egyes kutatóintézetek tematikus profilja közötti különbségek. A témaelemzés eredményeit azonban mindig kritikusan kell kezelni. A minőségbiztosítás elengedhetetlen lépése az eredmények validálása. A legmegbízhatóbb módszer, ha a modell által azonosított témákat összevetjük egy kisebb, véletlenszerűen kiválasztott, de a hagyományos kódolás szabályai szerint elemzett mintával. Ha az emberi kódoló által azonosított témák és a gép által talált statisztikai klaszterek nagymértékben fedik egymást, az jelentősen megerősíti az eredményeink érvényességét. A témaelemzés tehát nem a hermeneutikai elemzés helyettesítője, hanem annak hatékony kiegészítője, egy makroszkóp, amely lehetővé teszi, hogy hatalmas szövegtömegekben is meglássuk a rejtett struktúrákat, amelyeket azután a hermeneutikai jellegű módszerek mélységével tárhatunk fel. Míg az LDA egy ma már szinte hagyományosnak tekinthető és robusztus módszer, érdemes megemlíteni a korszerűbb, transzformeralapú eljárásokat is. Ezek a BERT-típusú nyelvi modellekre építenek, amelyek a szavak kontextusfüggő jelentését is képesek kezelni. A témaelemzés területén egy ilyen népszerű eljárás a BERTopic, amely a BERT technológiáját használja fel a koherensebb témák azonosítására. Ezek a módszerek a szavak jelentésbeli hasonlóságát is figyelembe veszik, így gyakran jobban értelmezhető témákat eredményeznek, mivel nemcsak a szavak együttes előfordulására, hanem a szemantikai kontextusukra is támaszkodnak. Az alábbiakban egy kitalált neveléstudományi kutatáson keresztül mutatjuk be ezek használatát.

1. lépés: A kutatási kérdés és az adatok

Kutatási kérdésünk, hogy egy egyetemi kurzus online fórumán milyen rejtett témák és tanulási stratégiák merülnek fel a diákok hozzászólásaiban a szemeszter során. A kutatáshoz összegyűjtjük a kurzushoz tartozó online fórum összes diákok által írt bejegyzését. Az adatokat egy táblázatba (például CSV-fájlba) rendezzük, ahol minden sor egy-egy hozzászólást tartalmaz.

2. lépés: A szoftvereszköz kiválasztása

Ehhez a feladathoz egy erre specializált, kutatók által széles körben használt eszközt, a MALLET-et (MAchine Learning for LanguagE Toolkit) választjuk. Bár ez eredetileg egy parancssori eszköz, léteznek hozzá ingyenes, felhasználóbarát grafikus felületek, mint például a Topic Modeling Tool, amely a teljes folyamatot átláthatóvá és kattinthatóvá teszi. A példánkban erre a grafikus eszközre hivatkozunk.

3. lépés: Az adatok betöltése és előkészítése

A témaelemzéshez a dokumentumokat különálló fájlokként kell kezelnünk. A táblázatunk minden sorát (minden hozzászólást) külön txt-fájlként kell elmentenünk egyetlen mappába. A témaelemzés minősége nagyban függ az alapos előkészítéstől. A Topic Modeling Toolban lehetőségünk van egy magyar nyelvű stopszólista betöltésére, ami elengedhetetlen a zajszűréshez. A legjobb eredmény érdekében a szövegek lemmatizálása is erősen ajánlott, hogy a tanulás, tanulni, tanulok szavakat a modell egy fogalomként kezelje. Ezt a lépést gyakran a programba való betöltés előtt, egy különálló szkript segítségével kell elvégezni.

4. lépés: A témaelemzés futtatása

A Topic Modeling Tool felületén betöltjük a txt-fájlokat tartalmazó mappát. Itt kell meghozni a legfontosabb kutatói döntést: a témák számának (k) meghatározását. Mivel ez egy felfedező elemzés, érdemes több különböző k értékkel is kísérletezni (például 5, 10, 15 témával), hogy megnézzük, melyik beállítás adja a leginkább koherens és értelmezhető eredményt. A példában kezdjük 10 téma keresésével. Az algoritmus elindítása után a program a teljes korpusz alapján azonosítja a szavak együttes előfordulási mintázatait, és létrehozza a témamodellt.

5. lépés: Az eredmények feltárása és elemzése

Az algoritmus két fő kimenetet ad, amelyeket a szoftver interaktív formában jelenít meg, és amelyek a kutatói interpretáció alapját képezik:

a) Téma-szó listák: A program minden azonosított témához megjelenít egy listát a legjellemzőbb, súlyozott kulcsszavakból. A kutató feladata, hogy ezeket a szóklasztereket értelmezze, és címkével lássa el. Példaeredmény: Téma 3: vizsga, stressz, jegy, idő, készülni, nehéz -> Értelmezés/címkézés: „Vizsgastressz és időmenedzsment"; Téma 7: csoport, együtt, segítség, kérdés, megbeszélni -> Értelmezés/címkézés: „Kollaboratív tanulás és segítségkérés".
b) Dokumentum-téma megoszlások: Az LDA minden egyes hozzászóláshoz hozzárendel egy valószínűségi eloszlást, amely megmutatja, hogy az adott hozzászólás milyen arányban tartozik az egyes témákhoz. Ezt az adatot felhasználhatjuk a témák időbeli népszerűségének vizsgálatára, vagy azonosíthatjuk azokat a hozzászólásokat, amelyek a leginkább reprezentálnak egy-egy témát.

6. lépés: Az eredmények értelmezése és a korlátok

Az elemzés eredményeként nemcsak azonosítjuk a diákok diskurzusának fő témáit, hanem fel is tárjuk azok belső szerkezetét a kulcsszavak révén. Ez lehetővé teszi a kutató számára, hogy a felszín mögé nézzen, és adatalapúan térképezze fel a tanulási folyamat rejtett struktúráit. Ugyanakkor fontos tudatosítani a módszer korlátait:

a) Interpretációs felelősség: Az LDA csak statisztikai szóklasztereket ad, de a témák elnevezése és értelmezése teljes mértékben a kutató feladata és felelőssége.
b) A „k" kérdése: A témák számának megválasztása befolyásolja az eredményt. Egy rosszul megválasztott k érték túl általános vagy túlságosan széttöredezett, értelmezhetetlen témákat eredményezhet, ezért a validálás kiemelt jelentőségű.
c) A szórend hiánya: A modell nem veszi figyelembe a szavak sorrendjét. Az eredményeket ezért mindig a szövegek kontextuális olvasatával együtt érdemes értelmezni. A témaelemzés a nagy mintázatokat teszi láthatóvá, de a mélyebb megértéshez gyakran szükség van a hermeneutikai módszerek mikroszkópjára is.

Végül megjegyezzük, hogy több példánk grafikus felületű szoftvert mutatott be a könnyebb érthetőség kedvéért, de a társadalomkutató számára hosszú távon a programozási környezetben (például R-ben vagy Pythonban) végzett elemzés jelenti a legmegbízhatóbb utat. A kód írása ugyanis több alapvető előnnyel jár:

a) Átláthatóság és kontroll: Amíg egy grafikus szoftver fekete dobozként működik, ahol a háttérben zajló folyamatok rejtve maradnak, ezzel szemben a programkód minden egyes lépést – a szöveg tisztításától az algoritmus paramétereinek finomhangolásáig – nyilvánvalóvá és a kutató által ellenőrizhetővé tesz.
b) Reprodukálhatóság: Egy jól dokumentált programkód a tudományos munka legmagasabb szintű reprodukálhatóságát biztosítja. Bárki, beleértve a kutató jövőbeli önmagát is, ugyanazt az elemzést pontosan meg tudja ismételni, ami egy kattintásokon alapuló folyamatnál szinte lehetetlen.
c) Testreszabhatóság: Míg a grafikus szoftverek előre beépített lehetőségekre korlátozódnak, a programozás lehetővé teszi a legújabb módszerek alkalmazását és a folyamat teljes testre szabását a kutatási kérdés egyedi igényeinek megfelelően.

Végül megjegyezzük, hogy míg az LDA a szövegeket elszigetelten kezeli, a haladóbb modellek, mint a Strukturális Téma Modell (Structural Topic Model, STM), lehetővé teszik a kutatók számára, hogy a dokumentumokhoz tartozó metaadatokat (például szerző, keletkezés dátuma, politikai párt) is beépítsék az elemzésbe. Ezáltal olyan kérdéseket is megválaszolhatunk, mint: „Hogyan változott a gazdasági témák aránya a parlamenti felszólalásokban az idő múlásával?” vagy „Eltérően beszélnek-e a különböző pártok politikusai a környezetvédelemről?”. Az STM hidat képez a szövegben rejlő rejtett témák és a dokumentumok külső, strukturális jellemzői között, így mélyebb társadalomtudományi elemzést tesz lehetővé.

4. A minőség biztosítása

A minőségbiztosítás célja kettős: egyrészt validálni az eredményeket, vagyis felmérni, mennyire pontosak és megbízhatóak a modellek, másrészt átláthatóvá tenni a folyamatot, hogy más kutatók megismételhessék és ellenőrizhessék a munkát. A minőségbiztosítási eljárások eltérnek aszerint, hogy a használt NLP-módszer a felügyelt vagy a felügyeletlen gépi tanulás kategóriájába tartozik-e.

4.1. Felügyelt módszerek értékelése

A felügyelt módszerek (mint a névelem-felismerés vagy a gépi tanuláson alapuló szentimentelemzés) esetében a modell egy előre meghatározott kategóriába sorol be szövegeket. Itt a minőség alapja a modell teljesítményének tárgyilagos, számszerű mérése egy ember által létrehozott „igazsághoz" képest. A folyamat két fő lépésből áll: először létre kell hoznunk ezt az igazságot, majd szabványos metrikákkal meg kell mérnünk a modell teljesítményét.

a) Az igazság létrehozása: tanító és tesztadatbázisok

A modell értékelésének alapfeltétele, hogy az adatainkat legalább két, egymástól szigorúan elkülönített részre osszuk:

Tanító adatbázis: Ez az adatok nagyobb része (jellemzően 80-90%), amelyen a modell megtanulja a mintázatokat.

Tesztadatbázis: Ez az adatok kisebbik, érintetlenül hagyott része (10-20%), amelyen a modell még soha nem járt. Ezen mérjük le a modell végső, valós teljesítményét. Az itt szereplő adatokat a kutató(k) kézzel, személyesen kódolja (kódolják), létrehozva az aranyszabványt, amelyhez a modell jóslatai hasonlíthatók. A kódolók közötti egyetértés (például Cohen-kappa) itt is fontos mutatója az emberi kódolás megbízhatóságának.

Lényeges, hogy a tesztadatbázis teljesen érintetlen maradjon a modell tanítási szakaszában. Ha a tesztadatokból származó információ bármilyen módon átszivárog a tanítási folyamatba, a kapott teljesítménymutatók mesterségesen magasak és félrevezetőek lesznek. Megjegyezzük, hogy kisebb adathalmazok esetén egy egyszerű felosztás tanító- és tesztadatra túlságosan függhet a véletlentől (attól, hogy mely adatpontok kerültek a tesztadatbázisba). Ennek a

problémának a kiküszöbölésére szolgál a tízszeres keresztvalidáció. Itt az adatokat tíz egyenlő részre osztjuk, majd tízszer lefuttatjuk az elemzést úgy, hogy mindig egy másik részhalmaz a tesztadatbázis, a maradék kilenc pedig a tanító adatbázis. A tíz futás eredményeinek átlaga egy sokkal robusztusabb, az adathalmaz egészére jellemzőbb teljesítménymutatót ad.

b) A teljesítmény mérése

Miután létrehoztuk a tesztadatbázist, és lefuttattuk rajta a modellt, a teljesítményét szabványos metrikákkal mérjük. Először is fontos kiemelni a kiegyensúlyozatlan osztályok problémáját. A társadalomtudományi adatokban gyakori, hogy egy-egy kategória sokkal ritkábban fordul elő, mint a többi (például egy online fórumon a tiltott tartalmak aránya elenyésző). Ilyenkor a puszta pontosság félrevezető lehet, mert például egy modell, amely soha nem jósol tiltott tartalmat, akár 99%-os pontosságot is elérhet, mégis teljesen használhatatlan. Ezért van szükség olyan metrikákra, amelyek a kisebbségi osztály helyes felismerését is képesek mérni.
A modell teljesítményét a következő három metrikával mérjük:

Precizitás: A modell pozitív jóslatainak pontossága. (Azokból a szövegekből, amelyeket a modell pozitívnak címkézett, mennyi volt tényleg pozitív?) Segít felmérni a téves pozitív találatok arányát.

Felidézés: A releváns esetek megtalálásának képessége. (Az összes ténylegesen pozitív szövegből mennyit talált meg a modell?) Segít felmérni a téves negatív találatok arányát.

Például képzeljünk el egy modellt, amelynek az a feladata, hogy a beérkező e-mailek közül kiszűrje a levélszemetet (a pozitív kategória). Precizitás: Azokból az e-mailekből, amelyeket a modellünk levélszemétnek jelölt, mennyi volt valóban az? A magas precizitás fontos, hogy egy fontos levelet (például állásajánlatot) nehogy tévesen a levélszemét mappába tegyünk (kevés téves pozitív). Felidézés: Az összes valódi levélszemét e-mailből mennyit sikerült a modellünknek elkapnia? A magas felidézés fontos, hogy a postaládánkban a lehető legkevesebb kéretlen levél maradjon (kevés téves negatív).

Látható, hogy a két cél gyakran ellentétes. Ha mindent levélszemétnek jelölünk, a felidézés 100% lesz, de a precizitás borzalmas. Itt jön képbe a harmadik metrika:

F1-érték: A precizitás és a felidézés harmonikus átlaga, amely egyetlen számban összegzi a modell teljesítményét. Az F1-érték segít megtalálni az egyensúlyt a két metrika között, és a legtöbb esetben ez a legfontosabb mutató, mivel mindkét típusú hibát (a fontos levelek kidobását és a kéretlen levelek átengedését) figyelembe veszi.

4.2. Felügyeletlen módszerek értékelése

A felügyeletlen módszerek (mint a témaelemzés) esetében a minőségbiztosítás alapvetően más logikát követ, mint a felügyelt modelleknél. Mivel itt nincsenek előre meghatározott helyes válaszok, a hagyományos pontosság nem mérhető. A minőség itt a feltárt mintázatok koherenciáján, értelmezhetőségén és kutatói hasznosságán múlik. A validáció egy többlépcsős folyamat, amely ötvözi az emberi interpretációt és a statisztikai mutatókat.

a) Az emberi értelmezés

A validáció első és legfontosabb lépése a kutató szakértői tudásán alapuló minőségi ellenőrzés. Ez a legfőbb validációs eszköz, amely két központi elemből áll. Az első a szemantikai koherencia vizsgálata: a feltárt témákhoz tartozó kulcsszólisták értelmesek? A szavak logikusan összetartoznak? Egy jó téma szavai intuitívan lefednek egy közös fogalmi mezőt, míg egy rossz témáé egy véletlenszerű szóhalmaznak tűnik. A második, elengedhetetlen lépés a keresztellenőrzés (trianguláció). A kutatónak el kell olvasnia több, a modell által egy adott témába magas valószínűséggel besorolt dokumentumot. A szövegek tényleges tartalma valóban megfelel-e a téma kulcsszavai által sugallt jelentésnek? Ez a lépés kritikus fontosságú, hogy elkerüljük a statisztikai anomáliák túlinterpretálását, és biztosítsuk, hogy a témáink valódi, tartalmi jelentéssel bírnak.

b) Statisztikai mutatók

Az emberi értelmezés folyamatát támogathatjuk szoftverek által automatikusan számolt mutatókkal. Ezek közül a legfontosabbak a témakoherencia-mutatók. E belső metrikák azt mérik, hogy egy téma legfontosabb kulcsszavai milyen gyakran fordulnak elő együtt a korpusz dokumentumaiban. A magasabb koherenciaérték általában jobb, értelmesebb témákra utal. A mutatók hasznosak a modell paramétereinek, különösen a témák számának (k) kedvező beállításához.

Azonban nem helyettesítik az emberi értelmezést, mivel egy statisztikailag koherens téma nem feltétlenül hasznos vagy lényeges a kutatás szempontjából.

c) A témák külső érvényessége

A validáció legmagasabb szintjét a külső mutatók jelentik. Itt azt vizsgáljuk, hogy a témaelemzés eredményei hasznosak-e egy másik, független feladat megoldásában. Ha a témák valóban lényegesek, vagyis rejtett szerkezetet ragadtak meg az adatokban, akkor prediktív erejüknek kell lennie. Például a diákok fórumhozzászólásai mellett ismerjük a végső vizsgajegyeiket is. Ha a témaelemzés által feltárt „Vizsgastressz" téma aránya egy diák hozzászólásaiban képes bejósolni a rosszabb vizsgajegyet, akkor a témánk valószínűleg külső érvényességgel is rendelkezik. Ez a megközelítés bizonyítja a legmeggyőzőbben, hogy a modellünk nemcsak statisztikai mintázatokat, hanem a valósággal összefüggő, értelmes struktúrákat tárt fel.

4.3. Általános minőségi követelmények minden NLP-kutatásban

A fentieken túlmenően, minden NLP-alapú társadalomtudományi kutatás minőségét és tudományos hitelességét néhány általános alapelv szisztematikus alkalmazása alapozza meg. Az első és legfontosabb ilyen elv a megismételhetőség. Ez a gyakorlatban azt jelenti, hogy a kutatási jelentésnek olyan részletességgel kell dokumentálnia a teljes folyamatot, hogy más kutatók elviekben képesek legyenek ugyanazokat az eredményeket létrehozni. Ennek érdekében a leírásnak tartalmaznia kell az adatgyűjtés forrásait, az adat-előkészítés pontos lépéseit (például a tokenizáció módját vagy a felhasznált stopszavak listáját), a használt szoftver(ek) és modell(ek) pontos verziószámát, valamint minden beállított paramétert (például a témák számát).

A második alapelv a szisztematikus értékelés. Ahogy a korábbi alfejezetekben részletesen bemutattuk, ennek módszertana a feladat típusától függ: a felügyelt módszereknél a szabványos teljesítménymutatók (precizitás, felidézés, F1-érték) jelentését, míg a felügyeletlen módszereknél a hermeneutikai és matematikai validálás lépéseinek alapos dokumentálását jelenti.

A harmadik, gyakran elhanyagolt, de a minőség szempontjából kritikus elem a hibaelemzés. Egy hiteles kutatás nemcsak a sikereket, hanem a modell tipikus hibáit is bemutatja. A kutatónak fel kell tennie és meg kell válaszolnia a kérdést: milyen típusú eseteket kezel a modell rosszul? Például miután betanítottunk

egy szentimentelemző modellt, a hibaelemzés során célzottan megvizsgáljuk azokat a szövegeket, amelyeket a modell a legmagabiztosabban, mégis hibásan osztályozott. Kiderülhet például, hogy a modell következetesen félreérti az iróniát tartalmazó mondatokat, vagy nem ismeri fel egy szűk szakmai közösség szlengjét. A hibák mintázatainak feltárása növeli a kutatás hitelességét, és mélyebb megértést nyújt a modell működéséről.

Végül a negyedik alapelv a korlátok tudatosítása. A kutatónak reflektálnia kell a választott módszer beépített korlátaira (például arra, hogy a témaelemzés nem veszi figyelembe a szórendet) és az adathalmazból fakadó lehetséges torzításokra. Az eredmények értelmezése során ezeket a korlátokat mindig szem előtt kell tartani. Ennek a kritikus szemléletnek az átlátható bemutatása az eredmények árnyalt és tudományosan megalapozott interpretációjának előfeltétele.

4.4. Modellek összehasonlítása és statisztikai szignifikancia

A kutatások során gyakran merül fel a kérdés: az A modell (például egy bonyolultabb neurális háló) valóban jobban teljesít, mint a B modell (például egy egyszerűbb szózsákmodell)? A tesztadatbázison mért teljesítménykülönbség lehet, hogy csak a véletlen műve, azaz az adott tesztadatok sajátosságainak köszönhető. A statisztikai szignifikanciatesztek segítenek eldönteni, hogy a modellek közti különbség valós-e. A gyakorlatban ez azt jelenti, hogy teszteljük a nullhipotézist, miszerint a két modell valójában egyformán jó. Ha a teszt (például binomiális teszt vagy bootstrapalapú teszt) p-értéke egy előre meghatározott küszöb (például $p < 0{,}05$) alá esik, akkor elvethetjük a nullhipotézist, és a modellek közti különbséget szignifikánsnak tekinthetjük. Ezzel az eredmények tudományos megalapozottsága jelentősen növelhető.

5. Mikor használjuk és mikor ne használjuk a módszert?

A természetesnyelv-feldolgozás módszerei alapvetően más típusú kutatási kérdésekre adnak választ, mint a hagyományos hermeneutikai jellegű elemzések. Nem az egyedi, mély megértésre, hanem a nagy léptékű mintázatok feltárására, számszerűsítésére és a szöveges adatok önműködő osztályozására összpontosítanak. A megfelelő módszer kiválasztása a kutatási kérdés, a rendelkezésre álló adatok és a kutatás jellegének gondos mérlegelését igényli.

5.1. A kutatási kérdés

Az NLP akkor a leghatékonyabb, ha a kutatót nagy mennyiségű szöveges adatban rejlő, aggregált szintű mintázatok, trendek vagy kategóriák érdeklik. A kérdések a „mi?", „melyik?", „mennyi?" és a „hogyan oszlik meg?" kérdőszavakra épülnek, és gyakran összehasonlító vagy leíró jellegűek. A cél a szövegek tartalmának számszerűsítése és strukturálása. Néhány jó NLP-kérdés:

a) Melyek a domináns diskurzusok a klímaváltozásról a magyar politikai pártok online kommunikációjában? (Témaelemzés)
b) Hogyan változott egy új termék fogadtatása az online vásárlói véleményekben az idő múlásával? (Szentimentelemzés)
c) Mely cégek és intézmények jelennek meg leggyakrabban a mesterséges intelligenciáról szóló híradásokban? (NER)

Néhány példa kerülendő, nem NLP-jellegű kérdésre:

a) Hogyan éli meg egy páciens a krónikus betegséggel való együttélést a naplóján keresztül? (Ez mély, idiografikus megértést igényel, amire az IPA alkalmasabb.)
b) Miért vezet a politikai polarizáció a közösségi médiában nyelvi agresszióhoz? (Ez egy ok-okozati kérdés, amelyre az NLP önmagában nem ad választ, legfeljebb korrelációkat tárhat fel.)

5.2. A minta

Az NLP a nagy minták módszertana. A modellek statisztikai alapokon működnek, így a megbízható eredményekhez nagy mennyiségű szöveges adatra van szükség. A minta két fő jellemzője: A minta mérete a feladattól függ. Témaelemzéshez vagy más felfedező jellegű elemzéshez több száz, de inkább több ezer dokumentum (például hírcikk, hozzászólás, interjú) szükséges, hogy a statisztikai mintázatok megbízhatóan kirajzolódjanak. A felügyelt gépi tanuláshoz (például egy egyedi szentimentelemző modell tanításához) a tanító adatbázisnak több ezer vagy akár több tízezer kézzel címkézett példát kell tartalmaznia a jó teljesítményhez.

A reprezentativitással kapcsolatban meg kell jegyezni, hogy a valóságban az NLP-elemzések nagyon gyakran kényelmi mintákkal dolgoznak, például egy könnyen elérhető online fórum összes hozzászólásával. A kutatónak tisztában kell lennie azzal, hogy az ilyen minták nem reprezentatívak a teljes populációra nézve. Az eredmények kizárólag a vizsgált korpuszra érvényesek. A következtetések levonásakor ezt a korlátot mindig hangsúlyozni kell.

Az NLP esetében ritkán beszélhetünk túl nagy mintáról. A korlátot inkább a számítási kapacitás és a feldolgozási idő jelenti, nem pedig módszertani megfontolás.

5.3. A kutatás jellege

Az NLP a legjobb választás, ha a kutatást a következők jellemzik:

a) A kutatás célja a feltárás és leírás nagy léptékben. A cél a szövegtengerben rejlő mintázatok, struktúrák, trendek azonosítása, amelyeket kézzel lehetetlen vagy rendkívül erőforrás-igényes lenne feltárni.
b) A kutatás a szöveg tartalmának számszerűsítésére irányul. A kutató a szöveges információt strukturált adattá szeretné alakítani (például témák aránya, vélemények polaritása, szereplők gyakorisága), hogy aztán statisztikai elemzéseknek vesse alá.
c) A kutató elfogadja a felülnézeti szempontot. Az NLP nem az egyedi szövegek mély, kontextuális jelentését tárja fel, hanem az aggregált adatokban rejlő mintázatokat.
d) A kutatás a hermeneutikai és matematikai módszerek ötvözésére törekszik (mixed-methods). Az NLP kiválóan alkalmas egy vegyes módszertanú

kutatás első, feltáró szakaszára. A témaelemzéssel azonosított fő diskurzusokat például később mélyinterjúkban vagy fókuszcsoportokban lehet részletesebben vizsgálni.

Ne használjon NLP-t (önmagában), ha:

a) A kutatás a mély, szubjektív jelentésrétegekre kíváncsi. Ha a kutató arra kíváncsi, hogyan él meg egy személy egy életeseményt, vagy milyen rejtett, kontextuális jelentései vannak egy beszélgetésnek, a hagyományos hermeneutikai módszerek (például IPA, diskurzuselemzés) megkerülhetetlenek.
b) A kutatási minta rendkívül kicsi. Néhány interjú vagy dokumentum elemzésekor az NLP-modellek statisztikai alapja instabil, így az eredmények megbízhatatlanok. A felügyelt gépi tanulás logikája arra épül, hogy a modell a kutató által felcímkézett, értelmezett példákból tanulja meg a mintázatokat, majd ezt a tudást alkalmazza nagy mennyiségű új adatra. Lényegében az emberi döntések kiterjesztését és automatizálását végzi. Ez a folyamat azonban mindig kompromisszummal jár: a skálázhatóságért cserébe feláldozunk valamennyit abból a pontosságból és kontextuális mélységből, amelyre csak az emberi értelmezés képes. Kis adathalmaz esetén ez a kompromisszum értelmetlen. Ha a kutató el tudja végezni a teljes minta manuális elemzését (például néhány tucat interjú esetében), akkor a hermeneutikai módszerek, mint a tematikus analízis vagy a tartalomelemzés, nemcsak pontosabb, de árnyaltabb és módszertanilag megalapozottabb eredményt is fognak adni.
c) A kutatás ok-okozati összefüggéseket keres. Az NLP-modellek korrelációkat és együtt járásokat tárnak fel, de oksági kapcsolatokat nem bizonyítanak.
d) A kutatási kontextus rendkívül sajátos és a nyelvezet ironikus vagy szubtilis. A legtöbb önműködő eszköz (különösen a szótáralapúak) nehezen birkózik meg az iróniával, a szarkazmussal és az összetett nyelvi jelenségekkel.

Előnyök	Hátrányok
Nagy lépték és hatékonyság: hatalmas mennyiségű szöveges adat gyors és szisztematikus feldolgozását teszi lehetővé.	Felszínesség: A modellek a szavak statisztikai mintázatait elemzik, nem a mély, kontextuális jelentést. A hogyan és miért kérdésekre korlátozottan adnak választ.
Objektivitás és megismételhetőség: a szabályalapú, algoritmikus folyamat csökkenti a kutatói szubjektivitásból fakadó torzításokat, és biztosítja az eredmények reprodukálhatóságát.	Technikai korlátok és hibák: a modellek nem tévedhetetlenek (például rosszul klasszifikálnak), teljesítményük nagyban függ az adatok minőségétől és az előkészítés gondosságától.
Új mintázatok feltárása: képes olyan rejtett struktúrákat és korrelációkat azonosítani, amelyek az olvasás során rejtve maradnának.	Adatigény: megbízható működésükhöz nagy mennyiségű adatra van szükségük.
Hermeneutikai és matematikai kutatás ötvözése: kiválóan alkalmas vegyes módszertanú kutatásokban a hermeneutikai és matematikai szakaszok összekötésére	Kontextusérzéketlenség: sok modell (például BoW-alapúak) nem veszi figyelembe a szórendet, és nehezen kezeli az iróniát, szarkazmust és a nyelvi kétértelműséget.

18. táblázat: A módszerválasztás mérlegelése: az NLP előnyei és hátrányai

6. A módszer változatai és haladó alkalmazásai

Az elméleti alapok fejezetében rámutattunk a dokumentum-term mátrix (DTM) egyik fő gyengeségére: a modell nem tudja, hogy a „király” és az „uralkodó” szavak jelentése hasonló. A Word Embedding technológia erre a problémára ad választ. A módszer lényege, hogy ahelyett, hogy a szavakat egyszerűen csak megszámolná, a szóbeágyzó modellek (mint a híres Word2Vec vagy a GloVe) megtanulják a szavak jelentését a környezetükből. A „szó jelentését a környezete határozza meg” elv alapján minden szót egy több száz dimenziós numerikus vektorral (egy számsorral) reprezentálnak. A lényeg, hogy a hasonló

környezetben használt szavak vektorai ebben a többdimenziós jelentéstérben közel kerülnek egymáshoz.

A gyakorlatban a kutatónak ritkán van szüksége arra, hogy egy ilyen modellt a nulláról, saját maga tanítson be, ami hatalmas adathalmazt és számítási kapacitást igényelne. A modern NLP egyik legnagyobb előnye, hogy számos, gigantikus méretű korpuszokon (például a teljes Wikipédia szövegén) előtanított szóbeágyazás ingyenesen elérhető online. A kutató számára a valódi módszertani döntés az, hogy a kutatási kérdéséhez melyik megközelítés illik jobban:

a) Használjon egy általános célú, előtanított modellt, ami akkor ideális, ha a szövegei általános nyelvezetet használnak.
b) Vagy tanítson egy saját, témaspecifikus modellt a saját korpuszán, ami akkor lehet indokolt, ha a szövegek nagyon speciális szakzsargont tartalmaznak (például orvosi leletek, jogi szövegek), ahol a szavak jelentése eltér a köznyelvitől.

A klasszikus embedding modellek (mint a Word2Vec) egy szűk, lokális kontextusablakot (például a vizsgált szó előtti és utáni öt-öt szót) vesznek figyelembe, míg a transzformer architektúrára épülő modellek (lásd később) az egész mondat környezetét képesek dinamikusan kezelni, ami még árnyaltabb jelentésreprezentációt tesz lehetővé. A következőkben néhány olyan haladó alkalmazást mutatunk be, amely ezen a technológián alapul a társadalomkutatók számára:

a) Dokumentumok reprezentálása felügyelt tanuláshoz: A szóbeágyazások talán legfontosabb gyakorlati alkalmazása, hogy a klasszikus, ritka DTM-DFM mátrixokat egy sokkal sűrűbb, szemantikailag gazdag reprezentációra cserélik. Szövegosztályozási feladatoknál (például egy online komment pozitív vagy negatív hangulatának eldöntése) a dokumentumot már nem a benne lévő szavak listája, hanem a benne lévő szavak vektorainak átlaga vagy összege képviseli. Ez a jelentésvektor sokkal jobban megragadja a szöveg tartalmát, így a felügyelt modellek pontosabb előrejelzéseket tudnak adni.
b) Szemantikai hasonlóság mérése: Mérhetővé válik, hogy két fogalom milyen közel áll egymáshoz egy adott korpuszban. Például megvizsgálhatjuk, hogy a szabadság és a biztonság szavak milyen távol vagy közel helyezkednek el a különböző politikai oldalak kommunikációjában.
c) Analógiák és rejtett viszonyok feltárása: A szóbeágyzások képesek megragadni bonyolult viszonyokat: a modellből matematikailag kimutatható, hogy a (király – férfi + nő) vektor nagyon közel áll a királynő vektorához.

d) Társadalmi torzítások vizsgálata: Mivel a szóbeágyazások az emberi szövegekből tanulnak, a bennük rejlő sztereotípiákat is megtanulják. Kimutathatóvá válnak a nemi (például az orvos a férfihoz, a nővér a nőhöz áll közelebb) vagy etnikai alapú torzítások a közbeszédben.

A Word2Vec nagy előrelépés, de van egy lényeges korlátja: minden szónak egyetlen, fix vektora van, függetlenül a környezettől. A magyar „levél" szónak ugyanaz a reprezentációja, akár egy fáról esik le, akár a postás hozza. A transzformer architektúrára épülő modellek (mint a BERT) ezt a problémát oldják meg. A módszer lényege, hogy a transzformer modellek a szöveg olvasása közben, dinamikusan hozzák létre a szavak környezethez igazított beágyazásait. A „levél" szónak más lesz a vektora egy botanikai és egy postai témájú mondatban. Ezek a modellek a mondat (vagy akár több mondat) teljes környezetét figyelembe veszik, így képesek a nyelvi kétértelműséget és a finomabb jelentésárnyalatokat is kezelni. Haladó alkalmazások társadalomkutatóknak:

a) Pontosabb osztályozás: Mivel a modell jobban érti a környezetet, a korábban bemutatott feladatokat (például szentimentelemzés, NER) sokkal nagyobb pontossággal képes elvégezni. Képes például felismerni az iróniát, ahol a hangulatot a környezet fordítja az ellenkezőjére.
b) Szövegrészek közötti logikai kapcsolatok vizsgálata (Natural Language Inference – NLI): A modell képes megállapítani, hogy két mondat között milyen logikai viszony áll fenn (például az egyik következik a másikból, ellentmond neki, vagy független tőle). Ezzel bonyolult érvelési láncok is vizsgálhatóvá válnak.

A fejlődés jelenlegi csúcsát a nagy nyelvi modellek (Large Language Models, LLM-ek), mint például a GPT-család modelljei jelentik. Ezek a transzformeralapú modellek hatalmas, az internet nagy részét lefedő korpuszokon tanulnak. Különlegességük, hogy nemcsak elemzésre, hanem szöveg létrehozására is képesek, és általános célú nyelvi eszközként használhatók. A módszer lényege, hogy az LLM-ek az ember által megfogalmazott utasítás vagy kérdés alapján hajtják végre a feladatokat. A kutatónak nem kell modellt tanítania, hanem meg kell tanulnia hatékonyan kérdezni a modelltől. E területen jelenleg a haladó alkalmazások a következők:

a) Zero-shot osztályozás: A kutatónak nem kell több ezer hozzászólást kézzel címkéznie egy szentimentelemző tanításához. Ehelyett egyszerűen megkérheti az LLM-et, hogy klasszifikálja a szövegeket egy prompt segítségével (például „Döntsd el a következő hozzászólásról, hogy a hangulata pozitív, negatív vagy semleges! Hozzászólás: [...]").
b) Segédkódolás hermeneutikai kutatásokban: Az LLM használható a kódolás felgyorsítására. A kutató beadhat egy interjúrészletet, és megkérheti a modellt, hogy az előre meghatározott kódkönyv alapján javasoljon kódokat, vagy foglalja össze a szöveg főbb témáit. Ez nem helyettesíti, de rendkívüli módon felgyorsíthatja a kutatói munkát.
c) Adatszintetizálás és -bővítés: A modell használható kérdőíves kérdések vagy interjúkérdések átfogalmazására, variációinak létrehozására vagy akár kitalált, de realisztikus válaszok generálására egy pilotteszthez. (Ez utóbbi etikai megfontolásokat igényel.)

7. Összefoglalás

A természetesnyelv-feldolgozás egy lendületesen fejlődő, interdiszciplináris terület, amely a társadalomkutatók számára is új távlatokat nyit a nagy mennyiségű szöveges adat elemzésében. A fejezet bemutatta, hogy az NLP módszertani hídként működik, amely a strukturálatlan szövegeket a statisztikai elemzés számára is értelmezhető, strukturált adattá alakítja. Az elméleti alapoknál megismertük a folyamat kulcsát: a szöveg numerikus reprezentációvá alakítását, amely a hagyományos szózsákmodelltől a korszerű, jelentésalapú szóbeágyazásokig terjed. Három alapvető, a gyakorlatban is széles körben használt módszert vizsgáltunk meg: a névelem-felismerés a szövegek szereplőinek azonosítására, a szentimentelemzés a vélemények polaritásának mérésére, a témaelemzés pedig a rejtett tematikus struktúrák feltárására szolgál. Láthattuk, hogy bár grafikus felületű szoftverekkel ezek a módszerek már programozói tudás nélkül is elérhetők, a felelős alkalmazásuk tudatos módszertani döntéseket igényel, különösen az adatelőkészítés során. A minőségbiztosításról szóló rész hangsúlyozta a felügyelt és felügyeletlen modellek értékelése közötti különbséget. Míg a felügyelt modelleknél a pontosságot szabványos metrikákkal (precizitás, felidézés, F1-érték) mérjük egy kézzel kódolt aranyszabványhoz képest, addig a felügyeletlen modelleknél

a minőség a feltárt mintázatok koherenciáján és a kutatói értelmezésen múlik. A fejezet végül bemutatta a korszerű NLP legizgalmasabb fejlesztéseit, a nyelvi környezetet is értő transzformer modelleket és a nagy nyelvi modelleket, amelyek forradalmasítják a szöveges adatokkal való munkát.

8. Önellenőrző kérdések

1. Mi a legfontosabb különbség a szótáralapú és a gépi tanuláson alapuló szentimentelemzés között, és ez milyen gyakorlati következményekkel jár a kutató számára?

Válasz: A szótáralapú megközelítés egy előre meghatározott, kötött szabályrendszer (a pozitív/negatív szavak listája) alapján működik, míg a gépi tanuláson alapuló modell ember által címkézett példákból, statisztikai mintázatok alapján tanulja meg a vélemények felismerését. A gyakorlati következmény, hogy a szótáralapú módszer gyors és nem igényel tanító adatbázist, de pontossága erősen függ a szótár minőségétől és doménspecifikusságától. A gépi tanulás modellje sokkal rugalmasabb és pontosabb lehet, de létrehozása rendkívül erőforrás-igényes, mivel több ezer, kézzel címkézett példára van szüksége a tanításhoz.

2. Egy kutató azt tapasztalja, hogy az öngyilkossági szándékra utaló online hozzászólásokat felismerő modelljének pontossága 99,9%. Miért lehet ez az eredmény mégis teljesen félrevezető, és melyik két metrikát kellene helyette vizsgálni?

Válasz: Az eredmény azért lehet félrevezető, mert az öngyilkossági szándékot jelző hozzászólások rendkívül ritkák egy nagy online fórumon (ez a kiegyensúlyozatlan osztályok problémája). Ha egy modell soha nem jósol ilyen szándékot (minden szövegre azt mondja, hogy nem veszélyes), de ezek a bejegyzések az összes hozzászólásnak csak 0,1%-át teszik ki, a modell pontossága 99,9% lesz, miközben teljesen használhatatlan, mert egyetlen, kritikusan fontos esetet sem talál meg. Helyette a precizitást és a felidézést (és azok harmonikus átlagát, az F1-értéket) kell vizsgálni. Ebben az értelemben a felidézés a legfontosabb metrika: megmutatja, hogy az összes valódi, segítségkérő hozzászólásból mennyit

sikerült a modellnek azonosítania. Itt a cél a lehető legmagasabb felidézés, hogy egyetlen veszélyben lévő személy se maradjon észrevétlen. A precizitás (hogy amit veszélyesnek jelölünk, az tényleg az-e) másodlagos, mivel egy téves riasztás (egy nem veszélyes hozzászólás megjelölése) következménye sokkal kisebb, mint egy valódi segélykiáltás figyelmen kívül hagyása.

3. Milyen típusú kutatási kérdés esetén választaná az LDA-témaelemzést egy hagyományos, a kutató által személyesen végzett tematikus elemzéssel szemben? Adjon egy konkrét példát!

Válasz: Az LDA-témaelemzést akkor választanám, ha egy nagyon nagy korpusz (több száz vagy ezer dokumentum) rejtett, általános tematikus struktúráinak feltárására vagyok kíváncsi, és a célom egy felülnézeti, kvantitatív képalkotás. Például: „Melyek a domináns diskurzusok a klímaváltozásról több ezer magyar nyelvű hírcikkben az elmúlt tíz évben?" Ezzel szemben a személyes tematikus elemzést akkor választanám, ha egy kisebb mintán (például 20 interjú) a mély, kontextuális, szubjektív jelentésrétegeket szeretném megérteni. Például: „Hogyan élik meg az aktivisták a klímaszorongást?"

4. Mi a szózsák-feltételezés, és melyik a legegyszerűbb módszer a korlátainak enyhítésére?

Válasz: A szózsák-feltételezés az a radikális egyszerűsítés, hogy egy dokumentum jelentésének megragadásához elegendő a benne lévő szavak gyakoriságát ismernünk, a szavak sorrendje és a nyelvtani szerkezet pedig elhanyagolható. Ennek legismertebb korlátja, hogy a „kutya kergeti a macskát" és a „macska kergeti a kutyát" mondatokat azonosnak tekinti. A legegyszerűbb módszer ennek enyhítésére az n-gram modellek használata, amelyek nemcsak önálló szavakat, hanem szópárokat (bigramok), szóhármasokat (trigramok) stb. is figyelembe vesznek, így a kontextus egy kis részletét megőrzik.

5. Mi a legfontosabb különbség a hagyományos Word2Vec és a korszerű BERT-típusú szóbeágyazások között, és ez miért jelent forradalmat a szövegelemzésben?

Válasz: A legfontosabb különbség, hogy a Word2Vec minden szónak egyetlen, fix numerikus vektort (jelentést) ad, függetlenül a környezettől (a „levél" szónak ugyanaz a vektora, akár fáról esik, akár a postás hozza). Ezzel szemben a BERT-típusú modellek kontextualizált szóbeágyzásokat hoznak létre, vagyis

a „levél” szónak más lesz a vektora egy botanikai és egy postai témájú mondatban. Ez azért forradalmi, mert lehetővé teszi a modellek számára a nyelvi kétértelműség kezelését és a jelentés sokkal finomabb, a valós emberi nyelvhasználathoz közelebb álló megértését, ami drámaian megnöveli az olyan feladatok pontosságát, mint a szentimentelemzés vagy az irónia felismerése.

Ajánlott irodalom

Arseniev-Koehler, A. (2024). Theoretical foundations and limits of word embeddings: What types of meaning can they capture? *Sociological Methods & Research, 53*(4), 1753–1793. https://doi.org/10.1177/00491241221140142

Carlson, R., Bauer, J., & Manning, C. D. (2025). *A New Pair of GloVes*. https://doi.org/10.48550/arXiv.2507.18103

Dunivin, Z. O. (2025). Scaling hermeneutics: a guide to qualitative coding with LLMs for reflexive content analysis. *EPJ Data Science, 14*(1). https://doi.org/10.1140/epjds/s13688-025-00548-8

Eisenstein, J. (2019). *Introduction to natural language processing. Adaptive computation and machine learning*. The MIT Press.

Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alsahfi, T., & Alshemaimri, B. (2025). BERT applications in natural language processing: a review. *Artificial Intelligence Review, 58*(6). https://doi.org/10.1007/s10462-025-11162-5

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. https://doi.org/10.48550/arXiv.2203.05794

Hankar, M., Kasri, M., & Beni-Hssane, A. (2025). A comprehensive overview of topic modeling: Techniques, applications and challenges. *Neurocomputing, 628*, 129638. https://doi.org/10.1016/j.neucom.2025.129638

Johnson, S. J., Murty, M. R., & Navakanth, I. (2024). A detailed review on word embedding techniques with emphasis on word2vec. *Multimedia Tools and Applications, 83*(13), 37979–38007. https://doi.org/10.1007/s11042-023-17007-z

Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). *stm : An R Package for Structural Topic Models*. *Journal of Statistical Software, 91*(2). https://doi.org/10.18637/jss.v091.i02

Zhang, B., Zhou, Y., & Li, D. (2025). Can Human Reading Validate a Topic Model? *Sociological Methodology, 55*(1), 59–90. https://doi.org/10.1177/00811750241265336

VIII. MEGERŐSÍTŐ FAKTORELEMZÉS (CFA)

SOLTÉSZ-VÁRHELYI KLÁRA

A fejezet a következő kérdéseket válaszolja meg:

a) Milyen elméleti alapokon nyugszik a CFA?
b) Mik a CFA gyakorlati lépései, hogyan kell az elemzést helyesen beállítani, az eredményeket helyesen értelmezni?
c) Milyen minőségi ismérvek alapján különböztethető meg egy kiváló CFA-tanulmány egy átlagostól?
d) Milyen további elemzési lehetőségek vannak a CFA során?

1. Bevezetés

A megerősítő faktorelemzés (CFA) egy olyan statisztikai módszer, amelyet a látens változók (faktorok) és a megfigyelt változók (itemek, tételek, indikátorok) közötti kapcsolatok vizsgálatára használunk. Lényege egy előzetesen meghatározott elméleti modell tesztelése, amelynek során a kutató megadja, melyik itemek tartoznak egy-egy faktorhoz, és a CFA ellenőrzi, hogy az adatok menynyire illeszkednek ehhez a feltételezett szerkezethez. Ezáltal a CFA nem feltáró jellegű, hanem elsődlegesen megerősítő elemzés, amely elméleti hipotéziseket vizsgál.

A CFA gyakorlati haszna elsősorban kérdőívek és skálák pszichometriai vizsgálatában jelenik meg. Segít annak ellenőrzésében, hogy a mérőeszköz valóban azzal a struktúrával rendelkezik-e, mint ahogy azt tervezték (például visszaigazolják-e valós kitöltések a feltételezett alskálák létezését, és egy adott alskálához tényleg azok a tételek tartoznak-e, amelyeket a kutatók feltételeztek). Új mérőeszközök fejlesztése mellett fontos eleme a kérdőívek más nyelvre, más populációra való adaptációja során a validációnak is. A CFA így a társadalom- és viselkedéstudományi kutatások egyik alapvető eszköze, amikor a mérési modellek elméleti megalapozottságát szeretnénk statisztikai úton igazolni.

2. A CFA koncepciója

A kutatás során gyakran fordul elő, hogy olyan jelenséget, tulajdonságot szeretnénk vizsgálni, amely nem mérhető közvetlenül. Az emberek viszonylag pontosan meg tudják ítélni és jelölni egytől ötig terjedő Likert-típusú skálán az egészségi állapotukat, de nem tudják pontosan megítélni az extraverziójukat vagy az introjektált motivációjuk mértékét. Ez utóbbi esetekben a vizsgálandó jelenség helyett annak viselkedésben és attitűdökben megjelenő kivetüléseit mérjük. Azaz nem arra kérdezünk rá, hogy mennyire extrovertált valaki, mert azt nem tudja pontosan megítélni, hanem hogy mennyire szeret emberek között lenni, mennyire szeret kipróbálni új dolgokat, mennyire keresi az élményeket, és az ezekre a kérdésekre adott válaszaiból következtetünk a mérni kívánt jelenségre, jelen esetben az extraverzióra.

2.1. A modell specifikációja (model specification)

A vizsgálni kívánt tulajdonságot az elemzés során *látens változó*nak, a mérésére szolgáló változókat *megfigyelt* vagy *manifeszt változó*knak nevezzük. A közvetlenül nem mérhető látens változók mérhető *indikátorai* a manifeszt változók. A látens változók és manifeszt változók kapcsolatát leíró modellt *mérési modell*nek nevezzük.[1]

A 10. ábra egy példa egy ilyen mérési modellre. Látható, hogy két látens változót tartalmaz (téri és idői rendezettség), melyeket *ovális*sal jelölünk. Mindkét látens változóhoz három-három manifeszt változó (kérdőívtétel, *item*) tartozik, melyeket *téglalap*pal jelölünk. A látens és manifeszt változókat *egyirányú nyíl* köti össze, ezeket nevezzük *töltések*nek vagy *faktorterhelés*nek. A nyíl a látens változókból mutat a manifesztekbe, mivel a belső tulajdonság fejeződik ki a viselkedésben/attitűdben, úgy is mondhatjuk, hogy a látens változó „tölt" a manifesztre.[2] A két látens változót *kétirányú nyíl* köti össze, ez a két változó közötti *kovariancia* jelölése, ebben az esetben az összefüggést feltételezzük, de irányát nem teszteljük. (Mi a kovariancia? Két változó együtt járásának mértéke, megmutatja, hogy azok értékei hajlamosak-e együtt növekedni vagy csökkenni. A kovariancia standardizált formája a korrelációs együttható. Illetve ha egy változó önmagával való kovarianciáját számoljuk, akkor a változó *varianciáját* kapjuk meg, azaz azt, hogy mekkora változatosság van az adott változóban.)

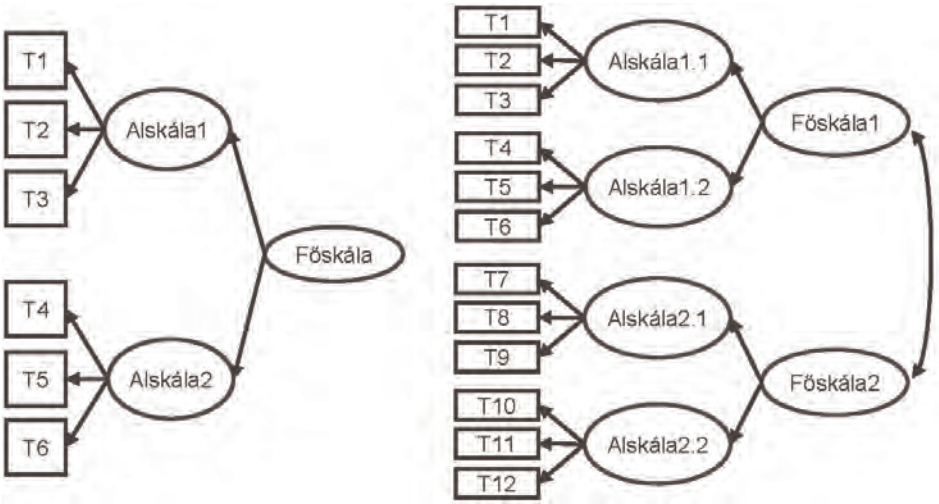
[1] A látens változók mérésének bemutatása során a klasszikus pszichometriai elmélet azon alapvetésére támaszkodunk, mely szerint a közvetlenül nem megfigyelhető konstruktumok becslése kizárólag indirekt úton, a manifeszt változók közvetítésével valósítható meg (Bollen, 1989).

[2] Az összefüggés ezen irányát reflektív modellnek nevezzük, és ez az alapvető feltételezése a klasszikus SEM- és CFA-modelleknek (Bollen, 1989; Brown, 2015). A SEM-fejezetben látunk majd példát az ellenkező irányra is, ahol a tételekből mutat majd nyíl a látens változóba, mely elrendezést formatív modellnek nevezzük majd.

10. ábra: Mérési modell két faktorral, mindkét faktoron három-három tétellel

Az előbbiekben példát láttunk az úgynevezett *elsőrendű (first-order) CFA*-ra, ahol a modell a látens változóknak csupán egyetlen szintjét feltételezi. A manifeszt változók ezekre az elsőrendű faktorokra töltenek, és a modell nem tételez fel egy főfaktort (másodrendű faktort), amely magyarázná az elsőrendű faktorok közötti kapcsolatot. Ezzel szemben alkotható *másodrendű (second-order) CFA* is, ahol az elsőrendű látens változókat egy magasabb szintű látens változó magyarázza. A mi példánkban elképzelhető, hogy a téri és idői rendezettség mögött egy általánosabb értelemben vett rendezettség áll. Ennek modelljét láthatjuk a 11. ábra bal oldalán. Az is elképzelhető, hogy az ilyen főskálából nemcsak egy, hanem több is van a modellben, melynek grafikus megjelenítése a 11. ábra jobb oldalán látható. Jó példa a másodrendű CFA-ra a széles körben ismert NEO-PI-R kérdőív,[3] ahol az öt nagy személyiségdimenzió mindegyike hat alskálára bomlik, és minden alskálát nyolc tétel mér.

[3] NEO-PI-R: Costa & McCrae (2008). Továbbá érdemes megfigyelni a MBI-kérdőívet (Maslach et al. 2018). A kérdőív a kiégést három alskála mentén méri, melyek a kimerülés, a deperszonalizáció és a teljesítménycsökkenés (a nevek az évek során változtak). A szerzők eredeti állásfoglalása szerint ez a három alskála önálló, és nem összegződik egyetlen általános kiégés főskálába. Ennek megfelelően a faktorszerkezete az eredeti elképzelés szerint elsőrendű. Ezzel szemben a BAT (Schaufeli et al., 2019) a kiégést négy alskála mentén méri, és hivatalosan is létezik a főskálája, mely az alskálák összegzéséből jön létre, így ez a kérdőív másodrendű faktorstruktúrát alkalmaz. Végül figyeljük meg a TCI-személyiségkérdőívet (Cloninger, 1994), mely négy temperamentum- és három karakterdimenziót különít el, és minden skálához három-öt alskála tartozik alskálánként öt-tíz itemmel. Elsőre úgy tűnhet, ez egy harmadrendű struktúra, de a négy temperamentumskála nem összegezhető egy általános



11. ábra: Másodrendű CFA-modellek

De hogyan vizsgálható, hogy ez a modell helytáll-e a valóságban? Ehhez nézzünk egy jóval egyszerűbb felépítésű modellt részletesebb bemutatással (12. ábra A panelje)! A modellben egyetlen látens változó (X) tölt három manifeszt változóra (A, B és C). A töltések jele a λ (*lambda*). A látens változó varianciáját is megjelöltük a modellben, erre utal az X-ből önmagába visszamutató kétirányú nyíl, melynek jelölése VAR(X). A CFA leggyakoribb beállításának megfelelően a látens változó standardizált formáját használjuk, azaz átlaga nulla, szórása egy, és ennek megfelelően varianciája is egy, azaz VAR(X) = 1. Ezzel szemben a manifeszt változók varianciája nem jelenik meg közvetlenül a modellben, mert azt kettébontjuk: lesz egy rész, melyet a látens változó megmagyaráz (az előbbi példánál maradva, az emberek idői rendezettsége részben meghatározza, mennyire képesek követni egy fix napirendet). Lesz azonban a manifeszt változók varianciájának egy olyan része is, mely nem a látens változótól függ, hanem egyéb tényezőktől és a mintavételezési zajtól, például a személy idői rendezettségétől független, ha a munkája jellege teszi lehetetlenné a fix napirend követését. Ez a *meg nem magyarázott variancia*, más néven a *reziduális*, és az ábra nagy R-rel jelöli. Látható, hogy ezek a reziduálisok is oválisban helyezkednek el, mivel ezek is becsült értékek, és az adott tételre ható összes egyéb befolyásoló tényezőt

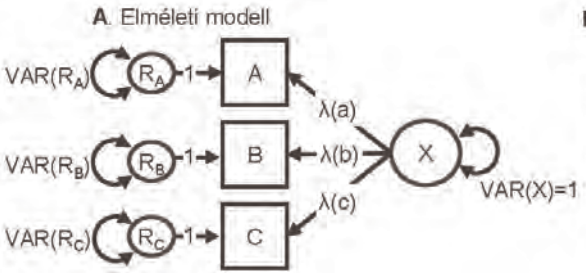
temperamentum-pontszámba, és hasonlóképp a három karakterskála sem, így a kérdőív tényleges faktorszerkezete valójában csak másodrendű.

jelképezik. A reziduálisok változatosságát VAR(R)-rel jelöltük. Egy item teljes változatossága mindig szétbomlik a látens változó által megmagyarázott és meg nem magyarázott részre.

A kérdőívvel csak a manifeszt változók mérhetők, a látens változóhoz és a reziduálisokhoz nem lesz oszlop az adatbázisban. De mi tudható A, B és C változóról? A varianciájuk és a közöttük lévő összefüggés, azaz a közöttük lévő kovariancia. Ez három változó esetében hat *független információ*: VAR(A), VAR(B), VAR(C), COV(A,B), COV(A,C) és COV(B,C). Az elméleti modellnek ezeket kell megmagyaráznia. A mért varianciákat és kovarianciákat a *megfigyelt (mért) variancia-kovariancia mátrix*ban (observed variance-covariance matrix) szokás megjeleníteni (12. ábra B panelje).

Hogyan magyarázható a modellel az A és B item közötti összefüggés? Vizsgáljuk meg a 12. ábra A panelje segítségével! Mivel nem fut közöttük közvetlenül sem egyirányú, sem kétirányú nyíl, így a modellben csak az magyarázza az összefüggésüket, hogy mindkettőben kifejeződik a látens változó. Ha végigkövetjük az egyik változóból másikba futó útvonalat, akkor azt kapjuk, hogy COV(A,B)=λ(a)*VAR(X)*λ(b). Ugyanígy felírhatók a COV(A,C) és a COV(B,C) egyenletek is. Az útvonalak felírásának kiterjedt szakirodalma van, amit jelen könyvben nem tárgyalunk, de az angol „*tracing rules*" kifejezéssel kereshetünk rá.

Hogyan írható fel az A változó varianciája a modell alapján? Mint korábban már említettük, a variancia egy változó önmagával való kovarianciája, azaz az útvonalnak a kiindulási változóba kell visszatérnie. Például ha A varianciáját számoljuk, akkor kiindulunk A-ból, áthaladunk λ(a)-n, „megfordulunk" a látens változó varianciáján, azaz VAR(X)-en, visszahaladunk λ(a)-n, és megérkezünk vissza A-ba. De ez még csak az A varianciájának látens változó által megmagyarázott része. A meg nem magyarázott részhez elindulunk az 1-esen, megfordulunk az A változó reziduális varianciáján, azaz VAR(R_A)-n, áthaladunk újra az 1-esen, és visszaérkezünk A-ba. Azaz VAR(A) = COV(A,A) = λ(a)*VAR(X)*λ(a) + 1*VAR(R_A)*1. Egyszerűsítve az összevonható tagokkal: VAR(A) = $\lambda(a)^2$*VAR(X) + VAR(R_A).



B. Mért variancia-kovariancia mátrix

	A	B	C
A	VAR(A)		
B	COV(A,B)	VAR(B)	
C	COV(A,C)	COV(B,C)	VAR(C)

12. ábra: A CFA elméleti modellje és mért variancia-kovariancia mátrixa

Mindezek alapján felírható hat egyenlet:
$COV(A,B) = \lambda(a)*VAR(X)*\lambda(b)$
$COV(A,C) = \lambda(a)*VAR(X)*\lambda(c)$
$COV(B,C) = \lambda(b)*VAR(X)*\lambda(c)$
$VAR(A) = \lambda(a)^2*VAR(X) + VAR(R_A)$
$VAR(B) = \lambda(b)^2*VAR(X) + VAR(R_B)$
$VAR(C) = \lambda(c)^2*VAR(X) + VAR(R_C)$

2.2. A modell azonosíthatósága (model identification)

Az előbbi egyenletrendszerből ismert értékek az itemek varianciái és kovarianciái (illetve a látens változó varianciáját mi adtuk meg egynek), és ismeretlen a három töltés és a három reziduális variancia. Mivel hat egyenlet van hat ismeretlennel, ez egy egyszerű többismeretlenes egyenletrendszer, melyet középiskolai szintű matematikatudással is könnyen meg lehet oldani. Ez egy *szaturált modell*, azaz minden lehetséges útvonalat behúzunk a modellben. Úgy is mondjuk, hogy a modell azonosítása szempontjából ez egy *„éppen azonosított modell"* (just-identified), ami annyit jelent, hogy pontosan annyi független információ van benne, mint amennyi becsülendő paraméter, ezért az egyenletrendszer egyértelműen megoldható. Ennek megfelelően egy háromitemes skála illeszkedését felesleges vizsgálni, mivel az mindig tökéletesen illeszkedik.[4]

[4] Bár e kontextusban a két fogalom egybeesik, Kline (2023) szerint érdemes megkülönböztetni őket: a szaturált modell a strukturális telítettséget (minden lehetséges kapcsolat meglétét) jelöli, míg a just-identified státusz a matematikai azonosíthatóságra (a becsült paraméterek és az elérhető információk egyezéséből fakadó nulla szabadságfokra) vonatkozik.

A független információk számának és a becsült paraméterek számának különbségét *szabadságfok*nak (degrees of freedom) nevezzük. A szaturált modell szabadságfoka nulla (df = 0), azaz pont annyi független információt tartalmaz, amennyi paramétert becsül.

Térjünk el ettől a szaturált modelltől először abba az irányba, amikor több a paraméter, mint a független információ. Erre lenne jó példa egy olyan skála, melynek csak két tétele van. Ekkor a szabad információk száma három: VAR(A), VAR(B) és COV(A,B), a becsült paraméterek száma azonban négy: λ(a), λ(b), $VAR(R_A)$ és $VAR(R_B)$. Ekkor a szabadságfok negatív lesz (itt df = -1). Ha felírjuk a három egyenletet, mely négy paramétert tartalmaz, akkor láthatjuk, hogy végtelen megoldása van az egyenletrendszernek. (Az egyszerűség kedvéért a varianciákat egy pillanatra félretéve, tegyük fel, hogy COV(A,B) = 10, ekkor a megoldása az egyenletnek a λ(a) = 10 és λ(b) = 1, vagy a λ(a) = 5 és λ(b) = 2, vagy a λ(a) = -20 és λ(b) = -0,5 is.) Az ilyen modellt *alulazonosítottnak* (underidentified) nevezzük. Egy olyan kérdőíven, melynek valamely skálája csak két tételből áll, problémába ütközünk a modell futtatása során, és a hagyományos CFA-hoz képest általában csak további megkötések mellett futtatható a modell.

Egy másik gyakori oka annak, ha alulazonosított modellt kaptunk, hogy ugyan van elég item a skálán, de a töltések mellé túl sok egyéb összefüggést is behúzunk, például túl sok reziduális kovarianciát engedünk meg (erről a hibáról később részletesen szó esik majd).

Térjünk át a másik oldalra, amikor több független információnk van, mint becsült paraméterünk. Erre egy példa, hogy ha a skála négy tételből áll, ekkor a független információk száma tíz, melyek a COV(A,B), COV(A,C), COV(A,D), COV(B,C), COV(B,D), COV(C,D), valamint az A, B, C és D változók varianciája. A becsült paraméterek száma azonban csak nyolc: a négy töltésérték és a négy reziduális variancia. A szabadságfok pozitív, df = 2. Ha felírjuk egyenletrendszerként, akkor tíz egyenletet kapunk, de csak nyolc ismeretlennel. Egy ilyen egyenletrendszernek nincs megoldása, ennek megfelelően a modell *túlazonosított* (overidentified).

2.3. Paraméterbecslés (model estimation)

Első olvasatra ez rossz hírnek tűnhet, de egyáltalán nem az, sőt itt kezdődik a CFA modelltesztelése. Példánkban az átláthatóság kedvéért leegyszerűsítve csak a kovarianciákat nézzük, és a varianciák egyenleteitől eltekintünk. A kovarianciák értéke a példa kedvéért legyen a fenti sorrendnek megfelelően 3, 4, 2, 5, 1, 4. Ezzel hozzuk létre a 13. ábra A paneljének megfigyelt

variancia-kovariancia mátrixát. Ekkor hat egyenlet írhatunk fel (továbbra is VAR(X) = 1 miatt kihagyható):
3 = COV(A,B) = λ(a)*λ(b)
4 = COV(A,C) = λ(a)*λ(c)
2 = COV(A,D) = λ(a)*λ(d)
5 = COV(B,C) = λ(b)*λ(c)
1 = COV(B,D) = λ(b)*λ(d)
4 = COV(C,D) = λ(c)*λ(d)

Bárhogy veszünk fel töltésértékeket, nem tudjuk a megfigyelt kovarianciamátrixot elérni. Például ha felvesszük az A változó töltését λ(a) = 1-re, a B változóét λ(b) = 3-ra, akkor a becsült COV(A,B) = 1*3 lenne, ami megegyezik a két változó között mért kovarianciával. Ekkor λ(c)-nek felvehetünk 4-et, hogy a COV(A,C) = 1*4 teljesüljön, és λ(d)-nek felvehetünk 2-t, hogy a COV(A,D) = 1*2 teljesüljön. Eddig úgy tűnik, hogy a töltések jól magyarázzák a mért kovarianciákat, azaz a *modell által implikált variancia-kovariancia mátrix* (13. ábra B panel) jól követi a megfigyelt variancia-kovariancia mátrixot (13. ábra A panelje).

Azonban ezek után COV(B,C) értéke 3*4=12 lesz a mért 5 helyett, azaz itt lesz 7 egység eltérés a mért értékektől. COV(B,D) értéke 3*2=6 lesz a mért 1 helyett, itt további 5 egység eltérés van, és COV(C,D) értéke 4*2=8 lesz a mért 4 helyett, ami további 4 egység tévedés. Látjuk, hogy a megfigyelt variancia-kovariancia mátrix eltér a modell által feltételezett variancia-kovariancia mátrixtól. Összesen ez a töltéskiosztás 0+0+0+7+5+4=16 egység tévedéssel járna.

Ha helyette a töltéseket úgy vesszük fel, hogy azok λ(a) = 1, λ(b) = 2, λ(c) = 3, λ(d) = 2 legyenek, akkor (ha végigvezetjük az előző gondolatmenetet), az össztévedés 0+2+0+1+3+2=8 lesz. Azaz ez a második töltéskiosztás a mért értékektől kisebb eltérést eredményez (13. ábra C és A paneljének eltérése). Így haladva, a paraméterek értékeit állítgatva el lehet jutni egy olyan paraméterérték-beállításhoz, amelynél már nem hozható ki jobb az aktuális modellt használva. Ha tovább akarjuk közelíteni a feltételezett modellt a megfigyelthez, már a modell szerkezetébe kell belenyúlni például változók ki- és bevételével (lásd később).

A. Mért var-kov mátrix

	A	B	C	D
A	VAR(A)			
B	3	VAR(B)		
C	4	5	VAR(C)	
D	2	1	4	VAR(D)

B. Modell által feltételezett var-kov mátrix, ahol λ(A) = 1; λ(B) = 3; λ(C) = 4; λ(D) = 2.

	A	B	C	D
A	VAR(A)			
B	3	VAR(B)		
C	4	12	VAR(C)	
D	2	6	8	VAR(D)

C. Modell által feltételezett var-kov mátrix, ahol λ(A) = 1; λ(B) = 2; λ(C) = 3; λ(D) = 2.

	A	B	C	D
A	VAR(A)			
B	3	VAR(B)		
C	2	6	VAR(C)	
D	2	4	6	VAR(D)

13. ábra: Példa mért és modell által feltételezett variancia-kovariancia mátrixra

Az előbb bemutatott paraméterérték-állítgatást természetesen nem nekünk kell kézzel elvégezni, a CFA paraméterbecslő eljárásai elvégzik ezeket a közelítéseket. A *paraméterbecslés* (parameter estimation) a CFA és SEM egyik központi lépése, amelynek során a modell paramétereit (töltések, varianciák, kovarianciák, reziduálisok stb.) egy algoritmus úgy határozza meg, hogy a modell által implikált variancia-kovariancia mátrix a lehető legjobban illeszkedjen a megfigyelt variancia-kovariancia mátrixhoz. A cél minden becslési módszernél ugyanaz, minimalizálni az eltérést a két mátrix között. A különböző eljárások abban térnek el egymástól, hogy milyen függvényt minimalizálnak, illetve milyen feltételezéseket tesznek az adatok eloszlására vonatkozóan.

A legelterjedtebb becslési eljárás a *Maximum Likelihood* (ML), amely feltételezi, hogy az adatok többváltozós normális eloszlásúak. Az ML-becslés előnye, hogy aszimptotikusan torzítatlan és hatékony, valamint elérhető minden főbb szoftverben. További előnye, hogy van olyan altípusa, amely jól kezeli a hiányzó értékeket (Full Information Maximum Likelihood, FIML), és amely az összes rendelkezésre álló információt felhasználja a hiányzó értékek kezelésére anélkül, hogy listwise vagy pairwise törlést kellene alkalmazni. A FIML ma a hiányzó adatok kezelésének egyik legajánlottabb módja SEM-ben.

Normalitási feltételek sérülésekor alternatív eljárások ajánlottak. A *Diagonally Weighted Least Squares* (DWLS) például kifejezetten nem normális eloszláshoz

és ordinális (Likert-típusú) itemekhez fejlesztett becslés. A DWLS kovarianciák helyett polikorikus korrelációkat használ, és csak a súlymátrix főátlóját veszi figyelembe, így kevésbé érzékeny a normalitás sérülésére és a kisebb mintaméretekre. Emellett olyan Likert-skáláknál, melyeknek maximum négy szintje van, pontosabb, jobb becslést ad, mint az ML, de ötfokú Likert-típusú skáláktól felfelé az ML is alkalmazható. Egy nagyon hasonló, gyakran használt eljárás a WLSMV (Weighted Least Squares Mean and Variance adjusted), amely a DWLS továbbfejlesztett változata. Robusztus a nem normális eloszlással szemben, és a becsült standard hibákat, valamint a khi-négyzet-statisztikát is korrigálja. Ez a megoldás vált szinte a leggyakoribbá a CFA becslő eljárásaként kérdőíves kutatásokban, ahol az ordinális skálák használata a leggyakoribb.

Bár e három paraméterbecslő eljárás a legelterjedtebb, ezeken kívül is léteznek egyéb eljárások. Speciális esetekben érdemes utánajárni más potenciális becslőknek is. Fontos kiemelni, hogy ugyanazt az eljárást különböző szoftverek más-más néven kínálják. Emiatt minden esetben érdemes ellenőrizni, hogy a választott program pontosan melyik algoritmust futtatja az adott parancs megadásakor.

Összességében tehát a paraméterbecslés célja minden esetben az, hogy a modell által feltételezett szerkezet minél inkább közelítsen a valós adatokhoz. A becslő kiválasztásához a kutatónak mérlegelnie kell, hogy milyen típusú adatokkal dolgozik (folytonos vagy ordinális), milyen a mintanagyság, illetve menynyire tartható a normalitás feltételezése. Ezek határozzák meg, hogy az ML, a DWLS vagy valamelyik más paraméterbecslő eljárás a legmegfelelőbb.

2.4. A modell értékelése (model evaluation)

Az előző fejezetben láttuk, hogy a paraméterbecslő eljárások megtalálják azokat a paraméterértékeket, amelyek mellett a modell által feltételezett variancia-kovariancia mátrix a lehető legközelebb kerül a megfigyelt variancia-kovariancia mátrixhoz. Ez azonban még nem jelenti azt, hogy a modell megfelelően illeszkedik az adatokhoz, csak azt, hogy az adott modell esetében ez a legjobb illeszkedés, ami elérhető. A következő lépés ezért a modell illeszkedésének vizsgálata különböző *illeszkedési mutatók* segítségével.[5]

[5] Az illeszkedési mutatók elvárt határértékei az évtizedek során jelentős változáson mentek keresztül – a korábbi, megengedőbb nézeteket felváltó, Hu és Bentler (1999) által javasolt szigorúbb küszöbértékeket az újabb szimulációs kutatások (például Marsh et al., 2004) ma már a modellbonyolultságtól és mintaelemszámtól függő rugalmassággal kezelik –, ezért bár kötetünkben a jelenleg irányadó

Az illeszkedés alapja a két variancia-kovariancia mátrix (a feltételezett és a megfigyelt) eltérése. Az eltérés mérőszáma a *khi-négyzet érték* (χ^2), amelyet úgy számítanak, hogy összevetik a megfigyelt és a modell által becsült kovarianciamátrixot. Ha a modell tökéletesen illeszkedne, a különbség nulla lenne, és a χ^2 is 0 lenne. Minél nagyobb a χ^2, annál nagyobb az összeltérés, tehát annál rosszabb az illeszkedés. (De vigyázat, egy komplexebb, több paramétert alkalmazó modellnek is lehet nagy a χ^2 értéke a sok kis eltérés kumulálódása miatt.)

Az eltérés statisztikai vizsgálata a *khi-négyzet-próba*, amely azt teszteli, hogy a két mátrix közötti különbség statisztikailag nulla-e. Ha a χ^2-próba nem szignifikáns, az azt jelenti, hogy a modell kellően pontosan illeszkedik az adatokhoz, a feltételezett értékek nem térnek el jelentősen a megfigyeltektől. Ugyanakkor a mutató erősen érzékeny az elemszámra, nagy minták esetén (akár 200-300 főtől) a próba szinte mindig szignifikáns lesz, így a modell túl szigorúan kerül elutasításra. Ezért szokás a χ^2 értéket a szabadságfokkal elosztani (χ^2/df). A χ^2/df arányánál általános irányelv szerint a 3 alatti érték elfogadható, a 2 alatti pedig jó illeszkedést jelez.

Mivel a χ^2 önmagában nem megbízható, elterjedtek a *komparatív illeszkedési mutatók*, amelyek a vizsgált modellt egy úgynevezett baseline modellel/nullmodellel vetik össze. A nullmodell általában az, ami ugyanazokat a változókat tartalmazza, mint a vizsgált modell, de minden változót (vagy majdnem minden változót) függetlennek tekint, így alapvetően a leggyengébb illeszkedéssel rendelkezik. Az ide tartozó legelterjedtebb mutatók a *CFI* (Comparative Fit Index), a *TLI* (Tucker–Lewis Index) és az *IFI* (Incremental Fit Index). A három mutató általában rendkívül hasonló eredményt mutat, de talán a legszigorúbb ítéletet a TLI hozza, mivel ez a számítás a modell parszimóniáját jobban figyelembe veszi, így a modellben lévő szabadságfokok csökkenését jobban bünteti. A mutatók értéke 0 és 1 között mozog, ahol a magasabb érték jobb illeszkedést jelez. Számításukból adódóan néha előfordul 0-nál kisebb vagy 1-nél nagyobb érték is, ekkor 0-ra és 1-re szokás kerekíteni. Általában 0,90 felett elfogadható, 0,95 felett pedig jó illeszkedésről beszélünk. Hátrányuk, hogy a nullmodell definíciója programonként eltérhet, ezért nem minden szoftver ad pontosan azonos eredményt. További probléma, hogy ezekhez a mutatókhoz nem kapcsolódik konfidenciaintervallum.

Széles körben használt mutató az *RMSEA* (Root Mean Square Error of Approximation). Az RMSEA *abszolút illeszkedésmutató*, mert önmagában méri, mennyire felel meg a modell az adatoknak, nem pedig egy másik (például null-)

tartományokat ismertetjük, a kutatóknak javasoljuk a mindenkori legfrissebb, kontextusfüggő módszertani ajánlások nyomon követését.

modellhez hasonlítja. Azt mutatja, hogy egy szabadságfokra vetítve mekkora átlagos eltérés maradt a modell és a valós adatok között, azaz a modell összetettségéhez képest mennyi a tévedés. Ennek megfelelően a kisebb RMSEA-érték jobb illeszkedést jelez, hiszen minél kisebb az RMSEA, annál kevesebb hiba marad a modellben. A 0 azt jelenti, hogy a modell tökéletesen illeszkedik (nincs maradék hiba). A 0,08 alatti érték jó, a 0,05 alatti érték kiváló illeszkedést jelez. Az RMSEA értékének nincs felső határa. Az RMSEA előnye, hogy gyakran megadják hozzá a 90%-os konfidenciaintervallumot, amelynek felső határa (upper level confidence interval, ULCI) szintén fontos, ha 0,10 alatt van, az illeszkedés jónak tekinthető.

Az *SRMR* (Standardized Root Mean Square Residual) a változók között a megfigyelt és a becsült korrelációk átlagos különbségeit mutatja (pontosabban a korrelációk közötti különbségeket négyzetre emeli, a négyzetes eltéréseket átlagolja, majd az átlagból gyököt von). A kisebb érték a becsült korrelációk kisebb eltérésére utal, ezáltal jobb illeszkedést jelez. Értéke 0,08 alatt elfogadhatónak tekinthető, 0,05 alatt pedig jónak.

Egyes kutatók említést tesznek még a GFI (Goodness-of-Fit Index) és az AGFI (Adjusted Goodness-of-Fit Index) mutatókról is, de ezek mára kevésbé ajánlottak, mivel erősen mintanagyságfüggők. Ezenfelül a módszertan folytonos fejlődésével a fent felsoroltak mellett rendre jelennek meg új megoldások az illeszkedés mérésére. Összefoglaló táblázatot az Útvonalelemzés fejezetben talál az olvasó.

Gyakran előfordul, hogy a különböző mutatók nem teljesen azonos következtetéshez vezetnek. Ez nem feltétlenül probléma, mivel mindegyik más szempontból értékeli a modellt. Az illeszkedésről mindig holisztikus döntést kell hozni, vagyis a modellt az összes mutató figyelembevételével, a kutatási kérdés és a minta sajátosságainak ismeretében kell értékelni.

2.5. A modell újraspecifikálása (model respecification)

Ha a CFA illeszkedési mutatói nem kielégítőek, a következő lépés a *modell újraspecifikálása*. A rossz illeszkedést több tényező is okozhatja. Például ha a kutató elképzelésével ellentétben az itemek mögött nem áll semmilyen egyértelmű látens változó, az az itemek közötti általános gyenge korrelációban és alacsony illeszkedésben mutatkozik meg. Az ilyen eset ritka, de előfordulhat, hogy a modell akár egyáltalán nem javítható.

Nem példa nélküli az sem, hogy az itemek nem a kutató által feltételezett struktúra mentén csoportosulnak (például étkezési preferenciákat vizsgáló kérdőívnél a kutató azt gondolja, hogy az emberek összetétel alapján döntenek a különböző termékek fogyasztásáról, és a kérdőívét ez alapján állítja össze, de a faktorstruktúra ellenőrzésénél szembesülnie kell azzal, hogy valójában ár és íz alapján történik a döntés). Ilyenkor feltáró faktorelemzéssel (EFA) felderíthető az itemek mögött ténylegesen álló faktorstruktúra, de egy második mintán ezt is vissza kell ellenőrizni CFA-val.

Az is lehetséges, hogy bár a faktorstruktúra a legtöbb tételre megfelelően áll, de van néhány egyik faktorra sem jól illeszkedő tétel a kérdőívben.

A leggyakoribb mégis az, amikor egy-két tétel túlzottan, a modell által meg nem magyarázható erősséggel korrelál más tételekkel. A túl erős korrelációkból adódó rossz illeszkedés feltárására gyakran használjuk a *modifikációs indexe*ket (MI), amelyek azt jelzik, mekkora lenne a χ^2 csökkenése, ha egy-egy plusz útvonalat engedélyeznénk a modellben. A MI-k gyakorlatilag végigiterálnak minden lehetséges módosításon, ezért a megfontolandó tippek mellett sokszor értelmetlen vagy szakmailag nem megalapozott javaslatokat is adhatnak.

A CFA-ban a modifikációs javaslatok általában két típusba sorolhatók:

Keresztt öltés (cross-loading) akkor jelentkezik, amikor egy item több faktorra is jelentősen tölt, tehát nemcsak a saját fő faktorára, hanem másik faktor(ok)ra is van kapcsolat. Erre lehet példa az 1. ábrában látott kérdőívnél a „Házimunkára mindig szánok időt", amely egyrészről téri rendezettséget eredményez, de mellette idői tervezést is igényel. A CFA-t legtöbbször kérdőívek validálásánál vagy SEM-modellek részeként alkalmazzuk. Mindkét esetben fontos, hogy a különböző látens változók között ne legyenek értelmezésbeli átfedések, ennek megfelelően a keresztt öltések miatt kialakuló rossz illeszkedés megszüntetésének megoldásaként szinte sohasem a keresztt öltés megengedését, hanem a tétel törlését választjuk.

A másik módosítási lehetőség a *reziduális kovariancia* (residual covariance) megengedése, amellyel azt feltételezzük, hogy a két item hibája korrelál, mert a mögöttük álló belső tulajdonságon túl is van valami, ami összeköti őket (például legyenek egy mozgásminőség-kérdőívben szereplő tételek többek között a guggolás, ugrás, lépcsőn lemenetel és lépcsőn felmenetel, ebben az esetben a lépcsőn való fel- és lemenetel között vélhetően kialakul és szakmailag megengedhető a reziduális kovariancia). Gyakran reziduális kovarianciához vezet két item hasonló szóhasználata, mert a hasonló megfogalmazás miatt az emberek a kelleténél hasonlóbb választ adnak a két tételre, erre példa a 10. ábra kérdőívének két „Nehezemre esik..." kezdetű tétele. Gyakran látható jelenség az egyenes és fordított tételek csoportosulása is. Ha azonos skálán belül van a két

tétel, melyek között valamilyen okból erősebb összefüggés lenne szükséges, és az összefüggés szakmai szempontok alapján érthető, akkor a reziduális kovariancia megengedhető. Azonban ha eltérő skálákon tölt a két tétel (mint ahogy azt az előző példában is láttuk), akkor a reziduális kovariancia nem engedhető meg. Jelentős illeszkedésbeli probléma esetében a kettő közül valamelyik tételt törölni szokás.

A skálavalidálás során a keresztt öltéseket és skálákon átívelő reziduális kovarianciákat általában itemtörléssel oldjuk meg, és a skálákon belüli reziduális kovarianciákat is csak akkor engedjük meg, ha az szakmailag indokolt. A modifikációs indexek segítségével addig javítgatunk, amíg a változtatás jelentős javulást eredményez az illeszkedésben, de nem szabad minden javaslatot mechanikusan elfogadni. Egy lépésben mindig csak egy változást eszközölünk, mert egyetlen változtatás megoldhat több illeszkedési problémát. Gyakori példa az, ha két item között nagyon erős az összefüggés, akkor a becslő eljárás úgy osztja ki a töltéseket, hogy ezt valami módon magyarázza. Ez eltolja a látens változó valós jelentéstartományát, ami viszont gyakran a többi item alacsonyabb töltését okozza, és összességében alacsony illeszkedést eredményez. Ilyenkor a két tétel közötti reziduális kovariancia megengedése láncolatszerűen oldja meg az összes problémát, és gyakran minden más is a helyére ugrik.

Ha az MI-k már nem nyújtanak további segítséget, érdemes áttekinteni az itemtöltéseket, különösen a gyengén töltő itemeket, amelyek gyenge kapcsolatot mutatnak a faktorral, és esetleg törlésre szorulnak.

További szempont lehet a minta mérete és a faktorok számának aránya, illetve a pszichometriai és elméleti szempontok figyelembevétele. Minden módosításnak szakmailag is megalapozottnak kell lennie, különben a modell *túlilleszt tett* (overfit) lesz, és az illeszkedés javulása csak matematikai trükk eredménye lesz, nem valódi mérési javulás.

Ha minden szakmai szempontból megengedhető lehetőséget kipróbáltunk, és a modell még mindig nem mutat megfelelő illeszkedést, érdemes elgondolkodni más faktorstruktúra kialakításán. Ez történhet elméletvezérelten, például ha van olyan elmélet, amely az itemek más struktúráját feltételezi, de történhet adatvezérelten is, az itemek közötti összefüggésekből kiindulva. Ilyenkor a mintát véletlenszerűen két részre kell bontani, az egyik félen feltáró faktorelemzést, a másik félen az új struktúrának megfelelően CFA-t futtatva.

2.6. A modell értelmezése (model interpretation)

A megfelelő illeszkedés igazolása után a következő lépés a modell eredményeinek részletes értelmezése. Elsőként célszerű áttekinteni az egyes itemek faktortöltéseit. Mindegyiknek szignifikánsnak kell lennie, megfelelő irányban kell állnia, és erősségüknek elég nagynak kell lennie ahhoz, hogy elfogadjuk, az adott tétel valóban a mögötte feltételezett látens konstrukciót reprezentálja. Ha egy item gyenge vagy nem szignifikáns töltést mutat, akkor megfontolható annak elhagyása vagy a modell módosítása. Ezt követően érdemes megvizsgálni az egyes tételek magyarázott varianciáját (R^2), valamint a reziduálisuk értékét, mely utóbbi a tételekben megfigyelhető változatosságnak azon részét reprezentálja, amelyet a látens változó nem képes magyarázni. A kirívóan alacsony R^2 érték és az ezzel szükségszerűen együtt járó magas reziduális variancia a mérési modell gyengeségére utal, mert azt jelzi, hogy az adott állítást döntő részben a mérési hiba vagy a modellben nem specifikált egyéb hatások (zaj) vezérlik, így az csak elhanyagolható mértékben tükrözi a vizsgált konstruktumot. Végül a látens változók közötti korrelációk és kovarianciák értelmezése is fontos, mert ezekből derül ki, mennyire különülnek el vagy mennyire kapcsolódnak össze a vizsgált konstrukciók, ami elengedhetetlen a modell elméleti hátterének értékeléséhez.

2.7. További elemzések

A CFA elvégzését követően a további elemzések szükségessége attól függ, milyen kutatási céllal készült a mérési modell. Amennyiben egy kérdőív validálásának részeként végezzük a CFA-t, a *strukturális validitás* igazolása után további lépések is indokoltak. Ellenőrizni kell a faktorstruktúra *konvergens validitását* (például az AVE-mutató alapján), valamint a *diszkriminánsvaliditást* (például a HTMT-mutatóval). Ezek mellett a *megbízhatóság* is kiemelt fontosságú, amelyet a Cronbach-alfa és a McDonald-omega értékek segítségével mérhetünk. A validálási folyamat szerves része továbbá a *kritériumvaliditás* vagy más néven a konstruktum konvergens validitásának vizsgálata, amely más, elméletileg kapcsolódó kérdőívekkel való korrelációk elemzésén alapul. (Könnyen félreértéshez vezethet, hogy a konvergens validitás kifejezést egyrészt a modell belső, faktorokra vonatkozó strukturális mutatójára, másrészt a skála egészének egy külső mérőeszközzel való kapcsolatára is használjuk.)

Ha a CFA-t a strukturális egyenletmodellezés (SEM) részeként alkalmazzuk, a mérési modell illeszkedésének megerősítése után a következő lépés a *strukturális modell* elemzése, vagyis a látens változók közötti kapcsolatok vizsgálata.

Mindkét megközelítés esetében hasznos kiegészítés lehet a *mérési invariancia tesztelése*, amely azt vizsgálja, hogy a modell különböző csoportok (például nemek, életkori kategóriák) esetében is azonos módon működik-e. Ennek részletes tárgyalására a SEM-ről szóló fejezetben térünk vissza.

18. doboz: A megerősítő faktorelemzés (CFA) kulcsfogalmainak definíciói

I. Változók és statisztikai alapfogalmak

Látens változó (latent variable): olyan elméleti konstrukció, amelyet közvetlenül nem tudunk mérni (például szorongás, elégedettség), de több megfigyelt változó mögött húzódó közös tényezőként feltételezünk.

Manifeszt változó (observed/manifest variable): közvetlenül mérhető, megfigyelhető változó (például kérdőívtételre adott válasz, tesztpontszám), amely a látens változó jeleként, indikátoraként szolgál.

Item, tétel, indikátor: a kérdőív vagy teszt egyetlen eleme (például egy kérdés, állítás vagy feladat), amelyre a válaszadó választ ad. Az item a manifeszt változó szerepét tölti be a CFA-ban, mivel közvetlenül mérhető, és a mögötte húzódó látens változó indikátoraként szolgál.

Kovariancia (covariance): két változó együtt járásának mértéke, mely megmutatja, hogy azok értékei hajlamosak-e együtt növekedni vagy csökkenni. Ábrán kétirányú nyíl jelöli.

Variancia (variance): egy változó értékeinek szóródását méri, a szórás négyzete, azt mutatja meg, mennyire térnek el az egyes megfigyelések az átlagtól. Ábrán önmagába visszatérő kétirányú nyíl reprezentálja.

Reziduális (residual): az item értékének azon része, amelyet a látens változó nem magyaráz meg (a mérési hiba és az egyedi sajátosságok együttese).

Reziduális variancia (residual variance): a reziduális szóródása, vagyis annak mértéke, hogy az adott item mekkora részben tartalmaz nem magyarázott, egyedi vagy véletlen hatásokat.

Reziduális kovariancia (residual covariance): két item közötti azon kapcsolat, amelyet a modellbe bevont faktorok nem magyaráznak meg, vagyis a hibatagjaik közötti összefüggés.

II. A modell szerkezete és azonosítása

Mérési modell (measurement model): a látens és a manifeszt változók közötti kapcsolat statisztikai leírása, amely megmutatja, hogy a mérhető itemek miként reprezentálják az elméleti konstrukciókat.

Elsőrendű (first-order) CFA: olyan modell, amelyben a manifeszt változókat egy vagy több látens változó (faktor) magyarázza, de a látens változók nem alkotnak magasabb rendű főskálát.

Másodrendű (second-order) CFA: olyan modell, amelyben az elsőrendű faktorokat (látens változókat) egy magasabb szintű, másodrendű látens faktor magyarázza. Ilyenkor az itemekre először az elsőrendű faktorok töltenek, ezekre pedig egy átfogóbb, másodrendű konstrukció tölt.

Töltés (faktorterhelés, loading): a látens változó és az item közötti kapcsolat erősségét és irányát kifejező paraméter a CFA-ban. Azt mutatja meg, hogy az adott item mennyire jól reprezentálja a mögötte álló faktort, minél nagyobb az abszolút értéke, annál erősebben tükrözi az item a látens konstrukciót. Ábrán egyirányú nyíl reprezentálja, mely a látens változóból mutat a manifesztbe. Matematikai jele a λ (lambda).

Független információ (independent piece of information): a kovariancia-mátrixban található olyan elemek száma, amelyek egymástól független információt hordoznak az adatok szerkezetéről. A varianciák és kovarianciák száma.

Szabadságfok (degrees of freedom): a CFA-ban a független információk számának és a modellben becsülendő szabad paraméterek számának a különbsége. Ez mutatja meg, hogy mennyi „mozgástér" marad a modell illeszkedésének vizsgálatára. Ha a szabadságfok 0, akkor szaturált modellről beszélünk, ha pozitív, túlazonosított modellről (itt az illeszkedés tesztelhető), ha negatív, akkor alulazonosított modellről (mely nem becsülhető egyértelműen).

Szaturált modell (saturated model): olyan statisztikai modell, amelyben annyi paramétert becsülünk, ahány független információ van az adatokban. Emiatt a modell tökéletes illeszkedést mutat (nincs hiba), de szabadságfoka sincs, így a modell illeszkedése nem tesztelhető.

Éppen azonosított modell (vagy pontosan azonosított modell, just identified model): a szaturált modell másik elnevezése, annyi szabad paraméter van, ahány független információ. A modell illeszkedése matematikailag szükségszerűen (triviálisan) tökéletes, így szabadságfok hiányában az érvényessége nem tesztelhető.

Alulazonosított modell (underidentified model): több a szabad paraméter, mint a rendelkezésre álló információ, ezért a modell nem becsülhető meg egyértelműen, CFA nem futtatható.

Túlazonosított modell (overidentified model): kevesebb a szabad paraméter, mint a rendelkezésre álló információ, így van szabadságfok, és a modell illeszkedése statisztikailag tesztelhető, ezen történik a hagyományos CFA.

III. Matematikai háttér és becslés

Útkövetési szabályok (tracing rules): olyan szabályrendszer, amelyeket a kovariancia-struktúra-modellezésben (például CFA, SEM) alkalmazunk annak levezetésére, hogy a látens és manifeszt változók közötti utak (útvonalak) mentén hogyan állíthatók elő a kovarianciák és varianciák. A nyomkövetési szabályok adják meg a szabályrendszert, amellyel a modell gráfjából a modell egyenletrendszere felírható.

Mért variancia-kovariancia mátrix (observed var-cov matrix): az adatokból számított mátrix, amely a manifeszt változók varianciáit és kovarianciáit tartalmazza.

Implikált variancia-kovariancia mátrix (implied var-cov matrix): a CFA-modell paraméterei (faktorterhelések, varianciák, kovarianciák) alapján előre jelzett variancia-kovariancia mátrix.

Paraméterbecslő eljárások CFA-ban (parameter estimators): olyan módszerek, amelyek a modell paramétereit úgy határozzák meg, hogy az implikált és a mért variancia-kovariancia mátrix a lehető legjobban közelítsen egymáshoz. A leggyakoribb eljárások a maximum likelihood (ML) és a diagonally weighted least squares (DWLS).

Maximum Likelihood (ML): a leggyakrabban használt paraméterbecslő eljárás. A paramétereket úgy becsüli, hogy azok mellett a megfigyelt adatok előfordulásának valószínűsége maximális legyen. Főként folytonos és normális eloszlású adatoknál ajánlott.

Diagonally Weighted Least Squares (DWLS) / Weighted Least Squares Mean and Vari-ance adjusted (WLSMV): ordinális (például Likert-típusú) adatokhoz fejlesztett robusztus paraméterbecslő eljárások, amelyek polikorikus korrelációkra épülnek, és kevésbé érzékenyek a normalitás sérülésére, valamint kisebb minták mellett is jól használhatók.

IV. A modell értékelése (illeszkedési mutatók)

Illeszkedési mutatók (fit indices): olyan statisztikai mutatók, amelyek azt jelzik, mennyire felel meg a feltételezett (modellből származó) variancia-kovariancia mátrix a megfigyelt adatok variancia-kovariancia mátrixának. Segítenek megítélni a modell általános érvényességét, mivel az egyes paraméterek szignifikanciája önmagában nem elegendő.

χ^2 (khi-négyzet): a megfigyelt és a modell által becsült kovarianciamátrix közötti eltérést méri. A kisebb érték jobb illeszkedésre utal. Optimális esetben a teszt nem szignifikáns, de a nagy elemszám miatt ez gyakran nem elérhető.

χ^2/df: a khi-négyzet és a szabadságfok hányadosa, a modell illeszkedésének eltéréseit a modellben lévő megkötések mértékéhez méri, a komplexebb és szigorúbb (több megkötést tartalmazó) modell esetében nagyobb összeltérést enged meg. A 3 alatti érték általában elfogadhatónak számít, míg a 2 alatti jónak.

CFI (Comparative Fit Index): komparatív illeszkedési mutató, a modell illeszkedését hasonlítja egy „nullmodellhez". Értéke 0-1 között értelmezhető. Mivel a teljesen megkötött, lehető legrosszabbul illeszkedő modellhez viszonyítja az aktuális modellt, így a nagyobb érték jobb illeszkedést jelent. Értéke 0,90 felett jó, 0,95 felett kiváló illeszkedést jelez.

TLI (Tucker–Lewis Index): komparatív illeszkedési mutató, az illeszkedésnél a modell bonyolultságát is figyelembe veszi. Értéke 0-1 között értelmezhető, a magasabb érték jobb illeszkedésre utal. Értéke 0,90 felett jó, 0,95 felett kiváló értéknek számít.

RMSEA (Root Mean Square Error of Approximation): illeszkedési mutató, mely azt méri, hogy egységnyi szabadságfokra vetítve mennyire „hibás" az illeszkedés. Ennek megfelelően a kisebb RMSEA-érték jobb illeszkedést jelez. A 0 azt jelenti, hogy a modell tökéletesen illeszkedik (nincs maradék hiba), 0,08 alatt jó, 0,05 alatt kiváló az illeszkedés. A hozzá számolt 90%-os konfidenciaintervallum felső határát (90% ULCI) 0,10 alatt tekintjük jónak.

SRMR (Standardized Root Mean Square Residual): a változók között a megfigyelt és a becsült korrelációk különbségeinek átlagát mutatja. A kisebb érték a becsült korrelációk kisebb eltérésére utal, ezáltal jobb illeszkedést jelez. Értéke 0,08 alatt elfogadható, 0,05 alatt jó illeszkedést jelez.

Modifikációs index (modification indices, MI): statisztikai mutató, amely azt jelzi, mennyivel javulna a modell illeszkedése (mennyivel csökkenne a χ^2 értéke), ha egy jelenleg nem engedélyezett paramétert (például keresztt öltést vagy reziduális kovarianciát) szabadon engednénk a modellben. Gyakorlatilag tehát azt mutatja, hol „feszül" a modell leginkább, és milyen módosítás oldaná ezt a feszültséget.

Keresztt öltés (cross-loading): amikor egy item nemcsak a saját (célszerűen hozzárendelt) faktoron mutat magas töltést, hanem egy másik faktoron is jelentős faktortöltéssel bír.

3. Kutatási folyamat

A CFA ritkán jelenik meg önálló elemzésként, leginkább két területen alkalmazzák. Az egyik a kérdőívek fejlesztése és validálása, amelyben a CFA egy viszonylag jól körülhatárolt folyamat részeként jelenik meg. Ez a folyamat rendszerint a kérdőív fordításával vagy adaptálásával indul, ezt követi a tartalmi validitás vizsgálata, majd a strukturális validitás ellenőrzése (itt kap szerepet a CFA), továbbá a konvergens (például AVE) és diszkriminációs validitás (például HTMT), a reliabilitás (például Cronbach-alfa vagy McDonald-omega), végül pedig a kritériumvaliditás (más skálákkal való korreláció) értékelése. A másik tipikus alkalmazási terület a strukturális egyenletmodellezés (SEM), ahol a CFA önállóan nem különül el, hanem a teljes modell részeként jelenik meg. Ebben az esetben a mérési modell a CFA-nak megfelelő részt képezi, amely a látens változók és

a mérőtételek kapcsolatát írja le, miközben a strukturális modell a látens konstrukciók közötti összefüggéseket vizsgálja.

3.1. Mintaválasztás

A CFA nem igényel különleges populációt a mintaválasztás során, az adatok gyűjtésének mindig abból a célpopulációból kell történnie, amelyre a kérdőívet adaptálni vagy validálni szeretnénk. A módszertan szempontjából nem a populáció összetétele a kritikus tényező, hanem sokkal inkább a mintanagyság. A szakirodalomban a szükséges mintanagyság meghatározására vonatkozóan nem léteznek abszolút érvényű kritériumok, csupán tapasztalati irányelvek (ökölszabályok) ismertek. Ilyen gyakran idézett ajánlás (például Kline, 2023) a paraméterek stabilitásához szükségesnek ítélt, tételenkénti legalább 10–20 fős arány, illetve a modellkonvergenciához javasolt minimum 200 fős minta. Ugyanakkor a modernebb, szimulációs vizsgálatok hangsúlyozzák, hogy a szükséges elemszám nagyban függ a modell komplexitásától és a változók típusától, így a megbízható és általánosítható becslésekhez a gyakorlatban gyakran a fenti minimumoknál lényegesen nagyobb, több száz (akár 500–1000) fős minta javasolt.

A mintaelemszám tervezésénél további szempont, hogy a gyakorlatban előfordulhat, hogy a kiindulásként feltételezett modell illeszkedése nem bizonyul megfelelőnek. Ilyenkor feltáró faktoranalízist (EFA) is szükséges lehet végezni, majd a kapott eredményeket újabb CFA-hoz felhasználni. Ehhez azonban a mintát célszerű kettéválasztani, az egyik részmintán végezhető az EFA, a másikon pedig az eredmények megerősítése CFA segítségével. Ez azt is jelenti, hogy mindkét almintában biztosítani kell a minimális elemszámot, vagyis a teljes mintának eredetileg is elegendően nagynak kell lennie ahhoz, hogy a két eljárás külön-külön is megbízható eredményeket szolgáltathasson.

3.2. Adatgyűjtés, az adatok előkészítése

Ugyan a CFA folytonos adatokon is alkalmazható, mégis a gyakorlatban leggyakrabban Likert-skálás kérdőívek validálására használják. Amennyiben egy már létező kérdőívet validálunk új nyelvre vagy új populációra, fontos szempont, hogy az eredeti kérdőívben szereplő skálakiosztást tartsuk meg, hiszen a mérőeszköz ekvivalenciáját ez biztosítja. Ha azonban saját fejlesztésű kérdőívet készítünk, mérlegelnünk kell, hogy milyen típusú Likert-skálát alkalmazunk,

mert a különböző skálaszintekhez eltérő paraméterbecslési módszerek kapcsolódnak. Például dichotóm tételek (igen/nem típusú válaszok) esetén a tetrakorikus (tetrachoric) korreláción alapuló eljárások, például a DWLS vagy a WLSMV becslő javasolt. A három- vagy négyszintes Likert-skálánál szintén ezek a legygyakrabban használt becslő eljárások. Öt vagy annál több szintnél (különösen, ha a mintaelemszám is nagy) az adatok kezelhetők kvázifolytonosként, így az ML- vagy a robusztus ML-becslések megfelelőek lehetnek. Valóban folytonos adatokon is az ML az elterjedt, ilyen például a teljesítménymutatók, a pszichofiziológiai adatok (például reakcióidő, pulzusszám), illetve különböző teszteredmények faktorstruktúrájának vizsgálata.

Az adatfelvételt követően az információkat hagyományos, táblázatos formában szükséges elrendezni, ahol az oszlopok az itemeket, a sorok a vizsgálati személyeket reprezentálják. A Likert-skálás tételek válaszszintjeit számszerűen kell kódolni (például 1–5 vagy 1–7), így közvetlenül elemezhetők a statisztikai programokban. Fontos megjegyezni, hogy a fordított tételeket nem szükséges előre visszakódolni, mert a CFA során ezek negatív faktortöltésként jelennek meg, ami pontosan tükrözi a mérőeszköz szerkezetét. Ez egyszerűsíti az adatelőkészítést, és lehetővé teszi, hogy a modell természetes módon számoljon az itemek irányultságának különbségeivel.

3.3. Az elemzés beállításai

A strukturális egyenletmodellezéshez több elterjedt szoftver érhető el, amelyek különböző előnyöket kínálnak, és amelyeken belül a CFA-t eltérő módon valósították meg. Az Mplus az egyik legprofesszionálisabb, licencdíjas kereskedelmi szoftver, mely elsődlegesen parancssoros alapú, de grafikus felület nélkül is rendkívül rugalmas és erős nagyobb adatelemzésekhez. Az AMOS (SPSS-hez kapcsolódva érhető el) szintén licencdíjas, és legfőbb erőssége a grafikus felület, amely a menüs vezérlést is ötvözi. A lavaan egy ingyenes R-csomag, amely szöveges szintaxisra épül, így nagy szabadságot ad, de grafikus felület nélküli. Ehhez felhasználóbarát, ingyenes menüs és grafikus környezetet nyújt a JASP és a Jamovi, amelyek lavaan alapokra építve lehetővé teszik a vizuális beállítást és eredménymegjelenítést. A könyv írója is ez utóbbiakat ajánlja. Emellett léteznek más, kevésbé elterjedt, de említésre méltó lehetőségek is, mint az EQS (licencdíjas, parancssoros és grafikus felülettel is), illetve a Lisrel (licencdíjas, inkább szöveges és technikai jellegű).

19. doboz: A SEM/CFA-elemzéshez használt legnépszerűbb szoftverek jellemzői

Mplus: Professzionális, kereskedelmi szoftver (magas licencdíjjal), amely elsősorban szintaxisalapú (szöveges parancsokkal vezérelhető). Rendkívüli rugalmassága miatt a legbonyolultabb modellek (pl. többszintű modellezés, kategorikus adatok) kezelésére is alkalmas. Hátránya a meredek tanulási görbe és a grafikus modellépítő felület hiánya.

IBM SPSS AMOS: Kereskedelmi, licencdíjas program, amelynek fő vonzereje a grafikus (rajzolós) felület és a menüvezérelt irányítás. Ideális választás kezdőknek és vizuális típusú kutatóknak, mivel a modell közvetlen lerajzolásával futtatható az elemzés. Korlátja, hogy a nagyon komplex modellek és finomhangolt szintaxisok beállítása nehézkesebb lehet benne.

R (lavaan csomag): Ingyenes, nyílt forráskódú, szintaxisalapú (programozást igénylő) megoldás. Legfőbb előnye a teljes reprodukálhatóság, a hatalmas nemzetközi közösségi támogatás és a folyamatosan frissülő módszertani eszköztár. Hátránya, hogy alapvető programozási ismereteket igényel, és nincs grafikus kezelőfelülete, így elsajátítása időigényesebb.

JASP / Jamovi: Ingyenes és nyílt forráskódú szoftverek, amelyek modern, menüvezérelt grafikus felhasználói felületet (GUI) kínálnak. Kiváló hidat képeznek a kezdők és haladók között: rendkívül felhasználóbarátok, miközben a háttérben a professzionális R (lavaan) motor dolgozik. Korlátjukat a legextrémebb, egyedi szintaxist igénylő speciális beállítások jelenthetik.

EQS / LISREL: Történeti jelentőségű, stabil kereskedelmi szoftverek, amelyek vegyes (szintaktikai és grafikus) felülettel rendelkeznek. Napjainkban már kevésbé elterjedtek, mivel fejlesztésük és módszertani megújulásuk lassabb, mint a modernebb vagy R-alapú alternatíváké.

Bármilyen módon is implementálták a CFA-t egy adott szoftverben, vannak olyan paraméterei az elemzésnek, melyeket minden szoftverben be lehet és be is kell állítani.

A CFA első lépése a hiányzó értékek kezelése, hiszen ezek elhanyagolása vagy rossz kezelése torzíthatja a becsléseket. A hiányzó értékek kezelésének több módja létezik. Az egyik leggyakoribb megoldás a listwise törlés (egy adatsor teljes kihagyása az elemzésből, ha bármely releváns változóban hiányzik az érték), azonban ez csökkenti a mintanagyságot, és torzító hatású lehet, különösen akkor, ha a hiányosság nem véletlenszerű. A másik megoldás a hiányzó értékek imputációja, melynek legegyszerűbb módja az itemátlaggal való helyettesítés,

de léteznek komplex eljárások is, melyek a többi változóból következtetnek a hiányzó adat potenciális értékére. Fontos tudni, hogy imputációkor fennáll a veszélye, hogy megváltoznak az összefüggések a tételek között, ami torzíthatja a CFA eredményeit (de kevés, például <5% hiányzó adat esetén e torzítottság elfogadható mértékű lehet). Harmadik megoldásként bizonyos paraméterbecslési eljárások beépített módon kezelik a hiányzó adatokat a modelltesztelés folyamatában. A FIML egy ilyen, hatékony módszer, amely az összes elérhető információt hasznosítja anélkül, hogy előzetes adatok helyreállítására lenne szükség. Más becslők, mint például a DWLS, nem támogatják ezt a fajta hiányzókezelést, és kizárólag listwise vagy pairwise törlést tesznek lehetővé.

Az adatok beolvasását követően, függetlenül attól, hogy grafikus (például AMOS), menüs (például JASP) vagy szöveges (például Mplus vagy lavaan) felületet használunk, a következő lépés a modell specifikálása. Ennek részeként fel kell venni és elnevezni a látens változókat, majd egyirányú nyilakkal hozzárendelni az egyes itemeket. Grafikus programokban általában külön meg kell rajzolni az itemek reziduumait is, valamint a látens változók közötti kovarianciát (kivéve, ha azok függetlennek tekintendők). Szöveges alapú programokban ezeket a rendszer általában automatikusan feltételezi, így éppen a függetlenséget kell igény esetén explicit módon jelezni. Az ennek megfelelő alapbeállítást minden programnál ellenőrizni kell, de megoldás lehet az is, ha a program alapbeállításaitól függetlenül mindenhol megadjuk, hogy melyik kovarianciákat engedjük meg, és melyeket tiltjuk. A kiinduló modell rendszerint nem tartalmaz reziduális kovarianciákat, de ha ezek indokoltak, akkor azokat is meg kell adni a specifikáció során.

A következő lépés a látens változók skálázásának beállítása, vagyis annak meghatározása, hogy a faktorok töltéseit milyen viszonyítási ponthoz képest számítottuk ki. Ennek két elterjedt megoldása van:

Beállítható a faktorok standardizált változatának használata megoldásként. Ekkor a látens változók eloszlását rögzítjük (átlag = 0 és szórás = 1), és minden faktor ebben a standard formában jelenik meg. Ennek eredményeként a faktortöltések szabályosan -1 és 1 közé esnek majd, és egymással közvetlenül összehasonlíthatóvá válnak, ami informatívabb, a látens változókat viszont csak standardizálva látjuk, átlagukban és varianciájukban nem lesz információ (hiszen azokat a beállításban rögzítettük 0-ra és 1-re).[6]

[6] Ritkán előfordul, hogy ennél a beállításnál is -1-nél kisebb vagy 1-nél nagyobb töltésérték jelenik meg, ezeket Heywood-eseteknek nevezzük, és legtöbbször valamilyen hibát jeleznek, például rossz illeszkedést, két item közötti túl erős kollinearitást vagy a manifeszt változóban lévő nem kezelt szélsőséges értékeket. Emellett eredhetnek kis elemszámból vagy skálázási problémából is. További részletekért lásd Kline (2023).

A másik gyakran alkalmazott megoldás a referenciapont kijelölése. Ekkor az adott alskála (faktor) egyik itemének faktortöltését rögzítjük 1-re. Ilyenkor a látens változó varianciája szabadon becsülhető, és ebből számítódik a többi item töltésértéke. Az outputban ezért a látens változó varianciája a modell által becsült értéket mutatja. A töltések értelmezése azonban relatívvá válik, minden item töltése az adott alskála első iteméhez képest számítódik. A töltéstáblázatban a referenciaként választott item töltése 1 lesz, az ennél nagyobb szám erősebb, az ennél kisebb pedig gyengébb kapcsolatot jelez a faktorral.

A következő alapvető döntés, hogy milyen paraméterbecslő eljárást alkalmazunk. Az elméleti alapoknál bemutattuk, hogy a leggyakrabban használt módszerek közé tartozik a maximum likelihood (ML), a diagonally weighted least squares (DWLS) és a weighted least squares mean and variance adjusted (WLSMV), de több más lehetőség is elérhető. A választás elsődlegesen a tételek mérési szintjétől és eloszlásbeli jellemzőitől függ. Folytonos, normális eloszlásúnak feltételezett változóknál az ML jól használható. Ha a multivariáns normalitás sérül, akkor ML helyett érdemes robusztus változatokat választani, ez lehet az MLR, MLM, MLMV. Emellett megoldás lehet a WLSMV is, mely egyáltalán nem igényli a normalitást. Ordinális skálájú adatoknál a WLSMV vagy a DWLS bizonyulhat megfelelőbbnek. Ez kifejezetten igaz a dichotóm és ötnél kevesebb szinten mérő Likert-skálákra, míg az ötfokú Likert-skálától felfelé az ML becslésének pontossága sem marad el.

Ezen túlmenően érdemes figyelembe venni a mintanagyságot. Az ML általában nagy mintán működik jól, kisebb mintán instabil lehet. Ezzel szemben a DWLS és WLSMV kifejezetten kis és közepes mintákra ajánlott. Ez befolyásolja a szükséges számítási erőforrásokat is, mivel bizonyos eljárások (például a WLSMV) nagy minta és komplex modell esetében jelentős számítási kapacitást igényelnek.

Fontos döntési szempont lehet a hiányzó értékek kezelésének képessége. Míg az ML képes FIML-lel kezelni a hiányzó értékeket, addig a DWLS és WLSMV általában listwise törléssel vagy imputációval működik.

A paraméterbecslő eljárás meghatározhatja azt is, mely illeszkedési mutatók érhetőek el (például DWLS-nél gyakran hiányzik a khi-négyzet-teszt), illetve befolyásolja azok számítási módját.

Végül a szoftveres támogatottság és a kutatási gyakorlat is szerepet játszhat, mert egyes módszerek bizonyos programcsomagokban (például Mplus, lavaan) könnyebben elérhetők, vagy ajánlott alapbeállításként szerepelnek, illetve nem minden szoftver kínálja az összes lehetőséget. A publikációkban leggyakrabban ML-lel találkozhatunk (főleg azért, mert ez volt az első megoldások egyike), de Likert-skálás adatoknál a bírálók gyakran WLSMV-t kérnek.

A paraméterbecslések mellett mindig fontos a hozzájuk tartozó standard error (SE) meghatározása is, hiszen ez mutatja a becslés bizonytalanságát. A hagyományos, analitikus SE-számítás (az úgynevezett delta-módszer) nem minden esetben megbízható, mivel torzított lehet, különösen, ha a normalitás vagy a függetlenség feltételei sérülnek. Ezt ellensúlyozzák a robusztus szendvicsesztimátorok (robust sandwich estimators), amelyek korrigálják a torzításokat és stabilabb SE-ket biztosítanak. Gyakran használt változataik például az MLR, az MLM, az MLMV és a WLSMV becslőkhöz kapcsolódó robusztus SE-k.

Egy másik megoldás a bootstrapping, amely a mintából történő ismételt mintavételezést alkalmazza a standard error becslésére. A bootstrap módszer előnye, hogy nem igényli a paraméterek normalitását, valamint jól használható komplex modellek vagy kis minták esetén is. Az így kapott SE-k és konfidenciaintervallumok megbízhatóbbak lehetnek, különösen nem normális eloszlások vagy heteroszkedasztikus adatok esetén. (De ne essünk túlzásba, a bootstrapping nem csodaszer a rendkívül kicsi elemszám vagy kifejezetten ferde/csúcsos adatok kezelésére.)

Végül az is fontos, hogy beállítsuk, melyik illeszkedési mutatókat szeretnénk megtekinteni (jó gyakorlat, ha lekérjük legalább a χ^2, CFI, TLI, RMSEA és a hozzá tartozó 90% CI és az SRMR értékeit). Emellett számos egyéb statisztikai mutató is lekérhető, amelyek a modell megfelelőségét vagy az adatok minőségét segítenek ellenőrizni. Ilyenek például a Kaiser–Meyer–Olkin (KMO) -teszt a minta megfelelőségvizsgálatához, a Bartlett-féle szfericitásteszt, az AVE a konvergens validitás mérésére, a HTMT a diszkriminánsvaliditás értékelésére (csak ott van értelme, ahol több alskálája is van a kérdőívnek), valamint a reliabilitási mutatók (Cronbach-alfa, McDonald-omega). További hasznos indikátorok lehetnek az R^2 értékek, melyek az itemek magyarázott varianciáját mérik, a reziduális mátrix statisztikái, melyek a modellhibák feltárására alkalmasak, a modifikációs indexek, melyekkel a lehetséges javítási irányokat azonosíthatjuk, valamint a skálaátlagok és szórások, amelyek segítenek a látens változók értelmezésében. Ezen mutatók kombinált vizsgálata adja a legteljesebb képet a modell megbízhatóságáról és validitásáról.

3.4. Elemzés értelmezése

Az elemzés értelmezésének első lépése a modell megfelelőségének ellenőrzése. Ehhez először a minta faktoranalízisre való alkalmasságát vizsgáljuk, tipikusan a KMO-mutatóval és a Bartlett-féle szfericitástesszttel. A KMO értéke általában

0,6 felett elfogadható, 0,7 felett jó, míg a Bartlett-tesztnek szignifikánsnak kell lennie.

Ezután áttekintjük a korábban kiválasztott illeszkedési mutatókat, mint a χ^2-teszt, a χ^2/df arány, a CFI, TLI, IFI, a RMSEA és annak 90%-os felső konfidenciahatára, valamint az SRMR. Ezek határértékeit a 2.4. fejezetben közöltük. A kapott értékek alapján mindig holisztikusan döntünk a modell illeszkedéséről.

Ha az illeszkedés nem megfelelő, a faktortöltések áttekintése helyett a modellt szakmailag átgondolt módon javítjuk, vagy megnézzük, mit javasolnak a modifikációs indexek (MI). Az MI-érték azt becsüli, hogy egy adott változtatás mennyiben csökkentené a χ^2-értéket, ennek megfelelően a magasabb MI nagyobb javulást jelent a modell illeszkedésében. Fontos, hogy az MI-értékekhez nem köthető merev, univerzális küszöbérték (mivel azok nagysága statisztikailag erősen függ a mintaelemszámtól), így mechanikus alkalmazásuk helyett a kiugróan magas értékeket érdemes keresni (Kline, 2023) és csak azokat a javaslatokat figyelembe venni, amelyek magas MI-érték mellett szakmailag is indokolhatók. Egyszerre mindig csak egy változtatást hajtunk végre. Ha egy tétel túl sok vagy túl erős kereszttöltésben vesz részt, illetve túl sok reziduális kovarianciája van, érdemes megfontolni a tétel eltávolítását. Két különböző skálán lévő tétel között felmerülő reziduális kovariancia esetén szintén az egyik tétel törlése javasolt, míg az azonos skálán belüli reziduális kovariancia megengedhető, ha szakmailag indokolt. Grafikus programokban (például AMOS) a két item reziduálisát kétirányú nyíllal kötjük össze, a szöveges felületeken (például lavaan) a dedikált jelölést (~~) használjuk, menüs felületen (például JASP) a reziduális kovariancia a megfelelő menüpontban adható meg. Ezeket a módosításokat addig végezzük, amíg a modell illeszkedése megfelelő nem lesz, de legfeljebb addig, amíg van olyan MI-javaslat, amely jelentős és szakmailag indokolható.

Ha a modell illeszkedése már megfelelő, áttekintjük a faktortöltéseket, ahol minden itemnek megfelelően erősen ($\lambda \geq 0{,}50$), helyes irányban és szignifikánsan kell töltenie, de elméleti szempontból fontos tételeknél néha $\lambda \geq 0{,}40$ is megengedett. Ha egy tétel töltése ennél is alacsonyabb, indokolt lehet a tétel törlése. Ellenőrizzük az R^2-értékeket is, a legtöbbször $R^2 > 0{,}20$ az elfogadott, de tartsuk észben az értelmezésnél, hogy a reziduális kovarianciában részt vevő tételek R^2-értéke természetesen alacsonyabb lehet, hiszen az adott item varianciájának egy része átkerül a másik itemmel közös reziduális kapcsolatra. A gyenge teljesítményű itemeket ilyenkor még törölni szokás, de törlést követően újra ellenőrizni kell a modellt.

A látens változók közötti kovarianciát is áttekintjük, ellenőrizve, hogy nincs-e szokatlanul magas vagy ellentmondásos kapcsolat.

Végül, ha minden fő szempont rendben van, vizsgálhatjuk az egyéb mutatókat, például a konvergens validitást (AVE ≥ 50%), a diszkriminánsvaliditást (HTMT ≤ 0,85), valamint a reliabilitást (Cronbach-alfa és McDonald-omega ≥ 0,70). Ezek a kiegészítő ellenőrzések biztosítják, hogy a kérdőív vagy a mérési modell statisztikailag és szakmailag is megbízható legyen. Ha terveztünk multigroupelemzést, azt mindezek után futtatjuk le (lásd a SEM-ről szóló tankönyvfejezet).

4. Mikor használjuk és mikor ne használjuk a módszert?

A CFA-t akkor érdemes használni, ha előzetes elméleti vagy empirikus modellünk van arról, hogyan szerveződnek a tételek látens változókba. Ilyen eset például, amikor egy már létező kérdőív struktúráját szeretnénk ellenőrizni, vagy amikor új mérőeszközt fejlesztünk, és világos hipotézissel rendelkezünk az egyes faktorokhoz tartozó itemekről. A CFA segítségével tesztelhető, hogy az adatok mennyire illeszkednek a feltételezett faktorstruktúrához, és lehetőség van különböző alternatív modellek összehasonlítására is.

Nem célszerű viszont CFA-t alkalmazni, ha nincsenek világos előfeltevéseink a változók közötti kapcsolatokról, vagy ha feltáró jelleggel vizsgáljuk az adatokat. Ilyenkor a feltáró faktoranalízis (EFA) a megfelelőbb módszer, amely lehetővé teszi a faktorstruktúra szabadabb feltérképezését.

Szintén problémás lehet a CFA használata túl kis mintanagyság mellett, mivel a modellilleszkedési mutatók megbízhatatlanná válhatnak, illetve ha túl sok a tétel és a paraméterek száma a minta méretéhez képest.

5. A minőség biztosítása

A CFA eredményeit úgy érdemes megírni, hogy a közölt információk lehetővé tegyék az olvasó számára a modell minőségének megítélését. A publikációs gyakorlatban ezért különbséget tehetünk a minimálisan szükséges, az ajánlott, illetve a kiváló szintű közlés között. A szükséges szint azokat a mutatókat és információkat tartalmazza, amelyek nélkül a tanulmány eredményei nem értékelhetők megbízhatóan. Az ajánlott szint ezenfelül olyan részleteket is közöl, amelyek teljesebb képet adnak a modell erősségeiről és gyengeségeiről. A kiváló szint pedig a lehető legátfogóbb közlést jelenti, amikor a szerző minden jelentős illeszkedési mutatót, megbízhatósági és érvényességi mutatót, valamint kiegészítő elemzést is bemutat, így az olvasó teljeskörűen át tudja tekinteni a modell megfelelőségét.

Szükséges közlés (minimum elvárás):

a) A modell leírása tartalmazza, hogy hány faktort vizsgált a kutató, és mely itemek melyik faktorhoz tartoznak.
b) A kutató közli a mintanagyságot és az adatok forrását.
c) A modelllilleszkedési mutatók közül legalább a legfontosabbak (χ^2, CFI, RMSEA, SRMR) közlésre kerülnek.
d) Bemutatásra kerülnek az itemek standardizálatlan és/vagy standardizált faktortöltései és azok szignifikanciája.
e) Megtörténik a módosítások (itemtörlések, reziduális kovarianciák) összefoglaló közlése.
f) Bemutatásra kerülnek a faktorok közötti korrelációk.

Ajánlott közlés (jó gyakorlat) a szükséges közlésen felül:

a) Több illeszkedési mutató is közlésre kerül, ezek értelmezése és összevetése is megtörténik.
b) Ha a modell módosításokon esett át (itemtörlések, reziduális kovarianciák), megtörténik a módosítások részletes közlése (mely változók között mi történt), a részmodellek illeszkedésének közlése és a módosítások szakmai értelmezése a diszkusszióban.
c) A kutató összehasonlít alternatív modelleket (például egy- és többdimenziós, first- és second-order).

d) Az itemek részletesebb bemutatása tartalmazza a standardizálatlan és/vagy standardizált faktortöltéseket, a faktortöltések standard errorját és konfidenciaintervallumát, valamint a szignifikanciáját.
e) Bemutatásra kerülnek az itemek R-négyzet-értékei.
f) A kutató vizsgálja a faktorok megbízhatóságát (például Cronbach-alfa, McDonald-omega, kompozit reliabilitás).
g) A kutató kitér a konvergens és diszkriminánsvaliditás elemzésére (például AVE, Fornell–Larcker, HTMT).
h) A tanulmánynak része a kritériumvaliditás és a demográfiai mutatókkal való összefüggések vizsgálata.

Kiváló közlés az ajánlott közlésen felül:

a) Szükség esetén Multigroup CFA-val vizsgálja az invarianciát a kutatás szempontjából releváns csoportok között (például nem, életkor, kultúra).
b) Megtörténik a keresztvalidálás (például minta felezése, újramérés).
c) Ha lehetséges, a kutató javasol rövidített vagy alternatív modelleket.
d) A modellt áttekinthetően megszerkesztett ábra segítségével is bemutatja (faktorok, itemek, útvonalak).
e) A tanulmánynak része a végső kérdőív tételszintű közlése és a kiértékelési útmutató megadása.
f) Megtörténik a hiányzó adatok kezelésének és normalitásvizsgálatnak a részletes leírása.

6. A módszer változatai és haladó alkalmazásai

A CFA-nak több kiterjesztett vagy haladó alkalmazása is van, amelyek túlmutatnak az „egyszerű” egycsoportos modellillesztésen. Ezeket gyakran használják akkor, amikor a kutatási kérdés bonyolultabb, vagy amikor összehasonlításokra, speciális hipotézisek tesztelésére van szükség. A legfontosabbak:

a) Multigroup CFA (MGCFA): ezzel tesztelhető, hogy ugyanaz a faktorstruktúra több csoportban (például nemek, életkori csoportok, kulturális minták) azonos módon működik-e. Az invariancia vizsgálata három (vagy négy) fő szinten zajlik: konfigurális, metrikus, skalár (és szigorú) invariancia. Erről részletesebben lásd a könyv SEM-fejezetét.
b) Hierarchikus CFA (vagy second-order CFA): amikor az alsóbb rendű faktorokat egy magasabb rendű faktor fogja össze, ennek megfelelően a first-order látens változók felett megjelenik egy second-order szint is.
c) Bifaktor CFA: minden item egyszerre tölt egy általános faktorra és egy specifikus faktorra, így különválasztható az általános és a tartományspecifikus hatás. Míg a hierarchikus CFA inkább a főskála létjogosultságát teszteli, addig a bifaktor CFA inkább az alskálákét.
d) Módszertényező (method factor) használata: a módszertényező olyan látens faktor, amelyet a mérési módszerből vagy a tesztelés körülményeiből fakadó közös variancia megmagyarázására vezetünk be a modellbe, nem pedig a vizsgált pszichológiai konstrukció részeként. Azaz a módszer kifejezetten a torzító tényezők hatását vizsgálja. Például ha egy kérdőívben sok fordított tétel van, vagy ha a válaszadási stílus torzíthat, külön módszertényezőt lehet beállítani ezekhez.
e) Longitudinális CFA: idősoros vagy paneladatoknál segít megvizsgálni, hogy a konstrukció ugyanúgy mérhető-e az időben (longitudinális invariancia).
f) Klaszterszempontot tartalmazó CFA: olyan beállítás, amely figyelembe veszi, hogy a vizsgálati személyek nem függetlenek, hanem valamilyen csoportosító egységhez (például családok, osztályok, munkahelyi csoportok, kórházi osztályok, országok) tartoznak.

7. Összefoglalás

A megerősítő faktoranalízis (CFA) egy olyan statisztikai módszer, amely teszteli a pszichológiai vagy szociológiai konstrukciók mérésének strukturális validitását azáltal, hogy a megfigyelt itemek és a látens faktorok közötti kapcsolatokat modellezi. A CFA-t leggyakrabban kérdőívek fordításakor, adaptálásakor vagy új skálák fejlesztésénél alkalmazzák, de a strukturális egyenletmodellezés

(SEM) mérési modelljének részeként is fontos eszköz. Az elemzés során először a modell specifikálása történik meg látens faktorok és itemek hozzárendelésével, majd a paraméterbecslés (például ML, DWLS) következik, figyelembe véve a skálázást és a feltételezéseket. Az illeszkedési mutatók (például khi², CFI, TLI, IFI, RMSEA, SRMR) ellenőrzését követően szükség esetén a modifikációs indexek segítségével finomítható a modell. A további értelmezés során áttekintjük a faktortöltéseket, reziduálisokat, a látens változók közötti kovarianciát, valamint a validitási és megbízhatósági mutatókat. A CFA tehát komplex, de jól strukturált eszköz a mérési konstrukciók érvényességének és megbízhatóságának ellenőrzésére, amely megfelelően tervezett mintán és adatszerkezeten, átgondolt paraméterbeállításokkal és a modell illeszkedésének holisztikus értékelésével megbízható eredményeket szolgáltat.

8. Önellenőrző kérdések

1. kérdés: Mi a manifeszt és mi a látens változó?

Válasz: A manifeszt változó közvetlenül mérhető, például kérdőíves tétel. A látens változó nem mérhető közvetlenül, csak a manifeszt változókon keresztül, hiszen abban fejeződik ki, például személyiségjegyek.

2. kérdés: Mit jelent, hogy az elméleti modell illeszkedik a mért adatokhoz?

Válasz: Azt, hogy a modell által feltételezett variancia-kovariancia mátrix jól közelíti a ténylegesen megfigyelt variancia-kovariancia mátrixot az illeszkedési mutatók elfogadható értékei alapján.

3. kérdés: Mik a legismertebb illeszkedési mutatók, mikor tekintjük elfogadhatónak az illeszkedést?

Válasz: khi-négyzet-teszt (ha nem szignifikáns), χ^2/df (< 3), CFI, TLI, IFI (≥ 0,90), RMSEA (≤ 0,08), SRMR (≤ 0,08).

4. kérdés: Mire jók a modifikációs indexek, mikor szabad hallgatni rájuk?

Válasz: Segítenek azonosítani, hogy mely lehetséges módosítások javítanák a modell illeszkedését. Csak szakmailag indokolt, értelmes változtatásoknál szabad követni őket, mert az automatizált, minden ajánlást követő módosítás torzíthat. Leggyakoribb változtatás a reziduális kovariancia megengedése az azonos skálán mérő itemek között. Más változtatási javaslatai (például keresztöltések) jó jelzések lehetnek arra, ha egy itemet törölni érdemes.

5. kérdés: Mit mutatnak meg az itemek faktortöltései?

Válasz: Hogy az adott tétel mennyire és milyen irányban kapcsolódik a faktorhoz. Az erősebb, szignifikáns töltés ($|\lambda| > 0{,}50$) azt jelzi, hogy a tétel jól méri a faktort.

Ajánlott irodalom

Bollen, K. A. (1989). Structural equations with latent variables. John Wiley & Sons. https://doi.org/10.1002/9781118619179

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.

Costa, P. T., Jr., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), The SAGE handbook of personality theory and assessment, Vol. 2. Personality measurement and testing (pp. 179–198). Sage Publications, Inc. https://doi.org/10.4135/9781849200479.n9

Doval, E., Viladrich, C., & Angulo-Brunet, A. (2023). Coefficient Alpha: The Resistance of a Classic. *Psicothema, 35*(1), 5–20. https://doi.org/10.7334/psicothema2022.321

Groskurth, K., Bluemke, M., & Lechner, C. M. (2024). Why we need to abandon fixed cutoffs for goodness-of-fit indices: An extensive simulation and possible solutions. *Behavior Research Methods, 56*(4), 3891–3914. https://doi.org/10.3758/s13428-023-02193-3

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Howard, M. C., Boudreaux, M., Cogswell, J., Manix, K. G., & Oglesby, M. T. (2025). A literature review of model fit and model comparisons with confirmatory factor analysis: Formalizing the informal in organizational science. *Applied Psychology, 74*(1), Article e12592. https://doi.org/10.1111/apps.12592

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kalkbrenner, M. T. (2024). Choosing Between Cronbach's Coefficient Alpha, McDonald's Coefficient Omega, and Coefficient H: Confidence Intervals and the Advantages and Drawbacks of Interpretive Guidelines. *Measurement and Evaluation in Counseling and Development, 57*(2), 93–105. https://doi.org/10.1080/07481756.2023.2283637

Kline, R. B. (2023). *Principles and practice of structural equation modeling.* The Guilford Press.

Marsh, H. W., Wen, Z., & Hau, K. T. (2004). Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction. Psychological Methods, 9, 275–300. https://doi.org/10.1037/1082-989X.9.3.275

Maslach, C., Jackson, S. E., & Leiter, M. P. (2018). Maslach Burnout Inventory manual (4th ed.). Mind Garden, Inc.

McNeish, D., & Dumas, D. (2025). Reliability representativeness: How well does coefficient alpha summarize reliability across the score distribution? *Behavior Research Methods, 57*(3), 93. https://doi.org/10.3758/s13428-025-02611-8

Roos, J. M., & Bauldry, S. (2022). *Confirmatory Factor Analysis*. SAGE. https://doi.org/10.4135/9781071938959

Rönkkö, M., & Cho, E. (2022). An Updated Guideline for Assessing Discriminant Validity. *Organizational Research Methods, 25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Savalei, V., & Reise, S. P. (2019). Don't Forget the Model in Your Model-based Reliability Coefficients: A Reply to McNeish. *Collabra: Psychology, 5*(1), Article 36. https://doi.org/10.1525/collabra.247

Schaufeli, W. B., De Witte, H., & Desart, S. (2019). Burnout Assessment Tool (BAT) – Test Manual. KU Leuven, Belgium: Internal report.

IX. ÚTVONALELEMZÉS (PA)

SOLTÉSZ-VÁRHELYI KLÁRA

A fejezet a következő kérdéseket válaszolja meg:

a) Milyen elméleti alapokon nyugszik a PA?
b) Melyek a PA gyakorlati lépései, és hogyan kell az elemzést helyesen beállítani, az eredményeket helyesen értelmezni?
c) Hogyan értelmezhetők a direkt és indirekt hatások, azaz mi a jelentősége a mediált útvonalaknak?
d) Milyen minőségi ismérvek alapján különböztethető meg egy kiváló PA-tanulmány egy átlagostól?

A fejezet tudásanyaga részben átfed a CFA-fejezetben tárgyalttal, így e fejezet megkezdése előtt ajánlott annak feldolgozása.

1. Bevezetés

Az *útvonalelemzés*[1] (vagy útvonal-modellezés, path analysis, PA) a változók közötti kapcsolatok feltárására és statisztikai vizsgálatára szolgáló módszer, amely lehetővé teszi, hogy a kutató az elméletileg levezetett, feltételezett okokozati modellek adatokra való illeszkedését tesztelje. A módszer lényege, hogy a vizsgált változókat nem külön-külön elemzi, hanem hálózatként, ahol az egyes utak (nyilak) irányított hipotetikus hatásokat jelenítenek meg. Ezáltal a PA sokkal árnyaltabb képet nyújt, mint a hagyományos regressziós elemzés, mert egyidejűleg több közvetlen és közvetett hatást is képes számszerűsíteni.

Egy lehetséges vizsgálat például azt célozhatná, hogy miként hat a vezető érzelmi intelligenciája a beosztottak közötti bizalmi légkörre és ezen keresztül a munkahelyi kiégés mértékére. A modellben az érzelmi intelligencia közvetlen hatással lehet a bizalomra, és közvetett hatással a kiégésre, miközben a bizalmi légkör közvetlenül befolyásolja a kiégést. Az útvonalmodell segítségével világosan elkülöníthetők ezek a direkt és indirekt kapcsolatok, valamint becsülhető a teljes hatás nagysága is.

Ahogy a megerősítő faktoranalízis (CFA) önálló elemzésként a *strukturális egyenletmodellezés* (SEM) mérési modelljének, úgy az útvonalelemzés önálló elemzésként a SEM strukturális modelljének feleltethető meg. A PA tehát nem tartalmaz látens változókat, kizárólag a megfigyelt változók közötti feltételezett kapcsolatokra összpontosít, ezzel megteremtve az alapot a későbbi, teljes strukturális egyenletmodellek felépítéséhez.

[1] Az útvonalelemzés módszertanát eredetileg Sewall Wright genetikus dolgozta ki az 1920-as években (Wright, 1921); a technika a későbbi társadalomtudományi adaptáció során vált a strukturális egyenletek modelljének (SEM) szerves részévé, melynek modern matematikai keretrendszerét és szintézisét Bollen (1989) nagy hatású munkája rögzítette.

2 Elméleti alapok

A PA módszertana abból áll, hogy először megfogalmazunk egy elméleti összefüggésrendszert a vizsgált változók között, majd ezt matematikailag, egyenletek formájában specifikáljuk. Ezt követően megvizsgáljuk, hogy a modell identifikálható-e, azaz rendelkezésünkre áll-e elegendő független információ a paraméterek becsléséhez. Ha az identifikáció teljesül, a paraméterek értékeit becsüljük, majd ellenőrizzük, mennyire illeszkedik a modell az empirikus adatokhoz. Szükség esetén módosításokat hajtunk végre a modellen, végül pedig értékeljük és értelmezzük az eredményeket. Ennek lépéseit tekintjük most végig.

2.1. A modell specifikációja (model specification)

Az útvonalelemzés változók közötti kapcsolatrendszer modellezéséből áll. Amikor önálló statisztikai módszerként alkalmazzuk, akkor úgynevezett *manifeszt (mért) változók* kapcsolatát vizsgáljuk, azaz nincsenek látens tényezők, minden változó egy-egy oszlopként jelenik meg az adatbázisunkban. Ennek megfelelően a grafikus modelleken a változókat téglalapban jelenítjük meg. (Ezzel szemben, amikor az útvonalelemzés a SEM strukturális modellezéseként jelenik meg, természetesen előfordulhatnak majd látens változók is az útvonal részeként, melyeket oválissal jelölünk majd.)

A legegyszerűbb statisztikai kapcsolat két változó között a korreláció. A strukturális egyenletmodellezésben hagyományosan nem a korrelációt, hanem ennek standardizálatlan formáját, a *kovarianciát* használjuk. A kovariancia (hasonlóan a korrelációhoz) azt fejezi ki, hogy a két változó értékei milyen irányban és mértékben mozognak együtt, tehát ha az egyik nő, a másik is inkább nő vagy éppen csökken. A strukturális egyenletmodellezésben a kovariancia egyfajta természetes alap, mert abból mind a korrelációk, mind a varianciák levezethetők.[2]

A kovariancia során a két változó kapcsolata irányítatlan, azaz a modell nem tételezi fel, hogy egyik magyarázza a másikat, csupán azt, hogy összefüggés van közöttük. Ezt a kétirányú kapcsolatot egy kétirányú nyíl jelöli a modellábrán.

[2] A kovariancia és korreláció matematikai viszonyát, illetve a kovarianciaalapú számítás módszertani elsődlegességét a SEM-irodalomban Bollen (1989) és Kline (2023) tárgyalja alapvetésként.

Erre láthatunk példát a 14. ábrán, ahol X1 és X2 változó között a kétirányú nyíl mutatja az együtt járásukat, vagyis a kovarianciájukat. A fejezetben használt példában szépség és önbizalom összefüggését vizsgáljuk majd. A kovariancia csak annyit feltételez, hogy a két változó együtt jár, de nincs arra feltételezés, hogy a szép embereknek jobb az önbizalma, vagy a jó önértékeléssel rendelkező embereket látjuk szebbnek.

Az ábrán ezenkívül minden változóból egy-egy kétirányú nyíl indul ki és tér vissza önmagába. Ez a változó *varianciáját* jelöli. A variancia egy változó értékeinek szóródását mutatja a saját átlaga körül. Matematikailag a változó önmagával vett kovarianciája adja a varianciát, ezért jele az önmagára visszamutató kétirányú nyíl. Az ábrán jelölve láthatjuk az X1 és az X2 változó varianciáját is. Mivel ebben az egyszerű modellben egyik változó sem magyarázza a másikat, mindkettő teljes változatosságát önmagában tüntetjük fel. Ezt a jelölésben úgy fejezzük ki, hogy VAR(X1) és VAR(X2), vagyis az X1 és X2 változók teljes varianciája megjelenik a modellben, függetlenül attól, hogy közöttük kovariancia is fennáll.

14. ábra: A kovariancia modellje

Az *egyszerű lineáris regresszió* a következő lépés a kovariancián alapuló modellezés után. Az útmodellt a 15. ábra A panelje szemlélteti. Itt továbbra is két változó szerepel, de már kijelöljük, melyik a *magyarázó* és melyik a *magyarázott változó*. Azt teszteljük, hogy a szépséggel lehet-e magyarázni az önbizalmat. Ez nem jelenti azt, hogy a valóságban csak ez az irány létezhet, csak annyit, hogy a jelen modellben mi ezt az irányt teszteljük. Mivel egyik változó magyarázza a másikat, ezért a kapcsolatot egyirányú nyíllal ábrázoljuk, amely a *regressziós együtthatót (útvonal-együttható, koefficiens, β)* jelöli. A két változó nem egyenrangú, az egyik a *prediktor/predikáló változó* (X), a másik a *kimeneti/predikált* változó (Y).

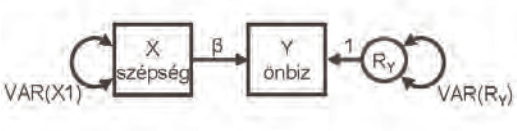
A prediktor *exogén változóként* szerepel a modellben, vagyis olyan tényezőként, amelyet nem a modell más változói határoznak meg, hanem kívülről, függetlenül kerül bevezetésre, tehát a modellel nem próbáljuk magyarázni, nem fut bele egyirányú nyíl. Ennek megfelelően varianciáját változatlanul egyben

kezeljük, azaz VAR(X) formában jelenítjük meg. A kimeneti változó ezzel szemben *endogén változó*, ami azt jelenti, hogy értékeit a modell más változói (jelen esetben az exogén prediktor) magyarázzák.[3] Ennek megfelelően varianciáját két részre bontjuk. Az egyik rész a *modell által megmagyarázott variancia*, amely a prediktor hatásából származik. A másik rész a *meg nem magyarázott*, az úgynevezett *reziduális variancia*, amelyet a tankönyvben nagy R-rel jelölünk. Így Y teljes varianciája a megmagyarázott és a reziduális komponens összegeként jelenik meg. A reziduum becsült érték, ennek megfelelően nem téglalapban, hanem oválisban jelenítjük meg a modellábrákon.

A kérdés ezek után az, hogyan számoljuk ki a modell fő elemeit, a regressziós együttható értékét, a kimeneti változó megmagyarázott varianciáját, valamint a fennmaradó, meg nem magyarázott, reziduális varianciát. Ezen értékek kiszámítása adja az egyszerű lineáris regresszió matematikai alapját, és megmutatja, hogyan formalizáljuk a magyarázó hatásokat az útvonalmodellekben.

Ahhoz, hogy a megbecsülendő paraméterek értékeit kiszámíthassuk, először fel kell írnunk a strukturális modellezés egyenletrendszerét. Ehhez viszont tisztában kell lennünk azzal, milyen információk állnak rendelkezésünkre. Egy egyszerű lineáris regresszió esetén három információról beszélhetünk: a prediktor változó teljes változatosságáról (VAR(X)), a kimeneti változó teljes változatosságáról (VAR(Y)), valamint a kettő közötti kapcsolatról (COV(X,Y)). Ezeket *független információk*nak nevezzük, és szokás őket egy táblázatban összefoglalni, amelyet a *megfigyelt (mért) variancia-kovariancia mátrixnak* nevezünk (15. ábra B panelje).

A. Egyszerű lineáris regresszió útmodellábrája



B. Egyszerű lineáris regresszió mért variancia-kovariancia mátrixa

	X	Y
X	VAR(X)	
Y	COV(X,Y)	VAR(Y)

15. ábra: Egyszerű lineáris regresszió modellje

Minden egyes független információhoz felírható egy egyenlet arra nézve, hogy miként áll elő az értéke a modell alapján. Az egyenletek felírásához speciális szabályokat követünk, ezeket *követési szabályok*nak (angolul *tracing rules*) hívják.

[3] A megértés segítheti, ha vizualizáljuk az összefüggéseket, és megfigyeljük, hogy az exogén változók kizárólag forrásként vannak jelen (csak indul belőlük nyíl), míg a szakirodalmi definíció szerint minden olyan változó, amelybe a modellben legalább egy nyíl érkezik – még ha közvetítőként tovább is adja a hatást –, technikailag már endogénnek minősül (Kline, 2023).

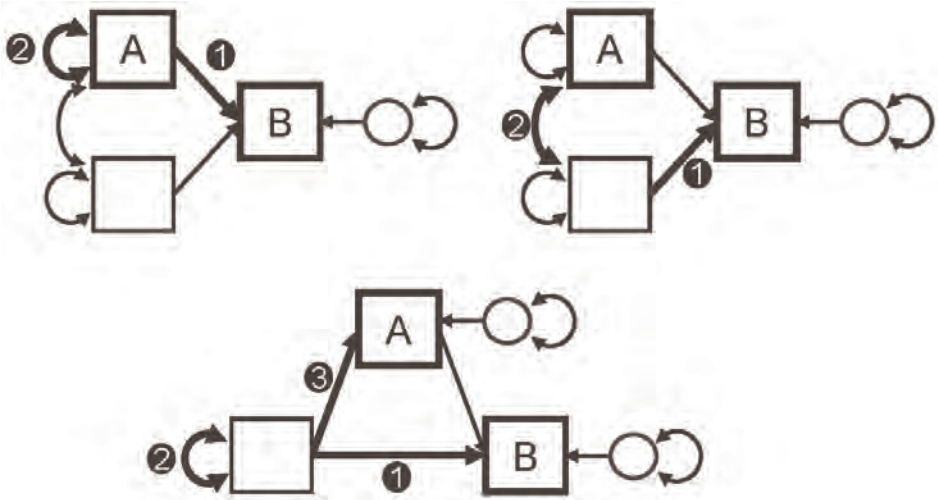
A lehetséges útvonalak bejárásához a számítások átláthatósága érdekében érdemes egy következetes eljárásrendet[4] követni: a két változó közötti kapcsolat levezetésénél célszerű a kimeneti változóból kiindulni, és a nyilak irányával szemben, hátrafelé haladva feltérképezni a prediktorhoz vezető utakat. Ezt a gyakorlati megközelítést alkalmazva háromféleképp haladhatunk:

1. Hátrafelé haladunk az egyirányú nyilakon egészen addig, amíg el nem érjük a célunkat, majd egy kétirányú nyíllal (a prediktor varianciájával) zárjuk le az utat.
2. Hátrafelé haladunk az egyirányú nyilakon, amíg el nem érjük a prediktorral kovarianciában lévő szomszédos változót, és a kovariancia kétirányú nyilán „átugrunk" a célunkhoz.
3. Hátrafelé haladunk az egyirányú nyilakon, egy kétirányú nyílon „megfordulunk", majd az egyirányú nyilakon előrehaladva érkezünk meg a célunkig. Itt már nem tesszük hozzá a célváltozó varianciájának kétirányú nyilát, mert minden útvonal csak egy kétirányú nyilat tartalmazhat, és ez az útvonal már tartalmaz egyet, azt, ahol megfordultunk.

Ha egy változó varianciájának képletét szeretnénk felírni, akkor abból indulunk ki, hogy a variancia egy változó önmagával vett kovarianciája, így a lehetséges útvonalak ugyanabba a változóba érkeznek vissza, ahonnan kiindultak.

A 16. ábra példát hoz a három lehetséges útvonalra. Fontos tudni, hogy a példák csak egy-egy útvonaldarabot mutatnak be, például a felső két ábrán látható mindkét útvonalon el lehet jutni A-ból B-be, és A és B kovarianciáját a kettő közösen adja majd.

[4] Bár Sewall Wright (1934) eredeti szabályai a lehetséges útvonalak érvényességét definiálják, és nem rögzítenek kötelező haladási irányt, a Bollen-féle strukturális megközelítés (Bollen, 1989) az egyenletek algebrai felírásakor, a dekompozíció átláthatósága és a szisztematikus levezetés érdekében, a kimeneti változóból kiinduló visszakövetés logikáját alkalmazza.



16. ábra: Útvonalpéldák a követési szabályokra

Ezeket alkalmazva három egyenlet írható fel, amelyeket a 15. ábra A paneljén érdemes végig is követni.

A szépség és önbizalom, azaz az X és Y közötti kovarianciának a számítása esetén Y-ból indulunk, visszafelé haladunk az egyirányú nyílon (β), amíg elérjük X-et, majd lezárjuk az utat X varianciájának kétirányú nyilával (VAR(X)). Ennek megfelelően az első képlet:

$$COV(X,Y) = \beta * VAR(X)$$

Y varianciájának felírásához Y-ból indulunk, és oda kell visszatérnünk. Ez kétféleképpen lehetséges. Egyrészt elindulhatunk visszafelé a β-n, megfordulunk VAR(X)-en, majd visszatérünk előrefelé β-n, ez Y varianciájának X által megmagyarázott része. Másrészt elindulhatunk a reziduális irányába is: elindulunk az 1-esen, megfordulunk a $VAR(R_Y)$-on és visszatérünk az 1-esen, ez az Y varianciájának azon része, amelyet X nem tud megmagyarázni. Ennek megfelelően a képlet:

$$VAR(Y) = \beta * VAR(X) * \beta + 1 * VAR(R_Y) * 1.$$

Vagy egyszerűsítve:

$VAR(Y) = \beta^{2*}VAR(X) + VAR(R_Y)$.

A harmadik egyenlet triviális. Mivel X exogén változó, és varianciáját nem bontjuk, ezért azt önmagával becsüljük, azaz VAR(X) = VAR(X).

Összefoglalva, ha a triviális egyenletet nem számoljuk, akkor két egyenletünk van, bennük két ismeretlennel, vagyis két becsülendő paraméterrel. Ez tehát egy kétismeretlenes, kétegyenletes egyenletrendszer, amely könnyen megoldható, és mindig egyértelmű megoldáshoz vezet. Ezt úgy fejezzük ki, hogy a modell *szaturált*, azaz pontosan annyi független információ áll rendelkezésre, mint amennyi becsülendő paraméter, így a modell mindig egyértelműen megoldható, és tökéletesen illeszkedik a megfigyelt adatokhoz. Éppen ezért egyszerű lineáris regresszió esetén nincs értelme az illeszkedést tesztelni, hiszen a modell mindig tökéletesen illeszkedik az adatokra.

A következő lépés a *többszörös lineáris regresszió* (17. ábra), amely különösen fontos tanulságot hordoz az útvonalmodellek megértésében. Amikor konceptuálisan ábrázolják, gyakran csak annyit rajzolnak fel, hogy a prediktorok egyirányú nyíllal mutatnak a kimeneti változóra, és azt mondják, „a prediktorok *kontrollálnak* egymás hatására". Ilyenkor azonban egy lényeges részlet lemarad az ábráról. Az, hogy a prediktorok között mindig megengedett a kovariancia. Ez nyitja ki igazán azt a fekete dobozt, amely a kontrollhatások mögött rejlik.

Vegyünk egy példát! Két prediktorunk van, a szépség (X1) és a népszerűség (X2), a kimeneti változó pedig az önbizalom (Y). Könnyen belátható, hogy a szépség növeli az önbizalmat, és hasonlóképp a népszerűség is. Ugyanakkor az is reális feltételezés, hogy a szépség és a népszerűség között kapcsolat áll fenn. Ebben az esetben a szépségből kétféle út vezethet az önbizalomhoz: közvetlenül (a szép emberek belenéznek a tükörbe, látják, hogy szépek, és ettől nő az önbizalmuk), illetve közvetetten a népszerűségen keresztül (a szépség növeli a népszerűséget, a népszerűség pedig az önbizalmat).

A korábban bemutatott útvonalkövetési szabályok segítségével a szépség és az önbizalom kovarianciáját a következőképp tudjuk felírni. Ha az önbizalomból indulunk visszafelé a β_1 nyílon, eljutunk a szépséghez, és ott lezárjuk az utat a VAR(X1)-gyel. De haladhatunk úgy is, hogy az önbizalomból a β_2 nyílon eljutunk a népszerűséghez, majd a kétirányú nyíllal átugrunk a szépséghez (COV(X1,X2)). Ebből következik, hogy

$COV(X1,Y) = \beta_1{}^*VAR(X1) + \beta_2{}^*COV(X1,X2)$.

Vagyis a szépség önbizalomra gyakorolt hatása két részre bontható, egy *közvetlen hatás*ra és egy *közvetett*, a népszerűségen keresztül érvényesülő hatásra. Amikor a többszörös regresszióban a szépség regressziós együtthatóját vizsgáljuk, akkor kizárólag a közvetlen hatást látjuk, és ezért mondjuk, hogy a szépség hatását a népszerűség kontrollálása mellett becsüljük. (A közvetett hatás explicit feltárása a következő lépésben, a mediációs modellekben történik majd meg.)

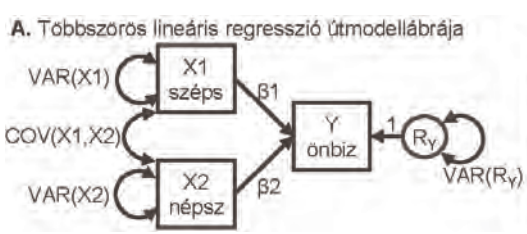
A példában szereplő lineáris regresszió egyenletrendszerének felírásához nézzük át a hozzá tartozó mért variancia-kovariancia mátrixot. A mátrixban szerepel a szépség, népszerűség és önbizalom varianciája, azaz VAR(X1), VAR(X2), VAR(Y), valamint a szépség-népszerűség, szépség-önbizalom és népszerűség-önbizalom kovarianciája, azaz COV(X1,X2), COV(X1,Y), COV(X2,Y). Ezek közül a VAR(X1), a VAR(X2) és a COV(X1,X2) exogén változókra vonatkoznak, ezért egyenleteik triviálisak lesznek. Így három független információ marad. Becsülni is három paramétert kell, a két regressziós együtthatót (β_1 és β_2), valamint azt, hogy az önbizalom varianciájának mekkora része megmagyarázott, és mekkora a reziduális rész.

A három egyenlet a következő (gyakorlásként kövessük végig a 17. ábrán!):

$COV(X1,Y) = \beta_1 * VAR(X1) + \beta_2 * COV(X1,X2)$
$COV(X2,Y) = \beta_2 * VAR(X2) + \beta_1 * COV(X1,X2)$
$VAR(Y) = \beta_1 * VAR(X1) * \beta_1 + \beta_2 * VAR(X2) * \beta_2 + \beta_1 * COV(X1,X2) * \beta_2 + \beta_2 * COV(X1,X2) * \beta_1 + 1 * VAR(R_Y) * 1$

Az első két egyenlet a kovarianciák felírását mutatja a már bemutatott módon. Az utolsót azonban érdemes külön végigkövetni az ábrán is. Az első négy tag az önbizalom megmagyarázott varianciáját adja, az utolsó tag pedig a meg nem magyarázott, reziduális rész.

Mivel ugyanannyi független információnk van, mint becsülendő paraméterünk, ezért ez a modell is szaturáltnak tekinthető, és egyértelműen megoldható.



B. Többszörös lineáris regresszió mért variancia-kovariancia mátrixa

	X1	X2	Y
X1	VAR(X1)		
X2	COV(X1,X2)	VAR(X2)	
Y	COV(X1,Y)	COV(X2,Y)	VAR(Y)

17. ábra: Többszörös lineáris regresszió modellje

A következő lépés a *mediációs elemzés*, amelyben a többszörös lineáris regresszióban impliciten meglévő indirekt útvonalat explicitté és tesztelhetővé tesszük. Ezzel további struktúrát adunk a változók közötti összefüggésrendszernek. Míg a lineáris regresszióban két szintet különböztettünk meg (a prediktorok magyarázták a kimeneti változót), a mediációban megjelenik egy harmadik, köztes szint is. Feltételezzük, hogy a prediktor hat a mediáló változó(k)ra, és a mediáló változó(k) közvetítik a hatást a kimeneti változóra. Más szóval, történetet adunk az összefüggésnek. Feltételezzük, hogy a szépség hat a népszerűségre és a népszerűség az önbizalomra.

A 18. ábra B. panelje alapján látjuk a megfigyelt variancia-kovariancia mátrix elemeit. Ezek a prediktor, mediáló változó és kimeneti változó varianciái, azaz VAR(X), VAR(M), VAR(Y), valamint a köztük lévő kovarianciák COV(X,M), COV(X,Y), COV(M,Y). Ez a készlet megegyezik a többszörös regressziónál látottal, azonban most csak a szépség (X) számít exogén változónak. Ennek megfelelően VAR(X) egyenlete triviális, a többi elem (VAR(M), VAR(Y), COV(X,M), COV(X,Y), COV(M,Y)) viszont mind olyan független információ, amelyet a modellnek kell megbecsülnie. Így a lineáris regresszióban látott három egyenlet helyett most öt egyenletet kell felírnunk. Kövessük őket ujjunkkal a 18. ábra A paneljén! (A β-értékek sorszámának szokatlan kiosztása csak azt a célt szolgálja, hogy az egyenletek könnyen összevethetők legyenek a lineáris regresszió egyenleteivel, más esetben a hagyományos balról jobbra történő számozás alkalmazható.) Az első egyenlet:

$$COV(X,M) = \beta_3 * VAR(X).$$

Fontos megjegyezni, hogy a népszerűségből (M) csak a szépség (X) irányába haladhatunk hátrafelé. Az önbizalom (Y) irányába nem, hiszen azzal megsértenénk a követési szabályt. A reziduális variancián keresztül sem tudunk eljutni a szépséghez, mert bár az irány hátrafelé mutat, ezen az úton soha nem érnénk el a prediktort.

Most nézzük a népszerűség és önbizalom kapcsolatát. Az önbizalomból kiindulva β_2-n haladunk a népszerűség felé, megérkezünk célunkhoz, M-hez, itt viszont azt vesszük észre, hogy nincs az ábrán kétirányú nyíl, amely lezárhatná utunkat. Azonban ez csupán azért van, mert a kétirányú nyíl, azaz VAR(M), fel van a modellben bontva egy X által megmagyarázott és egy reziduális részre. Éppen ezért a képletben a teljes varianciát, VAR(M)-et tüntetjük fel, még akkor is, ha az nem szerepel az ábrán.

Visszatérve a népszerűség és az önbizalom kapcsolatára, tehát az egyenlet első tagja az előzőekben végigkövetettnek megfelelően $\beta_2 * VAR(M)$. Van

azonban egy másik útvonal is a népszerűség és az önbizalom között. Ehhez elindulunk az önbizalomból β_1-en, megérkezünk a szépséghez, ott megfordulunk a szépség varianciájának kétirányú nyilán, és előrehaladunk β_3-on a népszerűségig. Mivel az útvonal már tartalmaz egy kétirányú nyilat, itt már nem kell még egy a lezáráshoz. Hasonlítsuk össze ezt a többszörös lineáris regressziónál látottal. Az egyenlet első fele (β_2*VAR(M)) azonos a két modellben, és a második fél is megfeleltethető egymásnak, főleg, ha emlékezünk, hogy a regresszióban látott COV(X1,X2) azonos a mediáció β_3*VAR(X)-ével. A végleges egyenlet:

$$COV(M,Y) = \beta_2*VAR(M) + \beta_1*VAR(X)*\beta_3.$$

Végül nézzük meg a szépség és az önbizalom kapcsolatát. Itt a kovariancia szétbomlik egy direkt és egy indirekt, népszerűségen keresztül megvalósuló ágra: $COV(X,Y) = \beta_1*VAR(X) + \beta_2*\beta_3*VAR(X)$, ahol β_1 a direkt hatás, $\beta_2*\beta_3$ pedig az indirekt hatás.

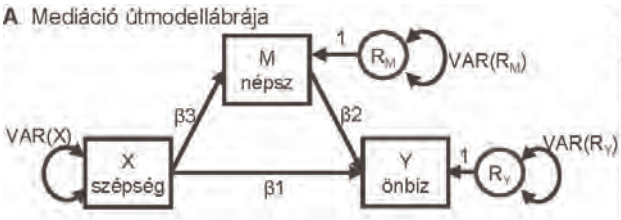
A kovarianciák után lássuk a varianciák egyenleteit. A prediktor varianciájának egyenlete továbbra is triviális VAR(X) = VAR(X). A mediáló változó varianciájának egyenlete a következő:

$$VAR(M) = \beta_3*VAR(X)*\beta_3 + 1*VAR(R_M)*1.$$

Az, hogy a népszerűségből további út vezet az önbizalomhoz, itt nem zavar bennünket, hiszen ez az összefüggés nem a népszerűség varianciáját magyarázza. Végül a legtöbb tagból álló egyenlet a kimeneti változó varianciájához tartozik (a képlet hosszú, de logikus, ha végigkövetjük az ábrán):

$$VAR(Y) = \beta_2*\beta_3*VAR(X)*\beta_3*\beta_2 + \beta_1*VAR(X)*\beta_1 + \beta_2*\beta_3*VAR(X)*\beta_1 + \beta_1*VAR(X)*\beta_3*\beta_2 + \beta_2*1*VAR(R_M)*1*\beta_2 + 1*VAR(R_Y)*1.$$

Érdemes még egyszer összevetni a többszörös lineáris regresszió és a mediáció képleteit, hogy lássuk, milyen mértékben hasonlítanak egymásra. A fő különbség az, hogy míg a lineáris regresszióban a szépség és a népszerűség közötti összefüggést pusztán statisztikai kovarianciaként kezeltük, a mediációban az összefüggés irányát elméleti feltételezéssel rögzítjük. Ennek köszönhetően a feltételezett indirekt útvonal számszerűsíthetővé válik, és a modell keretein belül statisztikai szignifikanciája (például bootstrap eljárással) tesztelhető.



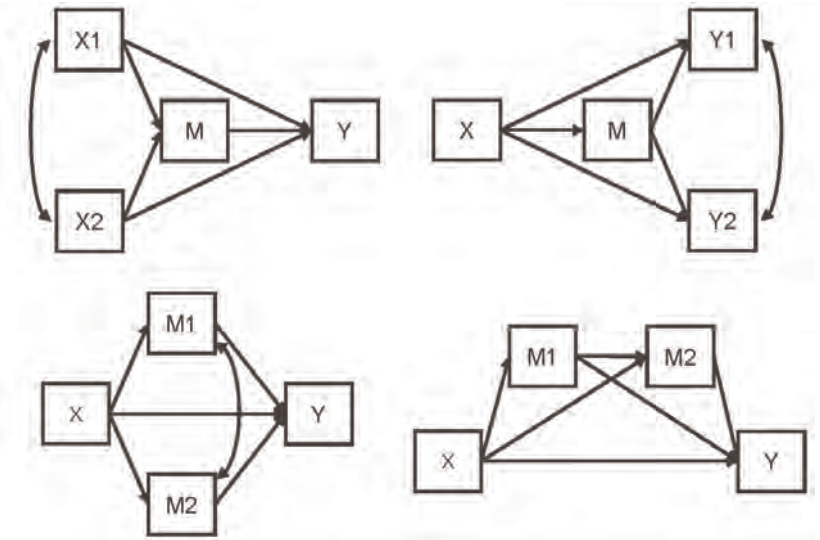
B. Mediáció mért variancia-kovariancia mátrixa

	X	M	Y
X	VAR(X)		
M	COV(X,M)	VAR(M)	
Y	COV(X,Y)	COV(M,Y)	VAR(Y)

18. ábra: A mediáció modellje

Innentől a modell komplexitásának csak a képzelet szab határt – persze észben tartva a parszimónia elvét és az olyan korlátozó tényezőket, mint az elemszám. Lehet egy modellben több prediktor, több kimeneti vagy mediáló változó. Ez utóbbiak lehetnek párhuzamosak vagy szeriálisak. Illetve a modell állhat ezek tetszőleges kombinációjából. A 19. ábrán láthatunk néhány példát. Ami közös ezekben a modellekben, hogy mindegyikben minden lehetséges egy- és kétirányú nyíl be van húzva, azaz ezek a modellek mind szaturáltak. Amíg szaturált modelleknél maradunk, addig – legyenek akármilyen összetettek – a mediációs modellek módszertanán belül maradunk, amelyek az útvonalmodellek egy speciális alcsoportját képezik.

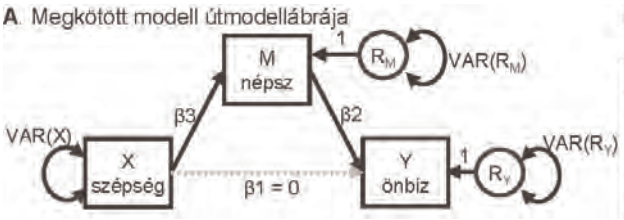
Érdekességként érdemes megjegyezni, hogy amíg a szaturált modelleken belül vagyunk, addig nem számít milyen módszerrel számoljuk ki a paramétereket (hiszen egyértelmű megoldása van az egyenletrendszernek), választhatunk bármilyen paraméterbecslő eljárást, de akár alkalmazhatjuk a hagyományos módszertant is, amely regressziók sorozatából építette fel a modellt (erre példa a mediációs szakirodalomban rendkívül elterjedt PROCESS-eljárás is). Mindez igaz a regressziós együtthatókra, de a hozzájuk tartozó standard errorokra, és ennek megfelelően a szignifikanciaértékekre már nem.



19. ábra: Komplexebb szaturált mediációs modellek

Amikor a szaturált mediációs modell után továbblépünk, a következő lépcső az, hogy elméleti megfontolások alapján bizonyos útvonalakat elhanyagolhatónak tekintünk. Ez azt jelenti, hogy ezeknek a hatásoknak az értékét nem szabadon becsüljük, hanem nullára állítjuk. Ezzel tulajdonképpen megkötéseket vezetünk be a modellbe, a szabad paraméterek számát csökkentjük, és így létrejönnek a *korlátozott* vagy *megkötött modellek* (*restricted models*). Korlátozásra példa, ha egy közvetlen útvonalat (X → Y) nullára állítunk, mert az elmélet szerint a hatás teljes egészében közvetett. Például feltételezhetjük, hogy a szépség csak a népszerűségen keresztül hat az önbizalomra, a közvetlen útvonal nélkül. Más szóval: hiába nézegetjük magunkat reggelente a tükörben, ez önmagában nincs hatással az önbizalmunkra.

Az így kialakuló modell (20. ábra) már statisztikailag tesztelhető, mert nem minden útvonalat engedünk szabadon illeszkedni az adatokhoz, mint a szaturált modellben. A megkötések miatt a modell illeszkedésének lesz egy bizonyos hibája, így lehetőség nyílik arra, hogy összevessük a modell és az empirikus adatok közötti eltérést. Ez adja a strukturális egyenletmodellezés egyik legfontosabb lényegét: azt, hogy megvizsgálhatjuk, hogy az általunk feltételezett, elméletileg indokolt egyszerűbb modell mennyire képes visszaadni a valóságot, és így eldönthetjük, hogy az elméleti megkötéseink tarthatók-e az adatok fényében.



B. Megkötött modell mért variancia-kovariancia mátrixa

	X	M	Y
X	VAR(X)		
M	COV(X,M)	VAR(M)	
Y	COV(X,Y)	COV(M,Y)	VAR(Y)

20. ábra: Megkötést tartalmazó modell ábrája

2.2. A modell azonosíthatósága (model identification)

Mielőtt részletesen bemutatnánk a megkötött modelleket, időzzünk még el egy kicsit a szaturált mediációs modelleknél. Mint már átbeszéltük, a szaturáltság azt jelenti, hogy a modellben minden lehetséges kapcsolat (útvonal) be van húzva a változók között. Matematikai szempontból ez azt eredményezi, hogy a rendelkezésre álló független információk száma pontosan megegyezik a becsülendő paraméterek számával. Másként fogalmazva, annyi egyenlet áll rendelkezésre, ahány ismeretlen van a modellben, így a megoldás egyértelmű.

Ennek kapcsán szokás bevezetni a *szabadságfok* (df, degrees of freedom) fogalmát, amelyet a szabad információk és a becsült paraméterek különbségeként határozunk meg. A szabadságfok tehát azt mutatja meg, mennyi pluszinformáció marad a modell ellenőrzésére. Amikor egy modell szaturált, akkor df = 0, vagyis nincs maradék szabadságfok, így nincs lehetőség az illeszkedés vizsgálatára sem, mert a modell tökéletesen reprodukálja az adatok kovarianciamátrixát. Az ilyen modellt nevezzük *pontosan azonosított* vagy *éppen azonosított* (just identified) modellnek, amelynek legfontosabb jellemzője, hogy bár mindig hibátlan illeszkedést mutat, ez önmagában nem szolgálhat az elméleti modell helyességének bizonyítékául.

Innen két irányba mozdulhatunk el. Az egyik eset az, amikor olyan modellt építünk, amelyben több a becsülendő paraméter, mint a független információ.

Ez előfordulhat például akkor, ha egy már meglévő egyirányú nyíl visszafelé is be van húzva (például nemcsak a népszerűség → önbizalom, hanem az önbizalom → népszerűség utat is berajzoljuk). Ilyenkor a szabadságfok negatívvá válik (df < 0), ami azt jelenti, hogy több az ismeretlen, mint az egyenlet. Az ilyen egyenletrendszernek végtelen sok megoldása lenne, ezért a modell matematikailag nem meghatározott. Ezt nevezzük *alulazonosított modell*nek (underidentified model), és ilyen esetben útvonalmodellel nem lehet tesztelést végezni.

Itt említjük meg a nem azonosítható modellek egy másik tipikus formáját, azt, amikor a nyilak zárt kört alkotnak, tehát egy változótól közvetve visszajutunk ugyanahhoz a változóhoz (A → B → C → A). Ez a körkörös oksági szerkezet hagyományos útvonalmodellekben nem tesztelhető, mivel az egyenletek rendszerében ilyenkor megszűnik az egyirányú oksági irány kijelölhetősége, és az azonosítási feltételek sérülnek. Az ilyen köröket csak speciális eljárásokkal, például dinamikus modellezési megközelítésekkel vagy visszacsatolásokat is kezelni képes módszerekkel lehet vizsgálni.

Most lássuk a másik irányt! Amikor egy modellben több független információ áll rendelkezésünkre, mint amennyi paramétert becsülni szeretnénk, vagyis több egyenletünk van, mint ismeretlenünk, az ilyen modellt *túlhatározottnak* (overidentified) nevezzük. Ez tipikusan úgy érhető el, hogy bizonyos útvonalakat rögzített értékre állítunk. Bár elvileg bármilyen számra állíthatnánk, a gyakorlatban majdnem mindig nullát választunk. Ez azt jelenti, hogy az adott útvonalat nem létezőnek/elhanyagolhatónak tekintjük. A modellben szereplő egyenletekben az adott kapcsolat helyet kap, de nullás súllyal, így az ábrán rendszerint fel sem tüntetjük. A korábban használt (szépség–népszerűség–önbizalom összefüggésére vonatkozó) példában a megkötött modell esetében azt feltételezzük, hogy a szépség hatása teljes egészében közvetetten a népszerűségen keresztül valósul meg az önbizalomra, azaz a szépség önbizalomra tett közvetlen hatása elhanyagolható, az egyenletben nullának tekintjük. A 20. ábra ezt a modellt szemlélteti, és a következő egyenletekkel írható le:

A független és a mediátor kovarianciája:
$COV(X,M) = \beta_3 * VAR(X)$. Ez megegyezik a korábbi mediációs modellel.
A mediátor és a kimeneti változó kovarianciája:
$COV(M,Y) = \beta_2 * VAR(M) + \beta_1 * VAR(X) * \beta_3$, de mivel a közvetlen hatást (β_1) nullára állítottuk, egyszerűsödik:
$COV(M,Y) = \beta_2 * VAR(M)$.

A független és a kimeneti változó kovarianciája:
$COV(X,Y) = \beta_2*\beta_3*VAR(X) + \beta_1*VAR(X)$, de β_1 itt is nulla, tehát:
$COV(X,Y) = \beta_2*\beta_3*VAR(X)$.

Ez azt jelenti, hogy három különböző egyenletünk van, miközben mindössze két ismeretlen paramétert (β_3 és β_2) szeretnénk becsülni.

Ugyanez történik a varianciák egyenleteiben is, itt is kiesik minden β_1-et tartalmazó tag.

VAR(X)=VAR(X) továbbra is triviális.
$VAR(M) = \beta_3*VAR(X)*\beta_3 + 1*VAR(R_M)*1$. Az egyenlet nem változik, hiszen β_1 megkötése nem érinti a mediáló változót, mert β_1 a kimeneti változót magyarázó paraméter. Viszont a kimeneti változó varianciáját felíró képlet jelentős mértékben rövidül:
$VAR(Y) = \beta_2*\beta_3*VAR(X)*\beta_2*\beta_3 + \beta_2*1*VAR(R_M)*1*\beta_2 + 1*VAR(R_Y)*1$.

A kieső tagok mind a kimeneti változó varianciájának modell által megmagyarázott részéhez tartoznak, míg a kimeneti változó reziduálisához tartozó rész természetesen érintetlen. Ennek megfelelően a modell megkötéseivel mindig csökken a megmagyarázott varianciarész aránya a meg nem magyarázotthoz képest. A strukturális egyenletmodellezés egyik kulcsgondolata viszont éppen az, hogy ha azok az útvonalak, amelyeket nullára rögzítünk, a valóságban is jelentéktelenek, akkor a megmagyarázott változatosság csökkenése is jelentéktelen lesz.

Összeszámolva azt kapjuk, hogy több az egyenlet, mint a becsült paraméter, ezért az így kapott egyenletrendszernek nincs pontos megoldása. Ezért mondjuk, hogy a modell túlhatározott.

2.3. Paraméterbecslés (parameter estimation)

Probléma ez? Természetesen nem. A strukturális modellezés esetében a túlazonosítottság nem hátrány, hanem a modelltesztelés alapja. Példánkhoz tekintsük a 20. ábrát. Legyenek a megfigyelt varianciák és kovarianciák:

COV(X,M) = 5
COV(M,Y) = 6
COV(X,Y) = 8

VAR(X) = 5
VAR(M) = 2
VAR(Y) = 10

Ezek adják a megfigyelt variancia-kovariancia mátrixot. Az egyszerűség kedvéért tekintsünk el a varianciák egyenleteitől, és koncentráljunk csak a kovarianciákra. A modell által feltételezett összefüggések:
$COV(X,M) = \beta_3 * VAR(X)$
$COV(M,Y) = \beta_2 * VAR(M)$
$COV(X,Y) = \beta_2 * \beta_3 * VAR(X)$

A mért variancia-kovariancia mátrix mellé felírható a modell által feltételezett értékeket tartalmazó mátrix is, és számítható a kettő eltérése. Ha például felvesszük $\beta_3 = 1$-et, akkor a modell szerinti feltételezett kovariancia X és M között 1*5 = 5 lesz, ami pontosan megegyezik a mért értékkel. Ha $\beta_2 = 3$, akkor az M-Y kovarianciát 3*2 = 6-nak feltételezi a modell, ami szintén helyes, de az X-Y kovarianciát kényszerűen 3*1*5 = 15-nek számítja, ami azonban 7 egységgel több a megfigyelt 8-as kovarianciaértéknél. Ezzel a paraméterbeállítással az összhiba a modell becslésében 0+0+7=7.

Ezzel szemben a $\beta_3 = 1$, $\beta_2 = 2$ paraméterbeállítás csak 0 + 2 + 2 = 4 összhibához vezet, ami még mindig nem tökéletes illeszkedés, de már közelebb áll a mért értékekhez, mint az első beállítás.

Látható, hogy a modell által implikált kovarianciák nem tudják tökéletesen visszaadni a megfigyelt értékeket. A különböző paraméterkiosztások azonban más és más mértékű eltérést eredményeznek a megfigyelt és a modell által implikált kovarianciák között. Bármely paraméterbeállításnál kiszámítható a teljes eltérés (össztévedés), és a cél az, hogy ezt minimalizáljuk. Összegezve, amikor a modellben bizonyos paramétereket megkötünk (értéküket nullára állítjuk), akkor a legtöbb esetben nem érhető el tökéletes illeszkedés, mert a modell szerkezete önmagában túl szűk ahhoz, hogy minden megfigyelt kapcsolatot visszaadjon. A paraméterbecslés célja csupán az eltérés minimalizálása.

A paraméterpróbálgatást természetesen nem a kutatónak kell kézzel elvégeznie, az útvonalelemzés *paraméterbecslő algoritmusai* éppen ezt a feladatot látják el. A *paraméterbecslés* (parameter estimation) a modellezés egyik központi lépése. Ekkor a modell paramétereit (útvonal-együtthatók, varianciák, kovarianciák, reziduálisok stb.) olyan módon határozza meg az algoritmus, hogy a modell által feltételezett variancia-kovariancia mátrix a lehető legjobban közelítse a megfigyelt variancia-kovariancia mátrixot. A cél minden eljárás esetében az eltérés minimalizálása, a módszerek közötti különbség pedig abban rejlik, hogy

milyen veszteségfüggvényt használnak, illetve milyen eloszlási feltételezésekkel élnek az adatokra vonatkozóan.

Az útvonalelemzés során a leggyakrabban használt becslési módszer a *Maximum Likelihood* (ML). Ez az eljárás aszimptotikusan torzítatlan, hatékony és konzisztens, valamint szinte minden szoftverben elérhető, így a PA gyakorlatában is alapértelmezettnek tekinthető. Az ML-becslés célja (hasonlóan más becslő eljárásokhoz), hogy a modell által implikált variancia-kovariancia mátrix a lehető legjobban illeszkedjen a megfigyelt mátrixhoz.

A PA esetében külön figyelmet kell fordítani az adatok mérési szintjére. Az ML ugyanis feltételezi, hogy az endogén változók (azok, amelyekre más változók hatással vannak a modellben) folytonos, skálaszintű változóknak tekinthetők. A legtöbb kérdőíves kutatásban ugyan elemi szinten Likert-típusú itemek szerepelnek, de ezeket rendszerint összegezzük vagy átlagoljuk, így olyan kompozit mutatókat (skálákat) hozunk létre, amelyek már kezelhetők folytonos változóként. Így a PA-ban az ML alkalmazása a gyakorlatban jól megállja a helyét.

Fontos ugyanakkor, ha egy változó ordinális szinten kerül a modellbe (például amikor az észlelt egészségi állapotot egyetlen ötfokú Likert-tétellel mérik), az gátja lesz az ML alkalmazásának. Ilyenkor, illetve akkor is, ha a skálákon nem feltételezhető a normál eloszlás, más paraméterbecslő eljárás választása javasolt. Jó választás lehet például a *Diagonally Weighted Least Squares* (DWLS).

Érdemes észben tartani, hogy az exogén változókra (a modell független prediktoraira) nincs ilyen szigorú feltétel, azok lehetnek ordinális változók vagy akár nominálisak is, ha a kategóriákat előzetesen dichotóm indikátorokra bontjuk.

Mindezekből következik, hogy a kutatás tervezési fázisában tudatosan kell gondolkodni a későbbi becslésről. Ha például az endogén változók mérését úgy alakítjuk ki, hogy azok skálaként értelmezhetők legyenek, akkor az ML használata nemcsak lehetséges, hanem kifejezetten ajánlott. Így biztosítható, hogy a modell illeszkedése megbízhatóan tükrözze az adatok szerkezetét, és ne kényszerüljünk alternatív, kevésbé elterjedt becslési eljárásokhoz folyamodni.

Mindez nemcsak az útvonalelemzésre igaz, a paraméterbecslés ugyanilyen alapvető eleme a teljes strukturális egyenletmodellezésnek is. A strukturális egyenletmodellezés két fő részből áll, a mérési modellből és a strukturális modellből. A mérési modell a látens változók és a megfigyelt indikátorok kapcsolatát írja le, ami erősen emlékeztet a megerősítő faktorelemzés (CFA) eljárására. Ezzel szemben a strukturális modell a változók közötti összefüggéseket ragadja meg, ami párhuzamba állítható az útvonalelemzés (PA) logikájával. Bár első ránézésre úgy tűnhet, hogy a két módszer teljesen különböző feladatot lát el, matematikai szempontból mindkettő ugyanarra az egyenletrendszerre vezethető

vissza. Emiatt a paraméterbecslés technikai részletei megegyeznek, további információért érdemes áttekinteni a CFA-ról szóló fejezet 2.3-as pontját is.

2.4. A modell értékelése (model evaluation)

Egy megkötött modell esetében a gyakorlatban soha nem érhető el tökéletes illeszkedés, a modell által feltételezett variancia-kovariancia mátrix mindig bizonyos mértékben eltér a megfigyelt mátrixtól. A modell értékelése éppen ezen eltérés nagyságának vizsgálatán alapul. Minél kisebb az eltérés, annál jobban illeszkedik a modell az adatokhoz, ami a valóság pontosabb, ugyanakkor parszimónikusabb (gazdaságosabb) leírását teszi lehetővé. Az eltérés mérésére az *illeszkedési mutatók* szolgálnak.

Az illeszkedés vizsgálatának alapmutatója a *khi-négyzet-statisztika* (χ^2), amely a megfigyelt és a modell által implikált mátrix közötti eltérést teszteli. Bár a χ^2 a hagyományos kiindulópont, nagy minták esetén túlérzékennyé válik, és szinte mindig szignifikáns eltérést jelez, ezért önmagában ritkán tekinthető döntő mutatónak. Emiatt javasolt a szabadságfokhoz viszonyított változatát (χ^2/df) is figyelembe venni, valamint egyre inkább előtérbe kerültek a további illeszkedési mutatók. A *komparatív mutatók* közül a *CFI* (Comparative Fit Index), a *TLI* (Tucker–Lewis Index) és az *IFI* (Incremental Fit Index) a nullmodellhez képest mutatják a javulás mértékét. Az *abszolút illeszkedést* az *RMSEA* (Root Mean Square Error of Approximation) méri, míg a reziduális alapú *SRMR* (Standardized Root Mean Square Residual) a megfigyelt és becsült kovarianciák átlagos eltérését számszerűsíti.

Fontos hangsúlyozni, hogy ezek az illeszkedési mutatók azonos logika szerint működnek függetlenül attól, hogy konfirmatív faktoranalízisben (CFA), útvonalelemzésben (PA) vagy teljes strukturális egyenletmodellezésben (SEM) használjuk őket. Mivel részletes bemutatásuk már a CFA-fejezet 2.4-es részében megtörtént, az olvasót arra biztatjuk, hogy ott tekintse át a mutatók értelmezésének részleteit. Itt csupán röviden jelezzük a leggyakrabban hivatkozott elfogadható értékeket.

Illeszkedési mutató	Irány	Elfogadható illeszkedés	Jó illeszkedés	Tökéletes illeszkedés
χ^2	kisebb szám jobb	-	nem szignifikáns	0
χ^2/df	kisebb szám jobb	< 3	< 2	0
CFI, TLI, IFI	nagyobb szám jobb	> 0,90	> 0,95	1
RMSEA	kisebb szám jobb	< 0,08	< 0,05	0
RMSEA 90% ULCI	kisebb szám jobb	-	< 0,10	0
SRMR	kisebb szám jobb	< 0,08	< 0,05	0

19. táblázat: Illeszkedési mutatók határértékei

2.5. A modell újraspecifikációja (model respecification)

Nem szabad elkeseredni, ha az első felírt modellünk nem illeszkedik megfelelően, sőt az első modell szinte soha nem illeszkedik elég jól. Ez teljesen természetes, hiszen a modellépítés egy iteratív folyamat, az elmélet és az adatok közötti párbeszéd. Ilyenkor szabad és kell is változtatni a modellen, de nagyon fontos három alapelvet szem előtt tartani. Az első, hogy mindig csak olyan változtatásokat tegyünk, amelyek elméletileg megalapozottak, soha ne csupán addig csiszolgassuk a modellt, amíg statisztikailag szép lesz. A második, hogy nem szabad túlzásba esni, megfelelő illeszkedés elérését követően ne változtassunk a még jobb illeszkedés elérése érdekében, mert az könnyen túlillesztéshez vezet. Az utolsó alapelv, hogy tegyük a folyamatot teljesen transzparenssé, és mindig pontosan dokumentáljuk, min változtattunk, miért tettük, milyen elméleti megfontolásból volt a változtatás megengedhető, és milyen eredménnyel járt. Ez biztosítja, hogy a modellépítés folyamatát az olvasó vagy bíráló is követni tudja. Ha így járunk el, akkor a változatok nem öncélú próbálkozások lesznek, hanem arra szolgálnak, hogy közelebb jussunk egy olyan modellhez, amely egyszerre elméletileg megalapozott és statisztikailag kielégítően illeszkedik.

A változtatásoknak két fő formája van. Az egyik lehetőség, hogy teljesen új struktúrát próbálunk ki, vagyis az előzőbe *be nem ágyazott* (non-nested) modellt állítunk fel. Ebben az esetben az új modell nem tekinthető az előző korlátozottabb/lazább változatának, hanem attól független, alternatív struktúrát képvisel. Például ha korábban azt feltételeztük, hogy a szépség a népszerűségen keresztül hat az önbizalomra, de most egy teljesen más teóriát követve az

önbizalmat tekintjük a népszerűség előzményének, akkor két egymásba be nem ágyazott modellt teszteltünk. Ennek egy speciális esete, amikor a struktúrát alapvetően megtartjuk, csak egy-egy változót veszünk ki vagy teszünk be a modellbe.

Több oka lehet annak, hogy bizonyos változókat elhagyjunk vagy éppen újakat beemeljünk. Indokolhatja a módosítást az, hogy a korábbi modellben szereplő változó nem mutatott szignifikáns vagy értelmezhető kapcsolatot, így megtartása inkább zajt, mintsem hasznos információt adna (de felhívjuk a figyelmet arra, hogy semmiképp sem kötelező a nem szignifikáns kapcsolatokkal rendelkező változókat törölni a modellből). Ugyanígy előfordulhat, hogy a modell illeszkedési mutatói vagy a modifikációs indexek arra utalnak, hogy egy új kapcsolat vagy rejtett tényező bevonása javítaná a modell magyarázóerejét. Az is lehetséges, hogy a kutató a módosítás során egy finomabb, a szakirodalomból vagy az adatokból újonnan kirajzolódó elméleti megfontolást követ, és ennek érdekében módosítja a modellt. Ilyen esetekben a módosítás mindig mérlegelés eredménye, ahol az empirikus adatok és az elméleti megalapozottság együtt indokolják a változtatást.

Fontos tudni, hogy ha módosítjuk a modell változókészletét, azzal nemcsak a becsült paraméterek, de a szabad információk száma is változik. Új változó bevonása új varianciát és kovarianciákat hoz, tehát növeli a független információk számát, változik a modell összetettsége. Éppen ezért nehéz általános szabályt megfogalmazni arra, hogy egy változó betétele vagy kivétele miként fog hatni az illeszkedésre, javítja vagy rontja azt. Továbbá ilyenkor a modell értékeléséhez a hagyományos illeszkedési mutatók (például CFI, TLI, RMSEA, SRMR) nem elegendők, hanem mellettük *információs kritériumok (AIC, BIC*) használata indokolt, amelyek a modell illeszkedését és a komplexitását együtt értékelik. Mindkét mutatónál a kisebb érték jobb egyensúlyt jelez az adatokhoz való illeszkedés és a parszimónia között.

A másik modellmódosítási lehetőség a beágyazott változtatás. A *beágyazott modell (nested model)* lényege, hogy a kisebb, egyszerűbb modell a nagyobb, komplexebb modell egy speciális esete. Úgy kell elképzelni, mint egy matrjoska babát, a kisebb baba be van ágyazva a nagyobba. A kisebb, korlátozottabb modell (ahol az útvonal együtthatója = 0) van beágyazva a nagyobb, szabadabb modellbe (ahol az útvonal együtthatóját a modell szabadon becsülheti). Technikailag ez azt jelenti, hogy az egyszerűbb modellt megkaphatjuk a komplexebből csupán egy vagy több paraméterének megkötésével (például nullára való megkötésével). Ennek megfelelően a beágyazott átalakítás során az előző struktúrát megtartjuk, a változók halmazán és összefüggésrendszerén alapvetően nem változtatunk, csak néhány útvonalat veszünk ki vagy teszünk hozzá.

A beágyazott módosítások a független információk számát nem, csak a becsülendő paraméterek számát változtatják meg. Ha egy útvonalat elhagyunk, akkor korlátozottabbá, szigorúbbá válik a modell, míg egy útvonal betétele növeli a modell szabadságát, lazábbá teszi azt. Ez utóbbi esetben a khi-négyzet-érték csökken (mert a szabadabb, rugalmasabb rendszerben javul az illeszkedés), a szabadságfok viszont csökken (azaz kevésbé parszimónikus a modell). Mivel a kiinduló modellhez képest jobb illeszkedést várhatunk, a legtöbb relatív illeszkedési mutató jellemzően nő az útvonal betételével. A CFI és TLI közül az utóbbi fog kevésbé változni, mivel ez az illeszkedés javulása mellett jobban figyelembe veszi a parszimónia csökkenését is. Az RMSEA kifejezetten érzékeny a szabadságfokra. Ha egy útvonal betétele után a khi-négyzet arányosan nem csökken annyit, mint a szabadságfok, akkor az RMSEA akár nőhet is (rosszabb illeszkedést jelezhet), tehát előfordul, hogy egy lazább modell rosszabb RMSEA-értéket produkál. Viszont ha olyan útvonalat tettünk be, melyre a modellnek „nagy szüksége volt", akkor az RMSEA értéke csökkenni fog. Az SRMR inkább a reziduumokra érzékeny, így a szabadságfok változását kevésbé jutalmazza vagy bünteti. Emiatt a lazább modellek majdnem mindig alacsonyabb SRMR-értéket (jobb illeszkedést) eredményeznek.

A beágyazott modellek közötti különbségek a hagyományos illeszkedési mutatók mellett khi-négyzet-különbségteszttel is tesztelhetők, ellenőrizve, hogy a lazább modell szignifikánsan jobb illeszkedést ad-e.

Érdemes külön megemlíteni a reziduális kovariancia betételét és kivételét a modellből, ami azt jelenti, hogy két reziduális között engedünk vagy tiltunk meg kapcsolatot. Reziduális kovarianciát általában csak azonos szintű, hasonló tartalmú változók között szokás engedni. A reziduális kovarianciák bevonása mindig az illeszkedést javítja, de könnyen túlillesztéshez vezethet, ezért különösen indokolt elméleti alap kell hozzá. Erre példa a párhuzamos mediáló változók reziduálisai közötti kovariancia. Például a népszerűséget úgy is lekérdezhettük volna, hogy külön rákérdezünk a népszerűségre az azonos és ellenkező nemű személyek körében. Ekkor két párhuzamos mediáló változót használva tesztelni tudjuk, hogy a szépség az azonos vagy ellenkező nemű ismerősök körében való népszerűségen keresztül hat inkább az önbizalomra. Feltételezhető, hogy a két népszerűségváltozó együtt jár, de a modellben ezt nem kívánjuk tesztelni. Viszont ha nem engedünk meg közöttük valamiféle kapcsolatot, akkor felesleges illeszkedési problémát okoz majd a két változó összefüggése, amit a modellünk csak a szépséggel tudna magyarázni. Ilyenkor megoldás a két változó reziduálisa közötti kovariancia megengedése.

A változtatások az alapján is csoportosíthatók, hogy mi a mögöttes inspiráció, ez alapján a változtatás lehet elmélet-, illetve adatalapú.

Az egyik legstabilabb út az elméletalapú újraspecifikáció. Ha többféle elméleti magyarázat létezik ugyanarra a jelenségre, akkor kipróbálhatjuk ezek alternatív modelljeit. Ilyenkor a statisztikai illeszkedés csak megerősít vagy cáfol, de a kiindulópont mindig az elmélet. Ez a legbiztonságosabb módszer, mert így elkerüljük, hogy a modellt túlzottan hozzáigazítsuk az adatokhoz.

Az adatalapú változtatás központi eszközei a *modifikációs indexek* (MI). A modifikációs index azt mutatja meg, hogy mennyivel csökkenne a modell khinégyzet-értéke, ha egy eddig nullára megkötött paramétert felszabadítanánk, szabadon engednénk. Az MI tehát előre jelzi, hogy egy adott változtatás mennyire javítaná az illeszkedést. Ez egy nagyon hasznos diagnosztikai eszköz, de óvatosan kell vele bánni. Az algoritmus ugyanis minden lehetséges paramétert végigiterál, így gyakran teljesen értelmetlen vagy elméletileg abszurd változtatásokat is felajánl. Például egy modellben A → B → C az út, de A és C között erős korreláció van. Az MI jogosan felajánlja A → C-t, de ugyanúgy felajánlja B → A-t és C → B-t is (úgy „okoskodva", hogy ha valahogy megerősíthetnénk az indirekt utat, az is megoldhatná az A és C kapcsolatának magyarázatában érezhető hiányt). Ez utóbbi változtatások nyilvánvalóan elméletileg értelmetlenek. Ezért a modifikációs indexeket nem szabad vakon követni, és csak azokat a változtatásokat szabad megengedni, amelyek szakmailag is értelmezhetők. Az MI tehát nem döntéshozó, hanem inspiráló, diagnosztikai eszköz. Megmutatja, hol lehet gond a modell illeszkedésével, de a végső döntést mindig az elméleti keret alapján kell meghozni.

Fontos szabály, hogy az újraspecifikáció során egyszerre csak egy változtatást engedjünk meg. Ennek oka az, hogy egyetlen új útvonal is megoldhat olyan feszültségeket a modellben, amelyek több ponton is rossz illeszkedést okoznak. Például egy modellben legyen az A → B → C → D a feltételezett útvonal. Az A és D között erős az összefüggés, de ez a közvetlen útvonal nincs megengedve, így az összefüggés csupán B és C közvetítésén keresztül tud megvalósulni. Tegyük fel azonban, hogy itt gyengék az összefüggések. Ekkor azt látjuk majd, hogy a modell nem tudja magyarázni A és D erős összefüggését. Az MI több javaslatot tesz ezen probléma orvoslására. A javaslatok között ott lesz a valós megoldás is, a direkt útvonal megengedése, de felajánl olyanokat is, amelyek az indirekt útvonalat erősítik, például azt, hogy húzzuk be az A → C és B → D útvonalat is. Ezek ugyan nem nonszensz megoldások, mint az előző esetben, de feleslegesek, mert ha megengedjük az A → D közvetlen utat, az egész lánc feszülése megszűnik, és a többi paraméter is közelebb kerül a valós értékekhez. Ha több változtatást engedünk egyszerre, nem látszik, melyik hozta a valódi javulást.

2.6. A modell értelmezése (model interpretation)

Amikor a modell megfelelő illeszkedését sikerült elérnünk, a következő lépés a modell értelmezése. Ez a lépés foglalja magában a regressziós együtthatók áttekintését. Ellenőrizzük, melyek bizonyulnak szignifikánsnak, milyen erős az adott összefüggés, és az elméletnek megfelelő irányban jelentkezik-e a kapcsolat. Fontos átnézni a reziduális kovarianciákat is, amennyiben volt olyan, amelynek szabadon hagyását megengedtük, hiszen ezek további információt nyújthatnak a modell értelmezéséhez. Szerepet kap továbbá az R^2-értékek vizsgálata, amelyek megmutatják, hogy az egyes endogén változók varianciájának mekkora részét magyarázza a modell. Az útvonalelemzés igazi sajátossága azonban a mediációk számszerűsítése, vagyis annak feltérképezése, hogy egy változó milyen közvetítő mechanizmusokon keresztül hat egy másikra.

A *mediációk* lényege, hogy több regressziós együttható láncolataként jönnek létre indirekt útvonalak. Ezek az útvonalak nem új, külön becsülendő paraméterek a modellben, hanem a meglévő regressziós együtthatók szorzataként számítható mutatók. Emiatt az, hogy hány indirekt hatást számolunk ki, semmilyen módon nem befolyásolja a modell illeszkedését, mert azt változatlan paraméterkészlet határozza meg, az indirekt hatások pedig csupán ezekből levezetett értékek.

A számítás maga egyszerű. Az indirekt hatás az útvonalat alkotó regressziós együtthatók szorzatából adódik. Például ha A hat B-re, B pedig C-re, akkor az A → B → C útvonal indirekt hatása az A–B és B–C együtthatók szorzataként adódik. Így akár kézzel, egy számológép segítségével is kiszámolhatnánk az értékeket, ha a becsült együtthatók rendelkezésre állnak.

Ami miatt azonban az útvonalelemzés mégis szükséges, az az, hogy az indirekt hatásokhoz nemcsak pontbecslést, hanem standard hibát, konfidenciaintervallumot és szignifikanciaértéket is szeretnénk kapni. Erre többféle módszer létezik. A hagyományos megközelítés a delta-módszer, amely azonban gyakran pontatlan vagy túlzottan optimista becsléseket ad, különösen kisebb mintákban. Ezért váltak elterjedtté a robusztusabb eljárások, mindenekelőtt a bootstrap, amely újra-mintavételezéssel ad megbízhatóbb standard hibákat és intervallumokat az indirekt hatásokhoz.

Összességében az útvonalelemzésben a közvetlen (direkt) és a közvetett (indirekt) hatások együtt adnak teljes képet a változók közötti összefüggésekről. A direkt hatások megmutatják az azonnali kapcsolatokat, míg az indirekt hatások feltárják azokat a köztes mechanizmusokat, amelyek révén egy változó hatást gyakorolhat egy másikra. A kettő együtt teszi lehetővé, hogy a modellezés

ne csak kapcsolatok kimutatására, hanem a mögöttes folyamatok megértésére is alkalmas legyen.

20. doboz: Az útvonalelemzés (Path Analysis) alapvető definíciói és modelltípusai

Útvonalelemzés (path analysis, PA): olyan statisztikai eljárás, amely a változók közötti direkt és indirekt kapcsolatok struktúráját vizsgálja.
Strukturális modell (structural model): a modell azon része, amely a látens és/vagy mérhető változók közötti oksági (irányított) kapcsolatokat tartalmazza.
Modellspecifikáció (model specification): annak pontos megadása, hogy a modellben mely változók állnak kapcsolatban, és milyen utakat, korlátozásokat feltételezünk.
Kovariancia: két változó együtt járásának mértéke, amely megmutatja, hogy a változások iránya mennyire van összhangban.
Variancia: fogalmilag egy változó értékeinek szóródása a saját átlagához képest, matematikailag egy változó önmagával vett kovarianciája.
Útvonal-együttható (path coefficient, β): egy irányított kapcsolat erősségét és irányát kifejező becslés a modellben.
Exogén változó: a modellben azok a külsőként kezelt változók, amelyek változatosságát a modell nem kívánja magyarázni, és ezért nem irányul rájuk egyirányú nyíl.
Endogén változó: a modell által magyarázott, más változóktól függő változó.
Reziduális variancia (residual variance): az endogén változó varianciájának az a része, amelyet a modell nem magyaráz meg.
Független információ (independent information): a mért adatokból származó, a paraméterek becslésére felhasználható, egymástól független mennyiségek száma.
Mért variancia-kovariancia mátrix (observed variance-covariance matrix): az adatokból számolt valós variancia- és kovarianciaértékeket tartalmazó mátrix.
Feltételezett variancia-kovariancia mátrix (implied variance-covariance matrix): a modell által implikált, elméleti paraméterekből számolt variancia- és kovarianciamátrix.
Paraméter: a modellben ismeretlenként kezelt mennyiség (például egy útvonal-együttható), amelyet becsülünk.
Követési szabályok (tracing rules): szabályok, amelyek alapján meghatározható, hogy egy adott útvonal, illetve kapcsolat hogyan járul hozzá a kovarianciákhoz.
Mediáció: amikor a modellben két változó kapcsolatát részben vagy egészben egy harmadik, közvetítő változón (mediátoron) keresztül, indirekt hatásként írjuk le.
Szaturált modell (saturated model): olyan modell, amely minden lehetséges kapcsolatot tartalmaz, ezért tökéletes illeszkedést ad (0 szabadságfok).
Korlátozott modell (restricted model): olyan modell, ahol egyes utak hiányoznak, vagy paraméterek rögzítve vannak, így kevesebb kapcsolatot enged, mint a szaturált.

Modell azonosíthatósága (model identification): annak ellenőrzése, hogy a paraméterek egyértelműen becsülhetők-e az adatokból.
Alulidentifikált modell (underidentified model): nincs elég információ a paraméterek egyértelmű becslésére (negatív szabadságfok), a modell nem futtatható.
Pontosan vagy éppen identifikált modell (just identified model): épp annyi független információ van, amennyi a paraméterek becsléséhez szükséges (szaturált modell, 0 szabadságfok, illeszkedés automatikusan tökéletes).
Túlidentifikált modell (overidentified model): több információ van, mint a paraméterek száma, így a modell illeszkedése tesztelhető (pozitív szabadságfok).
Direkt hatás: közvetlen kapcsolat két változó között.
Indirekt hatás: közvetett kapcsolat, amely egy vagy több közvetítő változón keresztül valósul meg.
Szabadságfok (degree of freedom, df): a független információk és a becsült paraméterek különbsége, amely meghatározza, hogy mennyire tesztelhető a modell.
Paraméterbecslés (parameter estimation): a modell ismeretlen paramétereinek meghatározása az adatok alapján.
Maximum Likelihood (ML): a leggyakoribb becslési módszer, amely az adatok legvalószínűbb paramétereit keresi a feltételezett eloszlás alapján.
Diagonally Weighted Least Squares (DWLS): alternatív becslési módszer, főleg ordinális adatokhoz, amely a kovarianciamátrix diagonális elemeit használja súlyozásként.
Illeszkedés (model fit): annak mértéke, hogy a modell által feltételezett variancia-kovariancia mátrix mennyire közelíti a megfigyelt mátrixot.
Illeszkedési mutatók (fit indices): számértékek, amelyek a modell illeszkedését jellemzik (például CFI, TLI, RMSEA, SRMR).
Abszolút illeszkedési mutatók (absolute fit indices): azt mutatják, hogy a modell közvetlenül mennyire képes visszaadni a megfigyelt adatokat (például χ^2, SRMR, RMSEA).
Relatív illeszkedési mutatók (comparative fit indices): a nullmodellhez viszonyítva mutatják a modell illeszkedésének relatív javulását (például CFI, TLI).
Beágyazott modellek (nested models): olyan modellek viszonya, ahol az egyik a másik speciális esete (például egy plusz út elhagyása).
Nem beágyazott modellek (non-nested models): olyan modellek viszonya, amelyek szerkezetükben nem vezethetők le egymásból.
Modifikációs index (modification indices): statisztikai mutató, amely megmutatja, mennyivel javulna a modell illeszkedése, ha egy adott fixált paramétert szabadon engednénk.

3. Kutatási folyamat

Az útvonalelemzés megjelenhet a strukturális egyenletmodellezés (SEM) részeként, amikor a mérési modell és a strukturális modell együttes vizsgálata történik, de önálló elemzésként is alkalmazható, kizárólag a változók közötti ok-okozati viszonyokra koncentrálva. Előbbi esetben a strukturális modell változókészletében lehetnek manifeszt és látens változók is, utóbbi esetben minden változó manifeszt formában szerepel. Amikor önálló elemzésként használjuk, érdemes megkülönböztetni a megkötésekkel rendelkező útvonalmodelleket a speciális, szaturált modelleket elemző komplex mediációs vizsgálatoktól, hiszen a két módszertan eltérő értelmezési és azonosítási kérdéseket vet fel.

3.1. Mintaválasztás

Mint minden elemzésnél, a PA-ban is az adatfelvétel mindig abból a célpopulációból történik, amelyre az elemzés eredményeit általánosítani kívánjuk. Módszertani szempontból a legfontosabb a megfelelő mintanagyság biztosítása. A stabil összefüggési mintázatok megbízható becsléséhez a szakirodalom szerint legalább 200 fő, illetve változónként 10-15 fő részvétele javasolt. Bonyolultabb elemzési tervek esetén ennél nagyobb elemszám válhat szükségessé. Emiatt a PA-vizsgálatokat jellemzően közepes vagy nagy mintákon (500-1000 fő) végzik, mert ezen a nagyságrenden megbízhatóbban és általánosíthatóbban határozhatók meg a változók közötti összefüggések.

3.2. Adatgyűjtés, az adatok előkészítése

Az adatgyűjtés tervezése során különös figyelmet kell fordítani a változók mérési szintjére. A PA leggyakrabban használt paraméterbecslési eljárása a maximum likelihood (ML), amely feltételezi, hogy a bevitt változók legalább intervallumszintűek. Ennek következtében a legelterjedtebb gyakorlat az, hogy a modell változóiként validált kérdőívekből származó skálapontszámokat adunk meg. Ezek mellett természetesen bevonhatók olyan egyetlen kérdéssel lekérdezhető háttérváltozók is, mint például az életkor, a nem vagy más, megfelelő mérési szintű demográfiai és egyéb tényezők.

Kevésbé ideális, de a gyakorlatban előforduló megoldás az, amikor a kutató saját maga állít össze tételeket, majd feltáró faktoranalízissel (EFA) alakít ki belőlük skálákat, és ezekkel futtatja le a PA-modellt. Ez ugyan technikailag elfogadott, azonban felveti a kérdést, hogy miért választunk különálló útvonalelemzést, ha a skálákból kiindulva a SEM is alkalmazható lenne. A SEM ugyanis lehetővé tenné a mérési hibák explicit kezelését, ami a tisztán manifeszt változókkal dolgozó PA-ból hiányzik.

Önálló PA alkalmazásakor tehát ajánlott külön figyelmet fordítani az endogén változók megfelelő mérési szintjére. Például ha a modellben az észlelt egészségi állapot mediáló tényezőként szerepel, akkor azt célszerű többtételes, validált skálával mérni, nem pedig egyetlen 1–5 fokú Likert-típusú tétellel. Az exogén változók esetében nincs ilyen szigorú kitétel, így azok között gyakrabban találunk dichotóm vagy ordinális mérési szintű változókat is, például nemet vagy iskolai végzettséget.

Ha elkerülhetetlen, hogy ordinális változó kerüljön endogén változóként a modellbe, akkor el lehet tekinteni az ML használatától, és helyette robusztusabb becslő eljárást (például Weighted Least Squares Mean and Variance adjusted, WLSMV) lehet választani. Különösen nehéz a helyzet akkor, amikor az endogén változó dichotóm vagy nominális mérési szintű. Ilyenkor többféle megoldás szokott előfordulni. Egyrészt alkalmazható *probit- vagy logit-linkkel működő generalizált strukturális egyenletmodell* (például Mplusban a categorical beállítás), amely kifejezetten a bináris vagy multinomiális kimenetekhez illeszkedik. Másrészt előfordul, hogy a kutatók a nominális változókat dummy-kódolással alakítják át, és ezeket folytonosként kezelve illesztik be a modellbe, bár ez utóbbi megoldás kevésbé elegáns, és bizonyos esetekben torzításhoz vezethet. Alternatívaként a *bayesi becslési eljárások* is alkalmazhatók, amelyek rugalmasabb keretet biztosítanak a kategorikus endogén változók kezelésére.

Az adatfelvételt követően az információkat hagyományos, táblázatos formában szükséges elrendezni, az oszlopok a változókat, a sorok a vizsgálati személyeket reprezentálják.

3.3. Az elemzés beállításai

21. doboz: Az útvonalelemzés szoftveres környezetének áttekintése

Az útvonalelemzés többféle statisztikai szoftverben elérhető, jellemzően a strukturális egyenletmodellezés (SEM) részeként.
Történetileg ki kell emelni a LISREL programot, amely az első olyan széles körben használt szoftver volt, amely a SEM teljes keretét lefedve kínált elemzési lehetőségeket. Bár a LISREL nagy hatással volt a módszertan fejlődésére, napjainkban már alig használják, helyét korszerűbb és rugalmasabb rendszerek vették át.

A kereskedelmi szoftverek közül az egyik legismertebb az Mplus, amely rendkívül széles körű lehetőségeket biztosít SEM-, PA-, valamint kevert és komplex modellek futtatására. Az Mplus szöveges szintaxison alapul, így használata eleinte tanulást igényel, de nagy előnye, hogy sokféle speciális becslési eljárást (például kategorikus változókra, hiányzó adatokra vagy többszintes struktúrákra) kínál. Hasonlóképp kereskedelmi termék az AMOS, amely az SPSS kiegészítőjeként vált ismertté. Az SPSS-integráció miatt sok kutató számára könnyebben hozzáférhető volt, de ma már kevésbé dinamikusan fejlődik, mint más alternatívák.

Az R környezetben futó lavaan csomag az egyik legfontosabb ingyenes lehetőség. Bár szöveges kódolásra épül, gyorsan fejlődik, széles körű funkcionalitást kínál, és ma már a legtöbb hagyományos és korszerű becslési eljárást támogatja. Azok számára, akik az R kódolását nehezebben kezelik, a JASP és a JAMOVI kínálnak alternatívát, amelyek felhasználóbarát, grafikus felületen futtatják a lavaan motorját. A lavaan szintaxisát ugyan továbbra is szövegesen kell megírni, de a beállításokat menüpontokkal lehet megtenni, és mindkét szoftver szép táblázatos eredménymegjelenítést ad. Szaturált modellek esetén teljes egészében menüvezérelt beállításokat is biztosítanak.

A szaturált modellekhez kapcsolódva meg kell említeni a PROCESS makrót, amely ugyan nem klasszikus értelemben vett útvonalmodell, hiszen nem egyszerre, hanem egymásra épülő regressziókból számítja ki a mediációs és moderációs hatásokat. Ez azonban szaturált modellek esetében nem jelent problémát, hiszen ott a megoldás egyértelmű. A PROCESS másik sajátossága, hogy a kutatónak nem kell magának megadnia a teljes modellt, hanem előre definiált, gyakran előforduló mediációs és moderációs struktúrákból választhat, amelyeket saját adataival tölthet fel.

Végül meg kell említeni a SmartPLS programot, amely ugyan szintén strukturális egyenletmodellezést kínál, de teljesen más filozófiát követ. Nem a modell és a mért adatok közötti eltérés minimalizálását (illeszkedés maximalizálását) tekinti fő céljának, hanem a variancia maximalizálására épülő partial least squares (PLS) megközelítést alkalmazza. Ez azt jelenti, hogy más típusú mutatókkal dolgozik, mint a hagyományos PA és SEM, és inkább prediktív célokra használható. A részleteket a SEM-ről szóló fejezetben fogjuk bemutatni.

A szoftverek ismertetését követően térjünk rá azokra a beállításokra, melyeket – bármely szoftvert is választjuk – mindenhol meg kell tennünk. Az elemzés első lépése a modell vizuális megtervezése és felrajzolása. Ideális esetben ez még az adatfelvétel előtt megtörténik, így már a kutatás tervezésekor ügyelhetünk arra, hogy a gyűjtött adatok megfeleljenek a modellben tervezett szerepüknek. Ha ez nem előzte meg az adatgyűjtést, akkor is szükséges a modell megrajzolása, az adatok és azok sajátosságainak figyelembevételével.

A modellt érdemes először papíron vázolni. A rajzon jelenjen meg a modellben szereplő összes változó az őket összekötő útvonalakkal együtt. A változókat célszerű balról jobbra vagy fentről lefelé elrendezni, hogy az ábra áttekinthető maradjon. Hagyományos PA esetén a változókat manifeszt formában, téglalapokban ábrázoljuk, míg SEM tervezésekor a látens változókat oválissal jelöljük. Az útvonal-együtthatókat egyirányú nyilak mutatják, és ahol nem feltételezünk kapcsolatot, oda nem kerül nyíl. Az exogén változók között szinte mindig megengedett a kovariancia, ezt is jelöljük a modellben. Amennyiben reziduális kovarianciákat feltételezünk (például párhuzamos mediáló változók között), ezeket kétirányú nyilakkal ábrázoljuk.

Az endogén változók reziduálisait a papíralapú vázlaton nem feltétlenül szükséges feltüntetni, azonban a szoftveres megvalósítás során mindenképp figyeljünk rá, hogy ezek a modellben megjelenjenek.

A következő lépés a modell átültetése a szoftverbe. Grafikus szerkesztésű programok esetében (például az AMOS grafikus felülete) ez többnyire triviális, a változókat „dobozként” vagy „ellipszisként” elhelyezzük, majd a nyilakat behúzzuk. Ugyanakkor itt is figyelni kell, hogy például a reziduálisokat ne felejtsük el berajzolni, illetve elnevezni.

Szöveges felületnél (például lavaan R-ben) a modell szintaxissal adható meg. Jó stratégia, ha a kutató balról jobbra halad, és minden endogén változónál egyesével felsorolja, milyen prediktorok mutatnak rá. A lavaan szintaxisban a regresszió jele a ~. Ez azt jelenti, hogy a bal oldali változót magyarázzák a jobb oldali változók.

Ha például egy M1 mediáló változót az X1 és X2 prediktor magyarázza, akkor ezt így írjuk:

M1 ~ b1*X1 + b2*X2
Itt M1 az endogén változó (erre futnak be a nyilak), X1 és X2 a prediktor változók, b1 és b2 a nyilak elnevezései (útvonal-együtthatók), a + jellel választjuk el a több magyarázó változót.
Minden endogén változóhoz külön sort kell felírni. Például ha az Y kimeneti változónkat M1 és X1 magyarázza, akkor:

Y ~ b3*M1 + b4*X1
Ezzel fokozatosan, soronként építhető fel a teljes modell. Miután az összes regressziós kapcsolatot megadtuk, a modell másik fontos eleme a kovarianciák szintaxisának megírása. A lavaanban ezt a ~~ jellel adjuk meg. Ha például két párhuzamos mediátor (M1 és M2) reziduálisai között kovarianciát feltételezünk, akkor azt a következő sorral tudjuk jelezni:

M1 ~~ M2
Következő lépésként mindenképp be kell állítani a paraméterbecslő eljárást és a standard errorok számításának módját. Emellett azt is meg kell adni, hogy a kimenetben milyen illeszkedési mutatókat szeretnénk látni, valamint az R^2-értékeket. Ha az illeszkedés nem megfelelő, célszerű a modifikációs indexeket is bekapcsolni, amelyek javaslatot adnak a modell javítására.

Amennyiben az illeszkedés elfogadható, minden szoftver automatikusan megjeleníti a standardizálatlan útegyütthatókat és azok standard hibáit (SE). Gyakran azonban külön beállítást igényel, hogy a kimenetben a standardizált értékek is szerepeljenek, illetve az, hogy láthatók legyenek a konfidenciaintervallumok és a p-értékek.

3.4. Az elemzés értelmezése

A modell értelmezésének első lépése az illeszkedésvizsgálat, melynek során több mutatót (például χ^2, CFI, RMSEA, SRMR) együttesen értékelünk. Mivel ezek a mutatók az eltérés mértékét, a modell méretét, összetettségét és megkötöttségét eltérően súlyozzák, nem mindig adnak egybehangzó eredményt. Az a legjobb, ha minden fontosabb mutató jó illeszkedést jelez, de az is elfogadható, ha a legtöbb jó illeszkedés mellett egy-egy mutató enyhén elmarad az elfogadhatótól. Amennyiben az illeszkedés nem megfelelő, a faktortöltések elemzése

helyett a modellt kell javítani elméleti megfontolások vagy a modifikációs indexek javaslatai alapján.

Az értelmezés második köre akkor kezdődik, amikor sikerült kialakítanunk egy megfelelően illeszkedő modellt. Az értelmezés ezen szakasza nem különbözik attól, ahogy bármely regresszión alapuló elemzés eredményeit értelmezzük. Ellenőrizzük, mely együtthatók szignifikánsak, és melyek nem, valamint azt, hogy a szignifikáns hatások közül melyik pozitív, melyik negatív. A standardizált regressziós együtthatók alapján a hatások erőssége is a megszokott módon összehasonlítható.

Miután az útmodell eredményeit értelmeztük, vissza kell térni a modell strukturális specifikációjához, de nem azért, hogy tovább változtassunk a modellen, hanem azért, mert itt lehet megadni, hogy pontosan mely indirekt hatásokat szeretnénk a programmal kiszámoltatni, hiszen ezek alapértelmezetten nem minden esetben jelennek meg a kimenetben.

Az *indirekt hatás* definíció szerint olyan hatás, amely nem közvetlenül, hanem közbeiktatott változókon (mediátorokon) keresztül valósul meg. Matematikailag az indirekt útvonal az őt alkotó útvonal-együtthatók értékének szorzataként adódik, tehát ha egy prediktor egy mediátorra hat (együtthatója legyen βa), az pedig a kimeneti változóra (βb), akkor az indirekt hatás a két együttható szorzata (βa*βb). Amennyiben több közvetett útvonal is lehetséges, mindegyik külön számolható indirekt útvonalként, és az összes potenciális indirekt útvonal összege adja a teljes indirekt hatást. Ezenfelül a prediktor változó teljes hatása megegyezik a direkt hatás és az összes indirekt hatás összegével.

Nézzünk néhány példát!

A 21. ábra A panelje egy egyszerű lineáris regressziót jelenít meg. Itt a1 az X1 prediktor változó hatása az Y változóra, később erre utalunk majd teljes hatásként. A B panelben egy többszörös lineáris regresszió látszik, ahol b1 az X1 hatása Y-ra úgy, hogy kontrollálunk X2-re. Fontos, hogy a1≠b1, mivel b1 az X1 hatásának csak azon része, amely független X2-től.

A C panel egy hagyományos mediációs modell (M változó megegyezik a B panel X2 változójával). Szaturált, hiszen minden lehetséges útvonalat tartalmaz. Itt c1*c2 adja az X1 M-en keresztül megvalósuló hatását, azaz az M-en keresztüli indirekt útvonalat X1 és Y között, míg c3 adja a direkt útvonalat. Itt több egyezőséget is érdemes megfigyelni. Egyrészt a B és C panelt összevetve c3 = b1, mivel X1 direkt hatása szempontjából mindegy, hogy csak megengedjük a kovarianciát X1 és X2 között, vagy vizsgáljuk is. Másrészt az A és C panelt összevetve c1 * c2 + c3 = a1, azaz az indirekt és direkt útvonal összege kiadja a teljes hatást.

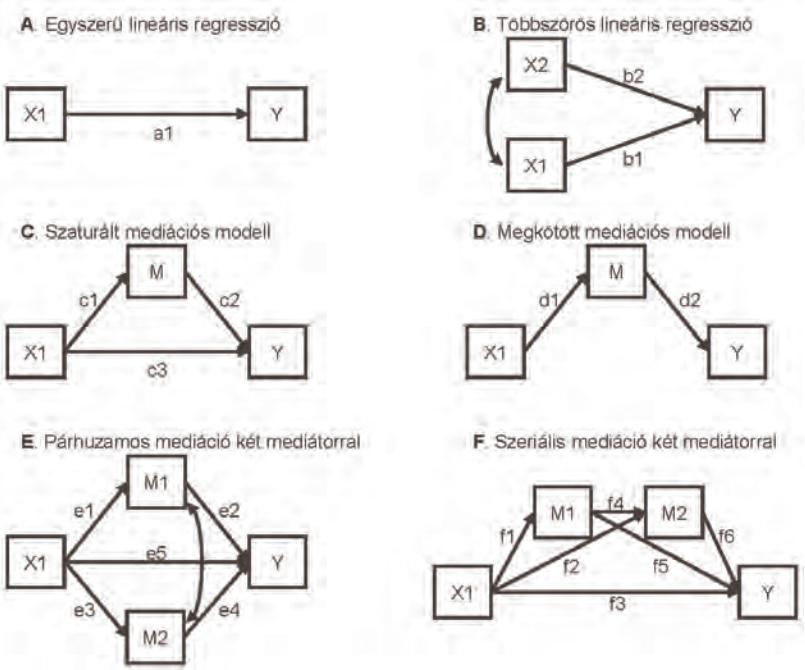
A D panelben egy olyan mediációs modell látható, mely tartalmaz egy megkötést, azt feltételezzük, hogy X1 közvetlenül nem, csak M-en keresztül, indirekten hat Y-ra. Fontos tudnunk, hogy a C panel c1 és c2 együtthatója nem egyezik meg a D panel d1 és d2 együtthatójával, hiszen a D esetben egy korlátozott modellt látunk, ahol a paraméterbecslő eljárás úgy állítja be d1 és d2 értékét, hogy az általuk létrejövő indirekt útvonal az X1-Y összefüggését is a lehető legjobban magyarázza.

Az E panel mediációs modellje két egymással párhuzamos mediátort tartalmaz. Erre példa a korábban már említett vizsgálat, ahol azt ellenőrizzük, hogy a szépség hatása az azonos és/vagy ellenkező nemű személyek közötti népszerűségen keresztül hat-e az önbizalomra. Ekkor külön ellenőrizhető az azonos neműek körében tapasztalt népszerűség mediáló hatása (e1*e2) és az ellenkező neműek körében tapasztalt népszerűség mediáló hatása is (e3*e4), valamint a kettő összegeként a népszerűség összmediáló hatása (e1*e2 + e3*e4). Az e5 természetesen továbbra is a szépség direkt hatását méri az önbizalomra, de nem feltétlenül egyezik meg c3-mal és b1-gyel, hiszen nem az azokban látható M-en keresztüli indirekt hatás van leválasztva, hanem két új változón (M1 és M2) keresztüli indirekt útvonalak.

Végül az F panel egy szeriális mediáció. Jó példa lehet, hogy a szépség hat a népszerűségre, a népszerűség folytonos megerősítést ad a szociális készségekről, és ennek a szociális kompetenciának a megerősítése vezet az önbizalomhoz. Ekkor egy szaturált modellben négy úton lehet eljutni a szépségből az önbizalomba. Direkt útvonal az, hogy a szép emberek reggel megnézik magukat a tükörben, és a látvány növeli az önbizalmukat. Az f1*f4*f6-os útvonal az előbb említett szeriális mediáció, azaz a szép emberek népszerűbbek, a népszerűség folytonos visszajelzés forrása szociális kompetenciáikról, és e kompetenciákban való megerősítettség vezet az önbizalomhoz. Létezik emellett viszont másik két indirekt útvonal is. Az egyik az f1*f5, mely útvonal kihagyja a visszacsatolást a szociális kompetenciákról, mert a népszerűség nemcsak ezen keresztül hathat az önbizalomra, hanem az is lehet, hogy a népszerű emberek több társas támogatottságot kapnak, és így ezek az emberek általában véve azt élik meg, hogy az életben mindent „könnyen el tudnak érni", ami önbizalomhoz vezet a társas kompetenciáktól függetlenül is. A másik pedig az f2*f6, mely a népszerűséget hagyja ki, és a szépségből közvetlenül a szociális kompetenciákhoz, onnan az önbizalomhoz ugrik. Ez utóbbinak van legkevésbé szakmai szempontból értelme, ha a szaturált modell helyett korlátozott modellt tesztelnénk, az f2-es útvonalat vélhetően el lehetne hagyni anélkül, hogy az illeszkedés jelentős mértékben csökkenne. Összegezve, az F panelben az egy direkt útvonal (f3) mellett három indirekt útvonal rajzolható fel, és ezek összege adja a teljes indirekt útvonalat

(f1*f4*f6 + f1*f5 + f2*f6). Végül, mivel az F panel modellje szaturált, az indirekt és direkt útvonalak összege megadná az A panel a1 útvonalának értékét.

Az E és F panel szaturált modellt mutat be, de természetesen egy hagyományos útvonalelemzésben korlátozásokat vezetnénk be. Legtöbbször a direkt útvonalat hagyjuk el, mivel pont azt szeretnénk tesztelni, hogy két változó összefüggése magyarázható-e közvetítő változókon keresztül. Ha igen, akkor a direkt útvonal elhagyása nem okoz nagy csökkenést a modell illeszkedésében, hiszen úgyis kicsi (esetleg nem is szignifikáns) az értéke. Emellett elhagyható bármely útvonal, amelyről szakmai szempontok alapján úgy véljük, hogy nem jelentősek, például az F panelből vélhetően az illeszkedés jelentős csökkenése nélkül elhagyható f2 is.



21. ábra: Szemléltető példák útvonalmodellekre

Lényeges hangsúlyozni, hogy az indirekt útvonalak nem külön paraméterek a modellben, hanem számított értékek (meglévő útvonalak szorzatai). Emiatt a szabadságfokokat nem érintik, hiszen nem új becslési feladatot hoznak létre, hanem a már becsült paraméterekből származtatott mennyiségek. Ennek megfelelően nem kell takarékoskodni a számításukkal, ha a 21. ábra F paneljén a

három X1-et és Y-t összekötő indirekt útvonal mellett a modell megértéséhez szükségét érezzük az X1-et M2-vel összekötő (f1*f4) részútvonal számításának is, következmények nélkül megtehetjük.

A gyakorlatban a szoftverekben (például lavaanban, R-ben) a közvetett hatásokat explicit módon meg kell adni. Ezt lavaanban a := operátorral tudjuk deklarálni, például a C panelhez:

indirekt := c1 * c2
total := c3 + (c1 * c2), de azt is írhatjuk, hogy total := c3 + indirekt.

4. Mikor használjuk és mikor ne használjuk a módszert?

Először érdemes világosan elkülöníteni a szaturált mediációs modelleket a korlátozott modellt tesztelő útvonalelemzéstől. Az elemzés módszertani megválasztása nagymértékben függ a kutatási kérdéstől és az adatok szerkezetétől. A hagyományos, szaturált mediációs megközelítés (például a PROCESS makró alkalmazása) akkor célszerű, ha a kutató egy jól körülhatárolt, néhány változót érintő hipotézist kíván tesztelni, ahol minden lehetséges útvonal becslésre kerül, és a modell lényegében szaturált. Ez egyszerű, áttekinthető és a gyakorlatban gyakran elegendő eszköz. Ezzel szemben útvonalelemzést korlátozott modellel akkor érdemes alkalmazni, ha a vizsgálat komplexebb, párhuzamos vagy szeriális mediátorokat, illetve strukturális összefüggések hálózatát tartalmazza. Ilyenkor a modell illeszkedésének statisztikai vizsgálata lehetővé válik, és a kutató explicit módon kontrollálja, mely utak kerülnek a modellbe, így tesztelve a hipotézisek elméleti megalapozottságát. További szempont, hogy a modell tartalmaz-e moderátor változót. Ha igen, és a moderátor folytonos, akkor arra speciálisan kialakított megoldások léteznek szaturált modell használata során (például a PROCESS makróban). Ha a moderátor kategorikus változó, az multigroup SEM keretében kezelhető (részletesen lásd a SEM-fejezetben).

Az elemzési stratégia megválasztásakor fontos különbséget tenni a pusztán mért változókkal dolgozó útvonalelemzés és a látens konstrukciókat is kezelő strukturális egyenletmodellezés között. Ha a kutatás középpontjában a változók közötti kapcsolatok feltárása áll, és a vizsgált változók megbízhatóan mértek, egytételes skálák és indexek alapján vagy validált kérdőívekből származnak, akkor a PA elegendő és célszerű megközelítés lehet. Ezzel szemben, ha a vizsgálatban hangsúlyos szerepet játszik a mérési hibák kontrollálása, a komplex vagy saját magunk által összeállított skálákkal reprezentált konstrukciók kezelése vagy épp a mérési modell illeszkedésének vizsgálata, akkor mindenképpen érdemes SEM-et alkalmazni. A SEM előnye, hogy egyszerre képes megbízhatóan kezelni a mérési és a strukturális modelleket, ugyanakkor nagyobb mintanagyságot igényel, és módszertanilag is komplexebb. Röviden: egyszerűbb, jól mérhető változók közötti kapcsolatokra a PA, míg látens konstrukciókat és mérési hibákat is figyelembe vevő kutatási kérdésekre a SEM a megfelelőbb választás.

5. A minőség biztosítása

Az útvonalelemzés eredményeinek közlése során a kutató felelőssége, hogy világos, reprodukálható és értelmezhető módon mutassa be a modellt és annak eredményeit. Az alábbiakban a közlés három szintjét különítettünk el: minimálisan szükséges, ajánlott és kiváló közlés.

Szükséges közlés (elvárt):

a) Biztosítja, hogy az olvasó megértse az elemzés logikáját.
b) A modell rövid bemutatása tartalmazza a vizsgált változókat, a köztük feltételezett kapcsolatokat és az elméleti háttér ismertetését.
c) Megtörténik az adatok és a minta jellemzése elemszámokkal, mérési szintekkel és leíró statisztikákkal.
d) Említésre kerül a használt paraméterbecslő eljárás neve.
e) Legalább három illeszkedési mutatót közöl a kutató (leggyakrabban χ^2(df), CFI és RMSEA), hogy megítélhető legyen a modell érvényessége.
f) Megtörténik az útparaméterek bemutatása a standardizált regressziós együtthatókkal (útvonal-együtthatókkal) és azok szignifikanciaértékével.

Ajánlott közlés (jó gyakorlat) a szükséges közlésen felül:

a) Lehetővé teszi az elemzés kritikus értékelését.
b) Részletes modellábra készül, amely vizuálisan megjeleníti a változókat és a köztük lévő kapcsolatok irányát és nagyságát standardizált együtthatókkal.
c) A használt paraméterbecslő mellett az SE becslő eljárás neve is megadásra kerül.
d) A kutató az illeszkedési mutatók szélesebb körét is bemutatja (χ^2(df), χ^2/df, CFI, TLI, RMSEA konfidenciaintervallummal, SRMR), hiszen ezek eltérő szempontokból értékelik a modellt.
e) A standardizált és standardizálatlan útparaméterek mellett megjelenik azok standard errorja, szignifikanciaértéke és konfidenciaintervalluma a becslések bizonytalanságának jelzésére.
f) Az indirekt hatásokat bemutatja, és közli azok nagyságát és szignifikanciáját.
g) Tárgyalja a módszertani döntéseket és indoklásukat (például normalitás feltételezése, hiányzó értékek kezelésének módja).

Kiváló közlés az ajánlott közlésen felül:

a) Lehetővé teszi a reprodukciót és mélyebb értelmezést.
b) Közli a teljes modellreprodukcióhoz szükséges adatokat, az illeszkedési mutatók mellett a reziduálisokat, kovarianciákat, vagy akár a teljes inputmátrixot (például kiegészítő anyagban) elérhetővé teszi.
c) Ha a modell változtatásokon esett át (például modifikációs indexek alapján), a változás folyamatát részletesen leírja, a különböző állomások illeszkedési mutatóit közli.
d) Vizsgálja az alternatív modelleket, a választott modellt összehasonlítja a hasonló, de elméletileg releváns modellekkel, a döntéseket indokolja.
e) Ellenőrzi a robusztusságot (például mintanagyság-érzékenység), tárgyalja a hiányzó adatok kezelését, szükség esetén bootstrap becslést alkalmaz.
f) Megtörténik a szignifikanciatesztelés mellett a hatásméretek értelmezése is a gyakorlati jelentőség hangsúlyozásával.
g) Közli a kódot vagy szintaxist az elemzés reprodukálhatósága érdekében (például Mplus, lavaan, Amos-script kiegészítő fájlban).
h) Az eredmények értelmezése mellett reflektál a módszertani korlátokra és a lehetséges továbblépésekre.

6. A módszer változatai és haladó alkalmazásai

A hagyományos útvonalelemzés mellett számos haladó eljárás is rendelkezésre áll, amelyek a bonyolultabb hipotézisek tesztelését vagy a módszertani korlátok áthidalását szolgálják.

a) A strukturális egyenletmodellezés lehetőséget ad arra, hogy a megfigyelt változók mellett látens konstruktumokat is modellezzünk, rugalmasan illesztve a módszert a kutatási kérdéshez.
b) A bayesiánus útvonalelemzés lehetővé teszi az a priori információk beépítését, és kis minták esetén is megbízhatóbb eredményt adhat.
c) A moderátort is tartalmazó modellek azt vizsgálják, hogy egy kapcsolat erőssége vagy iránya hogyan változik egy harmadik változó (moderátor) függvényében.
d) A multigroupelemzés lehetővé teszi, hogy ugyanannak a modellnek a paramétereit különböző csoportokban (például nem, életkori sávok, kultúra) hasonlítsuk össze.
e) A nemlineáris PA a lineáris hatások mellett képes kezelni görbült vagy interakciós összefüggéseket is.
f) A longitudinális modellek az ismételt mérések alapján elemzik a változók közötti összefüggéseket, így feltárhatók az időbeli folyamatok és ok-okozati mechanizmusok.

7. Összefoglalás

Az útvonalelemzés a strukturális egyenletmodellezés speciális formája, amelyben kizárólag megfigyelt változókkal dolgozunk. Lényege, hogy többváltozós regressziók hálózatát egyidejűleg vizsgáljuk, és így képesek vagyunk közvetlen, közvetett és összhatásokat feltárni. A módszer előnye, hogy áttekinthető, jól alkalmazható összetettebb hipotézisek tesztelésére, valamint lehetőséget ad

a változók közötti kapcsolatok irányának vizsgálatára. Ezzel pontosabb képet kaphatunk arról, hogyan illeszkednek egymáshoz az elméleti modell elemei, mint ha külön regressziókat futtatnánk.

Ugyanakkor a PA korlátja, hogy nem teszi lehetővé látens konstrukciók bevonását, így a mérési hibát sem tudja kezelni, ezért elsősorban akkor javasolt, ha a vizsgált változók megbízhatóan, közvetlenül vagy validált kérdőívekkel mérhetők.

Gyakorlati jelentősége abban rejlik, hogy egy viszonylag egyszerű, de rugalmas eljárással komplex elméleti modellek is vizsgálhatók, például mediációk, sorozatos hatások vagy többváltozós oksági láncok. A PA tehát jó kiindulópont a kutatóknak az összetettebb összefüggések feltárásához, és gyakran előszobája a bonyolultabb, látens változókkal is dolgozó SEM-elemzéseknek.

8. Önellenőrző kérdések

1. kérdés: Mit vizsgál az útvonalelemzés?

Válasz: A PA a változók közötti közvetlen és közvetett összefüggéseket vizsgálja egy előre meghatározott elméleti modell alapján.

2. kérdés: Miben kevesebb a PA egy teljes SEM-hez képest?

Válasz: A PA csak megfigyelt változókkal dolgozik, nem von be látens konstrukciókat, így a mérési hibát nem tudja kezelni.

3. kérdés: Mi a korlátozott modell a PA-ban?

Válasz: A korlátozott modell olyan PA, amelyben bizonyos útvonalakat nullára állítunk (nem szaturált), így a modell tesztelhető, és van szabadságfoka.

4. kérdés: Hogyan ellenőrizzük a modell illeszkedését?

Válasz: A PA-ban is illeszkedési mutatókat (például χ^2, CFI, RMSEA, SRMR) használunk, amelyek megmutatják, mennyire felel meg a modell az adatoknak.

5. kérdés: Mi a különbség a direkt és indirekt hatások között?

Válasz: A direkt hatás egy változó közvetlen kapcsolatát jelzi egy másikkal, míg az indirekt hatás közvetítőkön keresztül valósul meg.

6. kérdés: Milyen gyakorlati helyzetekben célszerű PA-t használni?

Válasz: Olyan esetekben, amikor világos hipotézisünk van a változók közötti kapcsolatok irányáról, és a mérési eszközök megbízhatók, így nincs szükség látens változók modellezésére.

Ajánlott irodalom

Bollen, K. A. (1989). Structural equations with latent variables. John Wiley & Sons. https://doi.org/10.1002/9781118619179

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models. *Sociological Methods & Research, 36*(4), 462–494. https://doi.org/10.1177/0049124108314720

Coutts, J. J., & Hayes, A. F. (2023). Questions of value, questions of magnitude: An exploration and application of methods for comparing indirect effects in multiple mediator models. *Behavior Research Methods, 55*(7), 3772–3785. https://doi.org/10.3758/s13428-022-01988-0

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.

Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating Model Fit of Measurement Models in Confirmatory Factor Analysis. *Educational and Psychological Measurement, 84*(1), 123–144. https://doi.org/10.1177/00131644231163813

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. The Guilford Press.

Lennox, C., & Payne-Mann, C. (2025). An Explanation of Path Analysis and Recommendations for Best Practice. *Contemporary Accounting Research*, Article 1911-3846.70019. Advance online publication. https://doi.org/10.1111/1911-3846.70019

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Rosseel, Y. (2012). *lavaan : An R Package for Structural Equation Modeling*. *Journal of Statistical Software, 48*(2). https://doi.org/10.18637/jss.v048.i02

Sathyanarayana, S., & Mohanasundaram, T. (2024). Fit Indices in Structural Equation Modeling and Confirmatory Factor Analysis: Reporting Guidelines. *Asian Journal of Economics, Business and Accounting, 24*(7), 561–577. https://doi.org/10.9734/ajeba/2024/v24i71430

Schoemann, A. M., Moore, E. W. G., & Yagiz, G. (2025). How and why to follow best practices for testing mediation models with missing data. *International Journal of Psychology : Journal International De Psychologie, 60*(1), e13257. https://doi.org/10.1002/ijop.13257

Williams, L. J., Williams, A. R., & O'Boyle, E. H. (2024). Assessment of Path Model Fit: Evidence of Effectiveness and Recommendations for use of the RMSEA-P. *Organizational Research Methods, 27*(1), 64–87. https://doi.org/10.1177/10944281221124946

Wright, S. (1921) Correlation and Causation. Journal of Agricultural Research, 20, 557–585.

Wright, S. (1934). The Method of Path Coefficients. Annals of Mathematical Statistics, 5(3), 161–215.

X. STRUKTURÁLIS EGYENLETMODELLEZÉS (SEM)

SOLTÉSZ-VÁRHELYI KLÁRA

A fejezet a következő kérdéseket válaszolja meg (feltételezve a CFA- és PA-fejezetek tudásanyagának ismeretét):

a) Milyen elméleti alapokon nyugszik a SEM?
b) Mi a különbség a kovarianciaalapú (CB-SEM) és varianciaalapú (PLS-SEM) eljárások között?
c) Mi a CB-SEM elmélete? (Mivel a CB-SEM-et alkotó két módszertan, a mérési és strukturális modell a CFA- és PA-fejezetben bemutatásra került, ez a fejezet az elmélet ismételt bemutatása helyett az előző két fejezet ismeretanyagának integrálására törekszik.)
d) Mi a PLS-SEM elmélete?
e) Mik a SEM gyakorlati lépései, hogyan kell az elemzést helyesen elvégezni, az eredményeket helyesen értelmezni?
f) Milyen minőségi ismérvek alapján különböztethető meg egy kiváló SEM-tanulmány egy átlagostól?

A fejezet erősen épít a CFA- és PA-fejezetek tudásanyagára, így e fejezet megkezdése előtt erősen ajánlott azon feldolgozása.

1. Bevezetés

A strukturális egyenletmodellezés (SEM) a modern statisztikai módszertan egyik legsokoldalúbb eszköze, amely lehetővé teszi, hogy egyszerre vizsgáljuk a látens változók közötti kapcsolatokat és a mérési modellek illeszkedését. A SEM tehát ötvözi a faktoranalízis és a regressziós elemzés előnyeit, egyszerre képes kezelni a mérési hibát, a többváltozós kapcsolatokat, valamint összetett elméleti modellek empirikus tesztelését. Emiatt a pszichológiától kezdve a szociológián és a neveléstudományon át a gazdasági kutatásokig széles körben alkalmazzák.

Amikor SEM-ről beszélünk, hagyományos értelemben általában a kovarianciaalapú SEM-re (covariance-based SEM, CB-SEM) gondolunk, amely a hagyományos (általában Maximum Likelihood becslésen alapuló) megközelítés. Az utóbbi évtizedekben azonban megjelent és egyre szélesebb körben terjed egy alternatív módszertan is, a varianciaalapú, parciális legkisebb négyzetek módszerén alapuló SEM (partial least squares SEM, PLS-SEM), amely más logikára épül, és más előnyökkel, illetve korlátokkal rendelkezik.

A CB-SEM célja a teljes modell illeszkedésének vizsgálata. Azt méri, hogy az adatok összefüggései mennyire felelnek meg az elméletileg feltételezett kovarianciastruktúrának. Ez a megközelítés erősen elméletvezérelt, így elsősorban hipotézisek tesztelésére, elméleti modellek megerősítésére használatos. Nagyobb mintát igényel, és érzékeny a normalitás sérülésére, ugyanakkor pontos információt ad az illeszkedés minőségéről.

Ezzel szemben a PLS-SEM inkább a predikciót helyezi előtérbe. Fő célja a magyarázott variancia maximalizálása a függő változókban. Ez a módszer kevésbé támaszkodik a szigorú eloszlásfeltételezésekre, kisebb minták mellett is stabil eredményeket adhat, és jól használható, ha a kutatás inkább feltáró jellegű, vagy amikor a kutató célja az előrejelzés és a magyarázóerő növelése. Ugyanakkor a PLS-SEM nem nyújt olyan részletes illeszkedést mérő mutatókat, mint a CB-SEM.

Összességében tehát bár a CB-SEM és a PLS-SEM célja hasonló (összetett modellek vizsgálata látens változókkal), a két módszer csak részben fedi át egymást. A CB-SEM inkább elmélet-ellenőrző, míg a PLS-SEM inkább prediktív jellegű eszköz. Emiatt a tankönyvben egymást követően tárgyaljuk őket, külön bemutatva sajátos előnyeiket és alkalmazási lehetőségeiket.

2. A kovarianciaalapú modellezés (CB-SEM)

A két módszer közül először a kovarianciaalapú modellezést tekintjük át. Ez az eljárás a régebbi, és ezt tekintjük a hagyományos SEM-nek.

2.1. A CB-SEM elmélete

A *kovarianciaalapú strukturális egyenletmodellezés* (CB-SEM) olyan statisztikai eljárás, amely összetett modellek vizsgálatát teszi lehetővé látens változókkal és többváltozós kapcsolatokkal. Lényege, hogy a teljes modell illeszkedését a változók közötti kovarianciastruktúra alapján értékeli. Azt vizsgálja, hogy a változók empirikusan mért összefüggése mennyiben egyezik azzal az összefüggésrendszerrel, amelyet az elméletileg feltételezett modell implikál.

Fontos megérteni, hogy egyetlen elméleti modell illeszkedése sem lesz soha tökéletes a valósághoz. A társadalomtudományi jelenségek rendkívül összetettek, és bármely modell szükségképpen leegyszerűsíti a valóságot, mert egyes változókat kihagy, a mérési hibát csak közelítő módon kezeli, és az összefüggéseket a valósághoz képest egyszerűbben írja le. De éppen ez a parszimónia az elméletalkotás lényege: megtalálni azt a lehető legegyszerűbb modellt, mely már képes viszonylag pontosan magyarázni a valóság összefüggéseit.

A SEM-nek nem az a célja, hogy teljes mértékben leképezze a valóságot, hanem az, hogy a lehető legjobban megragadja az elméletileg fontosnak tartott összefüggéseket. A módszer lényege, hogy a megfigyelt kovarianciák és a modell által feltételezett kovarianciák közötti eltérést számszerűsíti, és ennek nagyságát értékeli. Másként fogalmazva: a SEM mindig azt méri, hogy a modellünk mennyire van távol a változók között empirikusan megfigyelt összefüggéstől, és ezen keresztül ad képet arról, mennyire tekinthető az elméleti modellünk a valóság adekvát leírásának.

A CB-SEM egyik sajátossága, hogy a modell két jól elkülöníthető részből áll, a *mérési modell*ből és a *strukturális modell*ből. A mérési modell a látens változók és az őket mérő indikátorok kapcsolatát írja le, vagyis azt, hogy az elméleti konstrukciók hogyan manifesztálódnak a megfigyelt változókban. Ez a rész biztosítja, hogy a látens változók mérése megbízható és érvényes legyen. A strukturális

modell ezzel szemben a látens változók közötti összefüggéseket határozza meg, és azt mutatja meg, hogyan hatnak egymásra az elméleti konstrukciók.

A két komponens elkülönítése nem csupán elméleti, hanem módszertani szempontból is lényeges. A mérési modell vizsgálata önálló elemzésként a *megerősítő faktoranalízis* (CFA) keretében történik, amelyet korábban részletesen bemutattunk. A strukturális modell pedig önálló elemzésként megfeleltethető az *útvonalelemzésnek* (PA), amelyet szintén önálló fejezetben fejtettünk ki. A CB-SEM tehát e két korábban tárgyalt módszert ötvözi, és egyszerre biztosítja a mérési érvényesség ellenőrzését és a látens konstrukciók közötti kapcsolatok vizsgálatát.

Mivel a könyvben a CFA és a PA részletes bemutatása már megtörtént, a jelen fejezet a SEM tárgyalásakor ezek ismeretét adottnak tekinti, és nem ismétli meg részletesen a két elemzési logikát. Ehelyett arra helyezi a hangsúlyt, hogyan illeszkedik össze a két részmodell egy egységes keretben, miként értékelhető a teljes modell illeszkedése, és milyen módszertani szempontok merülnek fel akkor, amikor a mérési és a strukturális modellt együttesen, integráltan vizsgáljuk.

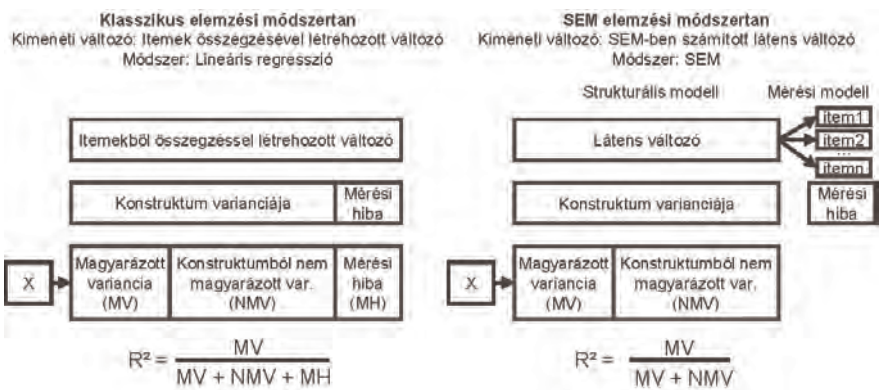
De miért olyan fontos a mérési és strukturális modell szétválasztása, ugyanakkor közös modellben való kezelése? Mert a SEM-ben a mérési hibát a mérési modellel kezeljük, és így az a strukturális modellben nem (vagy kevésbé) jelenik meg, ezzel pontosabb regressziós együtthatókat és R^2-értékeket eredményezve. Ennek logikáját kövessük végig a 22. ábra segítségével.

Hagyományosan a statisztikai elemzésekben úgy járunk el, hogy az itemekből átlagolással vagy összeadással skálapontszámot képzünk, és ezt a változót használjuk a további elemzésekben. Ilyenkor azonban a kapott skálapontszám varianciája két részből áll össze: egyrészt abból a varianciából, amely valóban a minket érdeklő konstruktumhoz tartozik, másrészt a *mérési hibából* (esetleges behatások, félrekarikázások, adatbeviteli pontatlanságok stb.). Példaként a depresszió változó pontszámában lévő változatosság egy része abból adódik, hogy vannak kevésbé és jobban depressziós személyek a mintában, ez a változatosság számunkra hasznos része. Másrészről két azonosan depressziós személy között is lehet eltérés, mert az egyikkel éppen aznap történt valami kellemetlen, így most a szokásosnál is rosszabbul érzi magát (szituatív ingadozás), de akár különbség adódhat abból is, hogy pontatlanul vittük be az egyik személy adatát (véletlen hiba). Amennyiben a stabil depressziószintet vizsgáljuk, ez mind számunkra haszontalan változatosság, azaz mérési hiba.

Ha ezt a skálapontszámot regresszióban használjuk függő változóként, és tegyük fel, hogy az egyik prediktor valóban magyarázza a konstruktum egy

részét, akkor intuitíven azt várnánk, hogy az R^2 pontosan megmutatja azt, hogy a konstruktum varianciájának hány százalékát magyarázza a prediktor. Valójában azonban az R^2 értéke nem ezt tükrözi. A számlálóban valóban a konstruktumból megmagyarázott variancia szerepel. A nevezőben viszont nemcsak a konstruktum teljes varianciája van jelen, hanem a mérési hibából származó variancia is. Másképp fogalmazva, az R^2 azt mutatja, hogy a konstruktum megmagyarázott része milyen arányt képvisel a (megmagyarázott + meg nem magyarázott + mérési hiba) teljes varianciából. Ennek következménye, hogy az R^2 kisebbnek látszik, mint amekkorának lennie kellene, és a regressziós együtthatók (b-értékek) is gyengébbnek tűnnek.

A SEM előnye, hogy a mérési modell révén a mérési hiba varianciáját elkülöníti, és a manifeszt tételek szintjén leválasztja a látens változóról. Így a látens változó varianciája sokkal tisztábban tükrözi a valódi konstruktum varianciáját. Amikor tehát a SEM-ben a látens változót magyarázza egy prediktor (ez már a strukturális modell része), az R^2 közelebb áll ahhoz az értékhez, amely elméletileg érdekel minket, a konstruktum varianciáján belül a magyarázott és meg nem magyarázott rész arányához. Ezért a SEM-ben általában nagyobb és pontosabb R^2-eket, valamint erősebb regressziós együtthatókat kapunk, mint a skálaátlagolásra épülő regressziós elemzésekben.



22. ábra: Mérési és strukturális modell elkülönítésének jelentősége

2.2. CB-SEM kutatási folyamata

A CB-SEM bemutatása után most rátérünk a módszer technikai folyamatának ismertetésére. Itt egymásba kapcsolódva jelennek meg mindazok a lépések, amelyeket korábban a CFA- és a PA-fejezetekben külön-külön megismertünk, hiszen a SEM ezek ötvözéséből építkezik. A következőkben tehát arra helyezzük a hangsúlyt, hogyan áll össze a két elemzés logikája egyetlen integrált vizsgálati keretbe.

2.2.1. Mintaválasztás

Az általános alapelv, hogy a minta jól kell reprezentálja az elméleti populációt, természetesen itt is érvényes, de emellett a további szempontokat is érdemes figyelembe venni.

Az elemzés legkritikusabb gyakorlati kérdése a mintanagyság. Ennek oka, hogy a SEM-modellek egyszerre tartalmazzák az összes látens változót és azok minden manifeszt indikátorát. A szakirodalomban gyakran alkalmazott ökölszabály, hogy legalább tíz fő jusson minden egyes indikátorváltozóra, ami már néhány látens konstruktum esetén is gyorsan több száz fős mintanagyságot jelent. Bár elméletileg már néhány száz fős mintával is futtatható egy SEM, a gyakorlatban az igazán megbízható eredmények inkább ezer körüli vagy azt meghaladó elemszám mellett születnek. Amennyiben a rendelkezésre álló minta ennél kisebb, sok kutató inkább a mérési és a strukturális modell külön-külön való elemzése mellett dönt, mert ez stabilabb és értelmezhetőbb eredményeket adhat.

A SEM érzékeny a minta heterogenitására, azaz arra, ha a vizsgált populáció valójában több alcsoportból áll eltérő mintázatokkal. Ilyenkor a modell „átlagolva" illeszkedik, ami elfedheti a valódi összefüggéseket. Ha várható csoportkülönbség, érdemes előre multigroupelemzést tervezni, és úgy gyűjteni mintát, hogy minden csoporthoz elegendő elemszám álljon rendelkezésre (ökölszabályként megfelelő arra törekedni, hogy minden elemzett alcsoportban legalább száz fő legyen).

2.2.2. Adatgyűjtés, az adatok előkészítése

Az adatgyűjtés tervezését érdemes úgy kivitelezni, hogy folyamatosan észben tartjuk a tervezett modellt.

A látens változókat mérő kérdések vagy indikátorok legyenek jól megfogalmazva, több tétellel lefedve. Egy zajos vagy pontatlan mérés sokkal erősebben rontja a SEM-et, mint egy egyszerű korrelációs elemzést. Érdemes validált kérdőíveket használni vagy azokat előzetesen validálni például pilotvizsgálattal. Ha mi állítjuk össze a tételeket, figyeljünk arra, hogy minden látens változóhoz legalább három tétel tartozzon, de érdemes ennél többet is tervezni, hogy akár a pilot-, akár a fővizsgálat alatt a rosszul mérő tételek törölhetők legyenek.

Külön figyelmet érdemes fordítani azokra a konstruktumokra, melyeket csak egy tétellel mérünk (például életkor). Ha ezek exogén változóként szerepelnek a modellben, akkor nem probléma, ha dichotóm, nominális vagy ordinális mérési szintűek, de ha endogén szerepben tervezzük vizsgálni őket, akkor skála vagy legalábbis kváziskála mérési szintűnek, például minimum öt, de inkább hét-tíz fokú Likert-típusú skálának kell lenniük. Ennek hiányában az általánosan elterjedt Maximum Likelihood (ML) paraméterbecslés helyett alternatív eljárást kell alkalmaznunk (például diagonally weighted least squares, DWLS).

Figyelni kell a hiányzó adatok kezelésére is. Kevés becslő eljárás tudja a hiányzó értékeket a modell részeként kezelni (ilyen például a Full Information ML, FIML eljárás), míg a robusztusabb becslő eljárások (például DWLS) a hiányzó értékek problémáját tipikusan listwise törléssel oldják meg. Az előző gondolatmenettel összekötve kifejezetten bosszantó helyzetbe kerülhetünk, ha néhány ordinális mérési szintű endogén változó miatt robusztus becslő eljárást kell választanunk, mely törli az összes olyan személyt a modellből, akinek akár egyetlen változóban is hiányzó értéke van. Tudnunk kell azt is, hogy az FIML is csak akkor kezeli jól a hiányzó értékeket, ha azok véletlenszerűek (nem szisztematikusak), és viszonylag kevés van belőlük. Végül meg kell említenünk egy harmadik ismert eljárást a hiányzó értékek kezelésére: a hiányzó értékek feltöltését, azaz az adatimputációt. Az imputáció többféle módszerrel működhet, de alapvetően kettő terjedt el: a hiányzó értékeket a skála átlagával helyettesítjük, vagy modern imputáló eljárással a személy többi adatából következtetünk rá. Az előbbi megoldás növeli a zajt a mérésben (mintha hibásan jelölt volna a személy), ezért csökkenti az összefüggések erősségét, utóbbi viszont (mivel a többi változóból becsüljük a hiányzó értéket) mesterségesen növeli az összefüggést a változók között. Természetesen kevés hiányzó adat esetében ezek a torzítások elfogadható mértékűek lehetnek. Összegezve, célszerű előre gondoskodni arról, hogy a kérdőív skálái, mérési módszerei megfelelőek legyenek (például sokszintű skálák, torzító hatások csökkentése). Ennek megfelelően célszerű a kérdőívet, mérési protokollt úgy kialakítani, hogy minél kevesebb kitöltetlen válasz maradjon (például kötelező mezők online kérdőívben).

Az adatfeldolgozás végén, a SEM-elemzés futtatása előtt át kell tekinteni a változók eloszlását. A CB-SEM (és főleg az ML becslő eljárás) erősen épít a normalitás feltételezésére. Ha a minta eloszlása torz, aszimmetrikus, vagy szélsőséges értékeket tartalmaz, az rontja az illeszkedést és a becslések megbízhatóságát. Kezelni kell a szélsőséges értékeket, és ha szükséges, transzformációval javítani az eloszláson, vagy robusztus becslési eljárást kell választani. Az ML-becslés feltételezi a többváltozós normalitást, amelyet a SEM-elemzésekben külön is ellenőrizhetünk. Ennek egyik legelterjedtebb módszere a Mardia-féle normalitásmutatók használata, amelyek a többváltozós ferdeséget és csúcsosságot vizsgálják. Ha ezek értékei szignifikánsan eltérnek a normális eloszlástól, az a feltételezés sérülésére utal. Ilyenkor a becslések torzulhatnak, a χ^2-próba és más illeszkedési mutatók pedig megbízhatatlanná válhatnak. A gyakorlatban ezért ilyen esetekben robusztus becslési eljárásokat (például Satorra–Bentler-korrekció, robust ML) vagy alternatív módszereket (például DWLS) szokás alkalmazni.

2.2.3. Az elemzés beállításai

22. doboz: A kovarianciaalapú (CB) SEM-elemzés szoftveres beállításai

A CB-SEM többféle statisztikai szoftverben elérhető. A leggyakrabban használt szoftvereket (Mplus, AMOS, az R lavaan csomagja, valamint az azt használó JASP és JAMOVI) a CFA- és PA-fejezetekben már ismertettük.

A SEM-elemzés mindig azzal kezdődik, hogy felrajzoljuk a tesztelendő modellt. Ahogyan a PA esetében is javasoltuk, ezt célszerű már az adatgyűjtés előtt megtenni, hogy a mérés során tudatosan biztosíthassuk a változók modellben betöltött szerepéhez szükséges tulajdonságokat, például a megfelelő mérési szintet. Ha a modell ábrája elkészült, el kell döntenünk, hogy az egyes változókat milyen szinten kezeljük. Egy változó lehet valódi manifeszt változó, vagyis olyan, amelyet egyetlen kérdéssel vagy méréssel rögzítettünk. Ilyenek többnyire a demográfiai jellemzők (például nem, életkor, iskolai végzettség), illetve a különböző csoportosító változók. Azokat a konstruktumokat viszont, amelyeket több tétellel mértünk, kétféleképpen is beépíthetjük: kezelhetjük őket látens változóként, vagy előzetesen kiszámított skálapontszám formájában is bevonhatjuk a modellbe. Ha kizárólag kiszámolt változókat használunk, akkor az elemzés tulajdonképpen a SEM egy speciális alesetét, az útvonalmodellezést (PA) jelenti, amelyet a korábbi fejezetben már részletesen bemutattunk. A gyakorlatban

minden változóról külön döntést hozhatunk, így egy modell tartalmazhat egyszerre számított változókat és látens konstruktumokat is. Általában célszerű számított változóként beilleszteni azokat a skálákat, amelyek soktételes, validált kérdőívekből származnak, hiszen ezek mérési struktúrája már ellenőrzött, és a sok tétel látens kezelésével csak feleslegesen növelnénk a modell komplexitását. Ezzel szemben indokolt látens változóként kezelni a rövidebb skálákat, illetve azokat a mérőeszközöket, amelyeket saját magunk állítottunk össze, és amelyeknél a mérési struktúra nem teljesen bizonyított. További szempont, hogy az exogén változókkal általában megengedőbbek lehetünk, és gyakran célszerűbb őket számított változóként szerepeltetni, hiszen ezek varianciáját a modell nem bontja megmagyarázott és meg nem magyarázott részre, így a bennük lévő mérési hiba kisebb problémát jelent. Az endogén változóknál azonban érdemes inkább törekedni arra, hogy látens formában épüljenek be, hiszen ezeknél különösen fontos a mérési hiba elkülönítése a tényleges konstruktumvarianciától.

Ezt követően a modell technikai beállítása következik, amelynek fő lépéseit a CFA és PA során már áttekintettük, de itt a teljes strukturális modell szempontjából kell őket egységesen megvalósítani. Először a modellt át kell ültetni a választott programba, amely során megadjuk a mért és látens változókat, utóbbiakhoz hozzárendeljük a mérési tételeket, valamint berajzoljuk a regressziós útvonalakat. Külön figyelmet igényelnek a reziduális kovarianciák. Az azonos szinten elhelyezkedő endogén változók között ezek általában megengedettek, de mindig indokoltan kell kezelni őket, eltérő szinten álló változók között (például egy prediktor és egy vele nem kapcsolatban lévő mediáló változó között) ritka a használatuk. A modellben szereplő közvetett hatásokat (indirekt útvonalakat) szintén deklarálni kell. A mérési modell beállításánál fontos döntés, hogy a látens változók varianciáját rögzítjük 1-re (elterjedtebb), vagy egy adott item töltését használjuk referenciaként.

Ezt követi a paraméterbecslési eljárás kiválasztása, amely a legtöbb esetben a Maximum Likelihood, de az eljárásnak több változata is elérhető, amelyek eltérő feltételezésekkel és robusztussággal rendelkeznek. A paraméterekhez kapcsolódó standard hibák becslése szintén kiemelt jelentőségű, hiszen ezek alapján állapítható meg a szignifikancia (beállításokat lásd: PA-fejezet).

Az eredmények között a standardizálatlan paramétereket általában automatikusan megjelenítik a programok, de célszerű mindig kikérni a standardizált paramétereket is, mivel ezek segítik az útvonalak összehasonlíthatóságát. Az illeszkedési mutatók kikérése szintén elengedhetetlen, mivel ezek jelzik, hogy a modell mennyire felel meg az empirikus adatoknak. A reziduálisok elemzése további támpontot ad arról, hogy hol tér el a modell az adatoktól, míg a modifikációs indexek a modell lehetséges továbbfejlesztésére adnak javaslatokat.

2.2.4. Az elemzés értelmezése

A SEM-elemzésekben elterjedt gyakorlat, hogy a teljes modell tesztelése előtt először kizárólag a mérési modellt vizsgáljuk. Ez azt jelenti, hogy az itemeket hozzárendeljük a látens változókhoz, de a látens változók közötti kapcsolatokat ekkor még nem regressziókkal, hanem csupán kovarianciákkal engedjük meg, és az összes látens változó között megengedjük a kovarianciát. E lépés célja annak ellenőrzése, hogy maga a mérési struktúra stabilan és megfelelően működik-e. Az illeszkedés javítása érdekében az ugyanazon látens változóra töltő tételek közötti reziduális kovariancia engedélyezhető, ha elméletileg indokolt az összefüggés. Ezzel szemben a kereszttöltéseket, illetve eltérő látens változók tételei közötti kovarianciát nem célszerű megengedni, mert ezek torzítanák a két látens közötti későbbi kapcsolat értelmezését. Ilyen esetben általában inkább az item törlését alkalmazzuk. Itt érdemes ellenőrizni még a látens változók érvényességét mérő mutatókat, mint az AVE vagy a HTMT (bár ezek a mutatók igazán kiemelt szerepet majd a PLS-SEM-nél kapnak). Ha a mérési modell illeszkedése megfelelő, akkor léphetünk tovább a strukturális rész beillesztésére.

A strukturális modell beépítését követően újra megvizsgáljuk az illeszkedési mutatókat. Ha romlik az illeszkedés, azt ekkor már többnyire inkább strukturális problémák okozzák, nem a mérési szerkezet. A modell változtatását megtehetjük elméleti alapon, vagy a javításához használhatunk modifikációs indexeket is, de a változtatásokat mindig elméletileg megalapozott módon kell végrehajtani. A módosítások lehetnek nem beágyazottak (amikor eltérő struktúrát próbálunk ki elméleti megfontolás alapján), illetve beágyazottak (nested modellek), amikor az eredeti modellhez egy-egy plusz útvonalat adunk hozzá. Fontos szabály, hogy egyszerre mindig csak egy útvonalat adjunk a modellhez, mivel egyetlen feszülés feloldása több illeszkedési problémát is orvosolhat (ahogy azt a PA-fejezetben részletesen tárgyaltuk). Arra is ügyelni kell, hogy ne közelítsük meg túlzottan a szaturált modellt, amelyben minden lehetséges kapcsolat szerepel, mert bár annak illeszkedése tökéletes lenne, kutatási szempontból érdemi információt nem hordoz ez a tökéletes illeszkedés.

Ha sikerült olyan modellt találni, amely megfelelően illeszkedik az adatokhoz, érdemes megvizsgálni még a reziduális varianciákat és az egyes endogén változók R^2-értékeit. Ezután térhetünk rá a regressziós együtthatók értelmezésére. Fontos megállapítani, melyek szignifikánsak, és melyek nem, milyen irányba mutatnak és mekkora erősségűek. A nem szignifikáns útvonalakat nem szükséges eltávolítani a modellből, hiszen azok a hipotézisek teszteléséhez fontos információt nyújtanak. Végül ellenőrizzük az indirekt hatásokat is, hiszen

a SEM egyik legnagyobb erőssége, hogy lehetővé teszi a direkt és indirekt útvonalak egységes értelmezését, amely a módszer valódi lényegét adja (szintén a PA-fejezetben tárgyaltuk).

2.3. A minőség biztosítása

A CB-SEM eredményeinek közlése során a kutató felelőssége, hogy világos, megismételhető és értelmezhető módon mutassa be a modellt és annak eredményeit. Az alábbiakban három szintet különítettünk el: minimálisan szükséges, ajánlott és kiváló közlés.
Szükséges közlés (minimálisan elvárt):

a) A kutató közli a mintanagyságot és az alapvető leíró statisztikákat.
b) A modellilleszkedéshez legalább 2-3 illeszkedési mutatót közöl (például χ^2, CFI, RMSEA).
c) Bemutatja a mérési modellt, a látens változókat és indikátoraikat pontos megnevezi, és közli a standardizált faktortöltéseket.
d) Bemutatja a strukturális modellt a standardizált útkoefficiensekkel és a szignifikanciaértékekkel.
e) Feltünteti a használt szoftver nevét.
f) Feltünteti a használt paraméterbecslő eljárás nevét.

Ajánlott közlés (jó gyakorlat) a szükséges közlésen felül:

a) A mintaméret és leíró statisztikák közlése mellett megtörténik a hiányzó adatok kezelésének módja, a normalitás és szélsőséges értékek vizsgálata.
b) Pontosan közli a teljes modell felépítését, a látens változók és a hozzá tartozó indikátorok pontos megnevezésével, a változók közötti útvonalegyüttható és reziduális kovarianciák közlésével.
c) a használt paraméterbecslő eljárás mellett a standard error becslő eljárását is megnevezi.
d) A modellilleszkedés többféle illeszkedési mutatóval történik (például χ^2, df, CFI, TLI, RMSEA, SRMR).
e) Indokolja a modellmódosításokat (például modification indices használata).
f) Megtörténik a mérési modell bemutatása a standardizált faktortöltésekkel és azok szignifikanciájával.

g) Megtörténik a strukturális modell bemutatása a standardizált és nem standardizált együtthatók közlésével (plusz SE, CI, p), valamint feltünteti az endogén változóknál a R^2-ek értékét is.

Kiváló közlés az ajánlott közlésen felül:

a) Részletesen bemutatja a mintavételi eljárást, vizsgálja a reprezentativitást. Előzetes teljesítményelemzést (power analysis) végez.
b) Az elemzés beállításait pontos közli, a paraméter és standard error becslésére használt eljárás megnevezésével és a mérési modell beállításainak közlésével.
c) Pontosan közli a teljes modell felépítését, a látens változók és a hozzá tartozó indikátorok pontos megnevezésével, a változók közötti útvonal-együtthatók és reziduális kovarianciák közlésével.
d) Szépen szerkesztett, olvasható modellábrát készít könnyen követhető betűméretekkel és elrendezéssel.
e) A modelllilleszkedés vizsgálatára többféle illeszkedési mutatót használ (például χ^2, df, CFI, TLI, RMSEA, SRMR), beleértve a parszimóniamutatókat (AIC, BIC).
f) A modellt összehasonlítja alternatív struktúrákkal és beágyazott modellekkel, és elvégzi az alternatív mérési modellek tesztelését (például konkurens faktorstruktúrák).
g) Multigroupelemzéseket végez a kutatás szempontjából releváns csoportok közötti inhomogenitások vizsgálatára.
h) Megtörténik a mérési modell bemutatása a standardizált faktortöltésekkel, többféle validitásvizsgálattal (például AVE, HTMT).
i) A strukturális modellben a közvetett hatások mellett bemutatásra kerülnek a mediációs és moderációs vizsgálatok is standardizált és standardizálatlan útvonal-együtthatókkal, SE- és p-értékkel, konfidenciaintervallummal.
j) A reprodukálhatóság és átláthatóság érdekében megosztja az adatokat és kódot (szintaxis), követi a nyílt tudományos elveket.
k) Publikációs kiegészítőket használ (például online függelékek) részletes táblázatokkal, ábrákkal.

3. A parciális legkisebb négyzetek módszerén alapuló SEM (PLS-SEM)

A parciális legkisebb négyzetek számításán alapuló strukturális egyenletmodellezés (Partial Least Squares Structural Equation Modeling – PLS-SEM) egy többváltozós statisztikai módszer, amelyet bonyolult összefüggések modellezésére használnak látens változók között. Gyökerei az 1970-es évekig nyúlnak vissza, amikor Herman Wold kifejlesztette a parciális legkisebb négyzetek (PLS) módszerét. Szélesebb körű elterjedése azonban csak a 2000-es években indult meg, amikor a számítási kapacitás növekedése és a felhasználóbarát szoftverek (például a SmartPLS) megjelentek, és az üzleti, menedzsment- és társadalomtudományi kutatásokban egyre inkább bevett gyakorlattá vált, de még ma sem ismert annyira, hogy módszertanának ismerete természetesen elvártnak tekinthető lehessen, ezért a tanulmányok módszertani részében részletesebb bemutatása szükséges. Továbbá amikor csupán SEM-et írunk, legtöbbször a CB-SEM-re gondolunk, míg a PLS-SEM-re hivatkozni általában teljes nevén szokás. Bár mindkét eljárás célja a (látens) változók közötti összefüggésrendszer vizsgálata, módszertanilag olyan mértékben eltérnek, hogy aligha tekinthetők azonos módszertan két alternatív megvalósításának. Az alábbiakban bemutatjuk a PLS-SEM elméletét mindenhol kitérve a CB-SEM-mel való hasonlóságokra és eltérésekre.

3.1. A PLS-SEM elmélete

A PLS-SEM is a változók közötti összefüggésrendszert vizsgálja, és lehetővé teszi, hogy ne csak közvetlenül mért változókat, hanem indikátorokon keresztül közvetetten mért látens változókat is a modellbe emeljünk.

Elsődleges célja a *predikció*, azaz a függő (endogén) változók varianciájának maximalizálása. Akkor ideális, ha a kutatás célja egy kulcsfontosságú konstruktum (látens változó) bejóslása vagy új, feltáró jellegű elméletek fejlesztése. A hangsúly a modell *prediktív erején* van. Ennek megfelelően a PLS-SEM egy predikcióorientált, feltáró (exploratív) módszertan, míg a CB-SEM egy konfirmáló (megerősítő) jellegű, elméletteszt elő eljárás.

Mindkét eljárás elkülöníti a mérési és strukturális modellt, de míg a CB-SEM esetében ez inkább elméleti jellegű szétválasztás, hiszen az eljárás egyben

teszteli a mérési és strukturális modell illeszkedését, addig a PLS-SEM esetében ez a szétválasztás a technikai megvalósításban is egyértelműen megjelenik.

A PLS-SEM paraméterbecslése egy kétfázisú iteratív algoritmussal történik: a külső és belső modell becslése váltakozik mindaddig, amíg a *konvergenciakritérium* nem teljesül. A konvergencia az a pont, ahol a további pontosítás már nem hoz jelentős változást az értékekben. Az első lépés a *külső (mérési) modell* becslése, mely során az eljárás a látens változók értékeit próbálja meghatározni az indikátorok súlyozott összegeként, és a súlyok állításának finomhangolása megtörténik. A második lépés a *belső (strukturális) modell* becslése, mely során az eljárás regressziós elemzések sorozatával becsli a látens változók közötti útvonal-együtthatókat. Az itt kapott eredmények adják a külső modell következő beállításának inputját. Lássuk e folyamatot algoritmikusan!

Tudjuk, hogy az algoritmusnak két dolgot kell meghatároznia: a külső súlyokat, melyek megadják, mennyire fontos egy-egy indikátor a saját látens változójának létrehozásában, és a belső útvonal-együtthatókat, melyek megmutatják, milyen erős a kapcsolat az egyes látens változók között. A probléma az, hogy az adatbázisunkban csak a manifeszt tételek léteznek, a látens változók nem. Ez utóbbiak értékének becsléséhez ismernünk kellene a külső súlyok értékét, de a külső súlyokat csak úgy tudjuk becsülni, ha már ismerjük a látens változók értékét. Ördögi kör. Ezt oldja fel az iteráció.

Az első lépés egy *kezdőérték megadása* (inicializálás). Az algoritmusnak el kell indulnia valahonnan, ezért fogja az összes indikátort, ad nekik egy kezdeti súlyt, és ezzel a kezdetleges súlyozással kiszámítja a látens változók első, durva becslését (pontszámát). A kezdeti súly általában minden indikátorhoz egy, ami megfelel annak, mintha a látens változókat egyszerűen az indikátoraik értékének összegzésével számolnánk. Végül a kapott skálapontszámokat standardizálja, mielőtt továbblépne a következő fázisba.

A második lépés a belső modell becslése. Most, hogy van egy (bár még pontatlan) becslésünk a látens változók értékeire, úgy tehetünk, mintha ezek valós, megfigyelt adatok lennének. Az algoritmus regressziós elemzésekkel megbecsüli a közöttük lévő kapcsolatok erősségét (a strukturális modell útvonal-együtthatóit, a β-értékeket).

A harmadik lépés a külső becslés frissítése (itt van az iterációs megközelítés lényege). Az algoritmus felhasználja a belső modellből származó információt a külső modell finomítására. De milyen információk vannak a belső modellben, amelyek segíthetnek finomhangolni a külső modellt? A második lépésben kiszámoltuk, hogyan hatnak egymásra a látens változók. Az algoritmus frissíti az egyes látens változók pontszámait a szomszédos, hozzájuk kapcsolódó látens változók figyelembevételével. Például ha A látens változóra erősen hat B és C,

akkor az A pontszámának új becslése már nemcsak a saját indikátorain fog alapulni, hanem B és C aktuális pontszámainak súlyozott összegén is.

Ezekkel a frissített, pontosabb látensváltozó-pontszámokkal az algoritmus újra megbecsüli a külső súlyokat (az indikátorok és a látens változójuk közötti kapcsolatot). Mivel a látens változók becslése megváltozott (pontosabb lett), a hozzájuk tartozó optimális indikátorsúlyok is változni fognak.

A negyedik lépés a konvergencia ellenőrzése. Az algoritmus összehasonlítja az újonnan számított külső súlyokat az előző iterációban számítottakkal. Ha a változás nagyon kicsi (egy előre meghatározott küszöbérték alatti), az azt jelenti, hogy a modell stabilizálódott, a súlyok már nem változnak érdemben. Ekkor az algoritmus leáll, és a végső súlyokkal és együtthatókkal számol. Ha a változás még jelentős, akkor az algoritmus visszaugrik a második lépésre, és a frissített külső súlyokkal és látensváltozó-pontszámokkal végzi el újra az egész folyamatot.

Ez az iteráló eljárás jelentősen eltér a CB-SEM módszertanától. Egy további jelentős eltérésről is kell még beszélnünk, mely a mérési modellt érinti. A CB-SEM-fejezetben azt tanultuk meg, hogy a látens változó és indikátorai közötti összefüggés úgy értelmezendő, hogy a látens változó fejeződik ki az indikátorokban, azaz a nyíl a látens változóból mutat az indikátorváltozók irányába. Ezt a mérési elrendezést *reflektív modellnek* nevezzük. A reflektív modellben azt feltételezzük, hogy létezik egy mögöttes, igazi látens változó, és az indikátorok e látens változó „tükröződései" (innen jön az elnevezés), azaz a konstruktum okozza a változást az indikátorokban. Ekkor az indikátorok egymással szorosan korrelálnak és felcserélhetők. Például a stressz kifejeződik olyan állításokban, mint hogy „Túlterheltnek érzem magam." vagy „Gyakran vagyok feszült.". A PLS-SEM is támogatja a reflektív modellek felállítását, azonban emellett lehetőség van annak feltételezésére is, hogy az indikátorok „formálják", hozzák létre a látens változót. Ezt nevezzük *formatív modellnek*. Ekkor az indikátorok mutatnak a látens változó irányába, és az indikátoroknak nem kell egymással korrelálniuk, és nem felcserélhetők. Például a stressz konstruktumot a munkahely elvesztése, a költözés és a válás indikátorai is okozzák (lásd 23. ábra), ahol természetesen nem várjuk azt, hogy aki elveszíti a munkahelyét, annak szükségszerűen válnia is kelljen. Összegezve, a PLS-SEM rugalmasan kezeli mind a reflektív, mind a formatív mérési modelleket, míg a CB-SEM leginkább a reflektív mérési modellekre épít, és formatív modell csak jelentős korlátozásokkal hozható létre benne.

23. ábra: Példa reflektív és formatív mérési modellre

Mit jelent ez az algoritmus szempontjából? A PLS-SEM-algoritmus mindkét esetben az indikátorok súlyozott összegeként, azaz kompozitként számítja ki a látens változó értékét (pontszámát), számára a látens változó mindig egy számított érték, egy proxy. A különbség nem a látens változó létrehozásának módjában, hanem az indikátorok súlyainak kiszámításában van a reflektív és formatív modellek között. A reflektív modellek esetén az algoritmus úgy számolja a súlyokat, hogy azok az indikátorok és a konstruktum közötti korrelációt tükrözzék. Azt feltételezi, hogy az indikátorok szorosan összetartoznak, és a közös varianciájuk a fontos. Ezzel szemben a formatív modellek esetén az algoritmus egy többszörös regressziót futtat, ahol a látens változó a függő, az indikátorok pedig a független változók. Itt az a lényeg, hogy az egyes indikátorok milyen egyedi mértékben járulnak hozzá a konstruktumhoz.

Látható, hogy a PLS-SEM iteratív, regresszióalapú algoritmusa a variancia maximalizálására törekszik, és nem a globális illeszkedésre. Nem az adatok kovarianciastruktúrájának teljes modellel való reprodukálására optimalizál. Ennek megfelelően a modell értékelésének mutatói is eltérnek a CB-SEM kapcsán megismertektől. A PLS-SEM-modell értékelése a prediktív képességen alapul. A legfontosabb mutatók a modell értékelése során a következők: A mérési modellben mindig ellenőrizni kell az indikátorok megbízhatóságát. Ezenfelül a reflektív mérési modellben ellenőrizendő a konstruktumok belső konzisztenciája, valamint a modell validitása. Formatív mérési modellnél a multikollinearitást kell ellenőrizni. A strukturális modell értékelésének alapjai a függő változók magyarázott varianciája, az útvonal-együtthatók nagysága és szignifikanciája, valamint a prediktív relevancia és a hatásméret.

Lássuk ezeket részletesebben!

Az *indikátormegbízhatóság* (Indicator Reliability) azt mutatja meg, hogy egy egyetlen, különálló indikátor (például egy kérdőíves kérdés) varianciájának mekkora részét magyarázza a látens konstruktum, amelyhez tartozik.

Mérőszáma az *indikátor külső töltésének négyzete* (squared outer loading, λ^2). Ennek értéke 0 és 1 között lehetséges, és ökölszabály szerint 0,50 felett kell lennie ahhoz, hogy a látens változó megfelelő indikátorának tekintsük. Ez azt jelenti, hogy a konstruktum az indikátor varianciájának több mint felét magyarázza, azaz a mérési hiba kevesebb, mint 50%. A gyakorlatban a kutatók sokszor nem a négyzetes értéket, hanem magát a külső töltést adják meg (λ), ekkor a határérték 0,50 gyöke, azaz $\lambda > 0{,}708$. A megengedőbb szakirodalom meghúz még egy határt a $\lambda = 0{,}40$-nál (ami 16% megmagyarázott varianciának felel meg), amely felett még mérlegelhető a tétel megtartása, ez alatt viszont a tételt egyértelműen törölni kell.

A *konstruktumszintű megbízhatóság* (Construct Reliability) a reflektív mérési modellben azt vizsgálja, hogy az indikátorok csoportja együttesen, egységként mennyire konzisztensen méri az adott látens konstruktumot. Ezt nevezzük *belső konzisztenciá*nak. Tehát nem egyenként, hanem csapatként vizsgáljuk az indikátorokat. Mérőszámai hagyományosan a Cronbach-alfa, ennek modernebb változata a McDonald-omega, de napjainkban ajánlottabb a kompozit megbízhatóság és a Dijkstra–Henseler-féle ρ_A számolása.

A *Cronbach-alfa* (Cronbach's α) a belső konzisztencia hagyományos mérőszáma. Azt vizsgálja, hogy a konstruktumot mérő indikátorok mennyire függenek össze egymással. Az elfogadott érték általában 0,70 felett van, de feltáró kutatásoknál a 0,60 is megengedett. A *McDonald-omega* (ω) a megbízhatóság egy modernebb és sokak szerint pontosabb mérőszáma. A PLS-SEM-ben az alfánál és omegánál javasoltabb mutatók a kompozit megbízhatóság és a Dijkstra–Henseler-féle ρ_A használata.

A *kompozit megbízhatóság* (vigyázzunk, több eltérő néven is hivatkozni szokás rá: Composite Reliability, CR, ρ_c vagy rho_C) az előző kettőhöz hasonlóan a belső konzisztenciát méri, de a PLS-SEM módszertanában előnyben részesítik, mert nem feltételezi, hogy minden indikátor egyformán jól méri a konstruktumot, hanem figyelembe veszi az egyes indikátorok eltérő súlyát (töltését). Az elfogadott érték itt is 0,70 vagy afeletti.

A *Dijkstra–Henseler-féle ρA* (vagy rho_A) egy megbízhatósági mutató, amelyet kifejezetten a PLS-SEM-modellekhez fejlesztettek ki. Pontosságban a Cronbach-alfa és a kompozit megbízhatóság között helyezkedik el, és kifejezetten a reflektív mérési modellek belső konzisztenciájának mérésére használják. Előnye, hogy kevésbé torzít, mint a Cronbach-alfa, és megbízhatóbb eredményt ad kis minták vagy eltérő faktorsúlyok esetén is. Szintén 0,70 felett tekintjük elfogadhatónak.

Szintén a reflektív modellek ellenőrzését szolgálja a konstruktum *konvergens validitása* (Convergent Validity), amely azt mutatja meg, hogy egy adott

konstruktum indikátorai valóban összetartoznak-e, azaz egyazon mögöttes jelenséget mérnek-e. A legfontosabb mérőszáma az *AVE* (Average Variance Extracted), vagyis az átlagos magyarázott variancia, amely azt fejezi ki, hogy a konstruktum varianciájának mekkora részét magyarázzák az indikátorai. Az elfogadott küszöbérték 0,50, ami azt jelenti, hogy az indikátorok átlagosan több mint 50%-ban magyarázzák a konstruktum varianciáját.

A *diszkrimináns validitás* (Discriminant Validity) azt biztosítja, hogy a modellünkben szereplő egyes konstruktumok valóban eltérő jelenségeket mérnek, és nem olvadnak-e össze. Tehát például a „vásárlói elégedettség" és a „márkahűség" két külön dolog-e a modellünkben. A *HTMT* (Heterotrait-Monotrait Ratio) a diszkrimináns validitás modern és megbízható mérőszáma. Azt vizsgálja, hogy a különböző konstruktumok indikátorai közötti (heterotrait) korrelációk mekkorák az azonos konstruktumhoz tartozó indikátorok közötti (monotrait) korrelációkhoz képest. Ha HTMT = 1, annak az lenne a jelentése, hogy a különböző konstruktumhoz tartozó tételek között ugyanolyan erős a korreláció, mint az azonos konstruktumhoz tartozó tételek között, tehát a konstruktumok nem válnak szét. Minél kisebb a HTMT értéke, annál jobban szétválnak a konstruktumok, mert a különböző látens változókhoz tartozó tételek korrelációi eltörpülnek az azonos látens változóhoz tartozó tételek közötti összefüggéshez képest. Az értéke 0,90 alatt kell hogy legyen (koncepcionálisan nagyon hasonló konstruktumok esetén), de a szigorúbb kritérium a 0,85 alatti érték.

A formatív modell értékelésére szolgál a *VIF* (Variance Inflation Factor), amely a multikollinearitást méri, megmutatja, hogy egy indikátor mennyire korrelál más indikátorokkal a modellben. Fontos különbség, hogy míg a reflektív modellnél az indikátorok közötti magas korreláció volt az elvárt, a formatív modellekben éppen az alacsony korreláció preferált, mert a magas multikollinearitás torzíthatja a formatív indikátorok súlyait és a strukturális utak becslését. Ezért a VIF-értékek ellenőrzése segít biztosítani a modell stabilitását és az eredmények megbízhatóságát. Elfogadható, ha VIF < 5, de ideális esetben VIF < 3.

Ha a mérésünk megbízható és érvényes, rátérhetünk a strukturális modell értékelésére, azaz a modellben felrajzolt ok-okozati nyilak (hipotézisek) vizsgálatára. Ennek mutatói a determinációs együttható, az útvonal-együtthatók mutatói, a hatásméret-mutatók és a prediktív relevancia. Lássuk ezeket is részletesen!

A *determinációs együtthatók (R^2-értékek)* azt mutatják meg, hogy a modellben egy adott függő (vagy más néven endogén) konstruktum varianciájának hány százalékát magyarázzák a rá ható konstruktumok. Például egy $R^2 = 0,65$ a „vásárlási szándék" konstruktumon azt jelenti, hogy a modellünk 65%-ban magyarázza, hogy miért alakul ki a vásárlási szándék. Az elfogadott értékek

szakterületfüggőek, de általában erősnek tekinthető 0,75 felett, közepesnek 0,50-0,75 között, gyengének 0,25-0,50 között, míg 0,25 alatt elhanyagolhatónak.

Az *útvonal-együtthatók* (Path Coefficients, β) a modell „nyilainak" az erősségét és irányát mutatják. Értelmezésük megegyezik a regressziós elemzések standardizált együtthatójánál megszokottal, illetve a CB-SEM esetében megtanulttal.

Önmagában a β-érték nagysága nem elég. Tudnunk kell, hogy ez a hatás statisztikailag szignifikáns-e, vagyis nem csak a véletlen műve. Mivel a PLS-SEM nem tesz feltevéseket az adatok eloszlásáról, a *szignifikancia* megállapításához a *bootstrapping* eljárást használja. Ez egy újra-mintavételezési technika, ahol az eredeti mintából véletlenszerűen, visszatevéssel több ezer (például 5000) új mintát hoz létre. Ezeken a mintákon újra és újra lefuttatja a modell becslését, és az így kapott β-értékek eloszlása alapján megbecsüli a p-értéket. Ha a p-érték egy előre meghatározott szint (jellemzően 0,05) alatt van, akkor a hatás szignifikánsnak tekinthető.

Következő mutatóként a *hatásméret* (f^2) azt méri, hogy egy adott független konstruktum mekkora hatással van egy függő konstruktum R^2-értékére. Tehát ha kivennénk a modellből az adott független változót, mennyivel csökkenne a magyarázóerő. Itt 0,02 és 0,15 között kis hatásról beszélünk, 0,15 és 0,35 között közepes hatásról, 0,35 felett nagy hatásról.

Végül a PLS-SEM nagy hangsúlyt fektet a modell előre jelző képességére, azaz a *prediktív relevanciára*, amelyet hagyományosan a *Stone–Geisser-féle* Q^2-értékkel szokás mérni. Ez azt mutatja, hogy a modell mennyire jól képes előre jelezni egy függő konstruktum értékeit. Ezt egy blindfolding nevű eljárással számolják, amely során az adatpontok egy részét kihagyják, majd a modell megpróbálja megjósolni a kihagyott értékeket. Ha a Q^2 értéke 0-nál nagyobb, akkor a modellnek van prediktív relevanciája az adott konstruktumra nézve. Értelmezése a következő: ha $Q^2 > 0$, akkor van prediktív relevancia, amely 0,25-ig gyenge, 0,50-ig közepes, 0,50 felett erősnek tekintendő.

Az előbb bemutatott Stone–Geisser-féle Q^2 egy úgynevezett hagyományos, *in-sample mutató*, mely a modell magyarázóerejét vizsgálja a becsült adatokon belül, vagyis azt, hogy a független változók mennyire képesek előre jelezni az endogén konstrukciókat a meglévő mintán. Emellett léteznek olyan kiegészítő eljárások, amelyek az adatok egy részét *out-of-sample* (tesztmintaként) kezelik, és így azt értékelik, hogy a modell mennyire képes új, még nem látott megfigyelések előre jelzésére. Ilyen eljárás a *PLSpredict*, amelynek során az out-of-sample Q^2*predict*-mutató kerül kiszámításra, amely a valódi prediktív teljesítményt méri a tesztadatokon.

Összegezve a PLS-SEM esetében a fókusz a megmagyarázott variancia és a prediktív erő maximalizálásán van, és a modellt nem a valósághoz való illeszkedése mentén értékeli, mint ahogy azt a CB-SEM-ben megismertük. Bár napjainkban egyre-másra jelennek meg olyan mutatók, amelyek ötvözik a két eljárás módszertanát, például az SRMR PLS-re adaptált verziója.

3.2. PLS-SEM kutatási folyamata

A PLS-SEM elméletének bemutatása után most rátérünk a módszer technikai folyamatának ismertetésére.

3.2.1. Mintaválasztás

A PLS-SEM-nek lényegesen enyhébbek az előfeltételei. Kisebb (akár 100 fős) mintákkal is jól működik, bár a tízszeres szabály (az útvonaldiagramban egy látens változóra mutató nyilak száma szorozva tízzel) itt is gyakori iránymutatás. Ezért a PLS-SEM jól használható, ha az adatgyűjtés korlátozott, például mert nehezen elérhető a vizsgált populáció.

3.2.2. Adatgyűjtés, az adatok előkészítése

A PLS-SEM nem igényli a többváltozós normalitást, és nem érzékeny a ferde vagy csúcsos eloszlásokra, azaz ez egy eloszlásfüggetlen, nem parametrikus módszer. A kisebb mennyiségű hiányzó értékeket jól kezeli imputációval vagy középértékpótlással. Bár a PLS-SEM kisebb mintán is működik, az *adatminőség* itt is kritikus. A torzított, zajos vagy mérési hibákkal teli adatoknál az eredmények itt is gyengülnek. Kevésbé érzékeny szélsőséges értékekre, de a multikollinearitás problémát jelenthet (különösen formatív modelleknél).

3.2.3. Az elemzés beállításai

23. doboz: A varianciaalapú (PLS) SEM-elemzés szoftveres beállításai

Az elemzésekhez kifejezetten PLS-SEM-re fejlesztett és általános célú szoftverek egyaránt elérhetők. A célzott eszközök közül a SmartPLS a legelterjedtebb, felhasználóbarát grafikus felülettel. A WarpPLS különösen nemlineáris kapcsolatok modellezésére alkalmas, míg az ADANCO korszerű, intuitív környezetet biztosít. A PLS Graph mint korábbi generációs eszköz mára háttérbe szorult. Az elemzés általános statisztikai környezetekben is megvalósítható: R-ben (plspm, seminr, cSEM csomagok), Pythonban (pyPLS), valamint a JASP nyílt forráskódú szoftverben.

A PLS-SEM-elemzés futtatása is a modell felrajzolásával történik. Ennek szabályai jelentős részben megegyeznek a CB-SEM-fejezetben ismertetett ajánlásokkal. A PLS-SEM-re speciálisan jellemző a látens változók formatív vagy reflektív modellként való meghatározása. A formatív és reflektív mérési modell közötti döntés elsősorban a konstrukció és a mérőtételek elméleti természetétől függ. A reflektív modell akkor indokolt, ha a tételek egy mögöttes, közös látens változó megnyilvánulásai, vagyis a tételek erősen korrelálnak, felcserélhetők, és elhagyásuk nem változtatja meg a konstrukció jelentését. A formatív modell akkor megfelelő, ha a tételek együtt formálják a konstrukciót, azok nem feltétlenül korrelálnak, és egy-egy indikátor elhagyása tartalmilag módosítaná a változó értelmét. Egyazon SEM-modell tartalmazhat olyan látens változókat is, amelyeket formatívan, másokat reflektíven emeltünk a modellbe.

A mérési és strukturális modell helyes megadását követően meg kell adnunk a becslés technikai beállításait. A PLS-SEM iteratív algoritmussal működik, így szükséges a maximális iterációk számának (például 300–500) és a konvergenciakritériumnak (például 0,0000001) a beállítása, ami biztosítja, hogy a modell stabil megoldáshoz jusson. Emellett minden szoftverben megadható a bootstrap eljárás beállítása, amely a paraméterek szignifikanciájának vizsgálatához szükséges. Ilyenkor a kutató kiválaszthatja a minták számát (általában 5000 vagy több javasolt), valamint azt, hogy a bootstrap konfidenciaintervallum aszimmetrikus vagy percentilis módszerrel készüljön.

Be kell állítania a súlyozási sémát (weighting scheme), amely a látens változók és indikátoraik kapcsolatát becsli. A leggyakrabban alkalmazott beállítás a path weighting scheme, de bizonyos helyzetekben a centroid vagy a factor weighting scheme is használható. Az ezek közötti választás általában kevésbé befolyásolja az eredményeket, de nagyobb modelleknél és formatív konstrukciók esetén érdemes körültekintően dönteni.

Végül dönteni kell a hiányzó értékek kezeléséről (például listwise vagy pairwise törlés, illetve imputáció).

A modell értékeléséhez be kell állítani a megfelelő mutatókat. A mérési modell értékelése reflektív konstrukcióknál: indikátormegbízhatóság (külső súlyok), konstrukciómegbízhatóság (Cronbach-alfa, McDonald-omega, ρ_A és ρ_c), konvergens (AVE) és diszkriminánsz (HTMT) validitás. A mérési modell értékelése formatív konstrukcióknál: indikátormegbízhatóság (külső súlyok), multikollinearitás ellenőrzése (VIF-érték). A strukturális modell értékelése: útvonal-együtthatók (path coefficients) és szignifikanciájuk (bootstrap alapján). Magyarázott variancia (R^2) az endogén konstrukcióknál, hatásméretek (f^2), prediktív relevancia (Q^2 és Q^2predict). Esetleg a modellilleszkedési mutatók (például SRMR), bár ez kevésbé hangsúlyos PLS-ben, mint CB-SEM-ben.

Ezek az értékelési szempontok minden PLS-SEM-elemzés alapját képezik, és a legtöbb szoftver automatikusan biztosítja a hozzájuk tartozó outputokat.

3.2.4. Az elemzés értelmezése

A PLS-SEM-modell értékelése egy szisztematikus, kétlépcsős folyamat, amely során először a mérés megbízhatóságát és érvényességét, majd a feltételezett belső struktúrát vizsgáljuk. A cél egy olyan modell létrehozása, amely nemcsak jól magyarázza a meglévő adatokat, hanem megbízható előrejelzésekre is képes.

Az elemzés mindig a mérési modell ellenőrzésével kezdődik, hiszen a strukturális modell hipotéziseit csak akkor érdemes tesztelni, ha a mérőműszereink, vagyis a konstruktumainkat alkotó indikátorok pontosak.

Az első lépés a reflektív modellek esetében az egyes indikátorok minőségének felmérése a külső töltések alapján. Azokat a tételeket, amelyek töltése nem éri el a 0,708-as küszöbértéket, jellemzően eltávolítjuk a modellből, mivel ez azt jelenti, hogy a hozzájuk tartozó látens változó a varianciájuknak kevesebb mint 50%-át magyarázza. Emellett bootstrapping eljárással ellenőrizzük, hogy a töltések statisztikailag szignifikánsak-e. A nem szignifikáns indikátorok szintén törlésre kerülnek. Formatív mérési modellek esetén más a logika, ott a multikollinearitást vizsgáljuk, és azokat a tételeket töröljük, ahol a VIF értéke túl magas, jellemzően VIF > 5.

Miután a gyengén teljesítő indikátorokat eltávolítottuk, a reflektív modelleknél konstruktumok szintjén is ellenőrizzük a megbízhatóságot és az érvényességet. A konvergens validitás azt vizsgálja, hogy az egy konstruktumhoz tartozó indikátorok valóban összetartoznak-e. Ennek igazolásához a Cronbach-alfa, a McDonald-omega, valamint a ρ_A és a ρ_c értékének 0,70 felett, az AVE

értékének pedig 0,50 felett kell lennie. Amennyiben az értékek nem megfelelők, további indikátor(ok) törlése lehet indokolt. Jellemzően a legalacsonyabb külső töltéssel rendelkező tételt távolítjuk el, majd újraértékeljük a mutatókat.

Még mindig a reflektív modelleknél maradva, ezt követően a diszkrimináns validitás vizsgálata következik, amely azt garantálja, hogy a modellben szereplő konstruktumok valóban eltérő fogalmakat mérnek. Ennek legmodernebb mérőszáma a HTMT, amelynek értéke a szigorúbb kritérium szerint 0,85 alatt kell hogy maradjon. Ha két konstruktum között az HTMT-érték túl magas, az erős átfedésre utal. Ilyenkor meg kell fontolni az egyik skála elhagyását, vagy ha elméletileg indokolt, a két konstruktum összevonását egyetlen, magasabb rendű változóvá.

Amint a mérési modell megbízhatónak és érvényesnek bizonyult, rátérhetünk a strukturális modell elemzésére, vagyis a feltételezett összefüggések tesztelésére. Először a regressziós útvonal-együtthatókat (β) és azok szignifikanciáját vizsgáljuk, amelyet a bootstrapping eljárás szolgáltat. Itt azt nézzük, hogy a feltételezett kapcsolatok (a nyilak) statisztikailag szignifikánsak-e (jellemzően $p < 0,05$), illetve hogy az együttható iránya és nagysága megfelel-e az elméleti várakozásainknak. Ezzel párhuzamosan értékeljük az R^2-értékeket, amelyek a modell magyarázóerejét mutatják meg az egyes függő (endogén) változókra. A hatások relatív fontosságának felmérésére az f^2-et használjuk, amely megmutatja, hogy egy adott független változó mekkora mértékben járul hozzá a függő változó R^2-értékéhez. Az elemzés során a direkt és indirekt útvonalakat is megvizsgáljuk, hogy feltárjuk a mediációs (közvetítő) hatásokat, és megértsük a változók közötti teljes kapcsolatrendszert, valamint szakmai és elméleti szempontból értékeljük az eredményeket.

Végül a modell prediktív, azaz előre jelző képességét értékeljük. A Stone–Geisser-féle Q^2-mutató a mintán belüli prediktív relevanciát méri, ha az értéke nullánál nagyobb, a modellnek van előre jelző képessége. A modern PLS-SEM-analízis csúcspontja a PLSpredict-eljárás, amely mintán kívüli előrejelzést ad. Itt azt vizsgáljuk, hogy a modellünk előre jelzési hibája alacsonyabb-e, mint egy egyszerű lineáris modellé. Ha a PLS-modell a legtöbb változó esetében pontosabb előrejelzést ad, akkor a modell prediktív ereje bizonyítottnak tekinthető a populációra nézve is.

3.3. A minőség biztosítása

A PLS-SEM közlése során az alábbi kritériumok határozzák meg a közlés minőségét.

Szükséges közlés (minimálisan elvárt):

a) A kutató bemutatja a modell vázát, azaz a látens és manifeszt változók neveit, valamint ezek kapcsolatát.
b) A modellt validálja néhány mutató segítségével (például mérési modellhez ρ_c, AVE).
c) Megadja az útvonal-együtthatókat és szignifikanciájukat (esetleg csak jelentős hatások feltüntetése).
d) Megadja az R^2-értékeket a fő kimeneti (endogén) változókhoz.

Ajánlott közlés (jó gyakorlat) a szükséges közlésen felül:

a) Leírja a teljes modellt és/vagy ábrázolja azt (például path-diagram), a látens és megfigyelt változókkal, az összefüggések irányaival.
b) Bemutatja a konstruktumvaliditás és megbízhatóság mutatóit (például ρ_c, AVE-értékek, VIF).
c) Bemutatja a diszkriminánsvaliditást (például HTMT-index).
d) Részletesen közli az útvonal-együtthatókat, ami magában foglalja az összes strukturális együtthatót (β), standard hibákat (SE), p-értékeket és konfidenciaintervallumokat.
e) Közli az R^2-értékeket minden endogén változó esetén és esetleg a Q^2-értéket is.
f) Tárgyalja a módszertani információkat: mintaméret, bootstrap-szimulációk beállításai (például futtatások száma), szoftver és verzió megnevezése.

Kiváló közlés az ajánlott közlésen felül:

a) Leírja a teljes mérési modellt, és indokolja, mely látens változót tekintette reflektívnek, melyet formatívnak.
b) Leírja a teljes strukturális modellt, bemutatja az útvonalakat és indokolja őket.
c) Szépen szerkesztett, olvasható modellábrát készít könnyen követhető betűméretekkel és elrendezéssel.
d) A mérési modell értékelésekor reflektív változók esetén értékeli a külső töltéseket, bemutatja a validitási és megbízhatósági mutatók széles körét (Cronbach-alfa, McDonald-omega, ρ_c, ρ_A, AVE), és ha a mutatók alapján törlés történt, részletesen ismerteti annak körülményeit.
e) A mérési modell értékelésekor formatív változók esetén értékeli a külső töltéseket, ellenőrzi a VIF-értékeket, és ha a mutatók alapján törlés történt, részletesen ismerteti annak körülményeit.
f) Ellenőrzi a diszkriminánsvaliditást (HTMT), és ha két változó túlzott átfedését mutatta ki, a probléma kezelésének lépéseit (például összevonás).

g) Strukturális modellnél teljes részletességre törekszik, azaz megjelenik minden útvonal-együttható standardizált és standardizálatlan formában, valamint közlésre kerülnek az az SE-, a p-érték, a CI- és a bootstraperedmények.
h) A direkt útvonalak mellett az indirekt útvonalakat is részletes ismerteti.
i) Minden R^2-, f^2-, Q^2-értéket, Q^2predict-értéket részletez és értelmez.
j) Esetlegesen illeszkedési mutatókat is bemutat (például SRMR).
k) Megtörténik a robusztussági vizsgálat alternatív modellek tesztelésével és multigroupelemzéssel.
l) Megtörténik az érzékenységi elemzés és a kontrollváltozók hatásának bemutatása.
m) A módszertani korlátokat átgondolja, valamint további kutatási kérdések vet fel.
n) Törekszik a transzparenciára és megismételhetőségre, ezért részletesen adat- és kóddokumentálást végez akár adatbázisra mutató linkkel.

4. Mikor használjunk CB-, és mikor PLS-SEM-et?

A strukturális egyenletmodellezés (SEM) két legelterjedtebb megközelítése, a kovarianciaalapú (CB-SEM) és a varianciaalapú (PLS-SEM) módszertan nem egymás riválisai, hanem inkább két különböző célt szolgáló, egymást kiegészítő eszközök a kutatók kezében. A köztük való választás nem azon múlik, hogy melyik a „jobb", hanem azon, hogy mi a kutatás elsődleges célja, milyen az elméleti háttér, és milyenek a rendelkezésre álló adatok.

A CB-SEM alapvetően egy megerősítő, elmélettesztelő eszköz. Akkor ideális a használata, ha a kutatónak van egy erős, jól megalapozott elméleti modellje, és a célja az, hogy ezt a modellt a valósággal, vagyis a mintából származó adatokkal összevesse. A CB-SEM a modellilleszkedés pontos mérésére fókuszál, olyan mutatókkal (például CFI, RMSEA), amelyek megmutatják, hogy a feltételezett modell mennyire képes reprodukálni az adatokban megfigyelt kovarianciamátrixot. Vagyis a fő kérdése: „Az elméletem összhangban van az adatokkal?" Emiatt szigorúbb feltételeket támaszt az adatokkal szemben, ez jellemzően nagyobb mintaméretet és a normális eloszlás feltételezhetőségét igényli.

Ezzel szemben a PLS-SEM egy predikcióorientált, elsősorban feltáró jellegű eszköz. Akkor a legjobb választás, ha a kutatás fő célja nem egy meglévő elmélet igazolása, hanem a függő változó(k) varianciájának minél jobb előre jelzése és a legfontosabb prediktorok azonosítása. Kiválóan alkalmas komplex modellek kezelésére, ahol sok a konstruktum és az indikátor, vagy ha az elméleti háttér még kiforratlan. A PLS-SEM sokkal megengedőbb az adatokkal szemben, jól működik kisebb mintákon, nem normális eloszlású változókkal, és a formatív mérési modelleket is képes kezelni. Itt a fő kérdés nem az illeszkedés, hanem az, hogy: „Mekkora a modell előre jelző ereje, és melyek a kulcsfontosságú kapcsolatok?"

Összefoglalva, a választás a kutatási célon múlik: ha egy szilárd elméletet szeretnénk a lehető legpontosabban tesztelni, és a modell egészének illeszkedése van a fókuszban, a CB-SEM a helyes út. Ha viszont a cél a prediktív pontosság maximalizálása, egy új elmélet feltárása vagy bonyolultabb modellek elemzése kevésbé ideális adatokon, akkor a PLS-SEM nyújtja a hatékonyabb megoldást. A két módszer tehát nem verseng, hanem a kutatási spektrum két különböző, de egyaránt fontos területét fedi le.

5. A módszer változatai és haladó alkalmazásai

Maga a strukturális egyenletmodellezés egy rendkívül rugalmas keretrendszer, amelyen belül számtalan alkalmazás létezik, amelyek a különböző minták és vizsgálatok eltérő tulajdonságaira optimalizálják az eljárást. Ebben a fejezetben a CB-SEM legelterjedtebb kiegészítőelemzését, a *multigroupelemzést* (MGA) mutatjuk be.

Az MGA a CB-SEM egyik fontos kiterjesztése, amely lehetővé teszi, hogy különböző csoportok (például nem, korosztály, kulturális háttér) strukturális modelljeit összehasonlítsuk. Ez a módszer különösen hasznos akkor, ha a kutatót az érdekli, hogy azonos elméleti modell eltérően működik-e különböző mintákban. Ily módon a multigroupelemzés a moderációs vizsgálatok egyik alternatívája: ahelyett, hogy egyetlen mintában tesztelnénk egy (általában folytonos) moderáló változó hatását, itt explicit módon csoportokra bontjuk a vizsgált mintát, és összehasonlítjuk e csoportokat a mérési és strukturális paraméterek mentén.

A multigroupelemzés logikája a következő. Tegyük fel, hogy az MG-elemzés csoportosító változójának két kategóriája van, például férfiakat és nőket hasonlítunk össze egymással. A multigroupelemzés során arra vagyunk kíváncsiak, hogy ugyanaz a SEM-modell ugyanazon paraméterekkel vajon használható-e mind a férfiak, mind a nők mintáján, vagy a két mintafélen eltérő modell illeszkedik jól. Ehhez először külön illesztjük a SEM-modellt a férfi és női mintafélen. Maga a modell ugyan azonos felépítésű (ugyanazokat a látens változókat tartalmazza, a látens változókhoz azonos tételek tartoznak, és ugyanazokat az útvonalakat becsüljük a strukturális modellben), de a paraméterek értékei eltérhetnek. Vagyis megengedjük, hogy mások legyenek a férfi és női minta töltésértékei, útvonalegyüttható-értékei, reziduálisai, reziduális kovarianciái, a manifeszt változók átlaga és szórása, a látens változók átlaga és szórása stb. Ha ez a modell jól illeszkedik, akkor egyesével megkötjük a különböző paramétereket. Akár ténylegesen egyesével is meg lehet kötni minden egyes becsült paramétert, de a gyakorlatban nem ez szokott történni, hanem például megkötjük a töltésértékeket, ezáltal arra kényszerítjük a modellt, hogy férfi és női mintán azonos töltéseket alkalmazzon a látens változók becslése során, míg a többi még meg nem kötött paraméter továbbra is eltérhet a két mintafél között. A töltések megkötése természetesen csökkenteni fogja az illeszkedést, de ha amúgy sem volt nagy különbség férfiak és nők mintájában a töltések között, akkor ez a csökkenés az illeszkedésben elhanyagolható lesz. Ekkor azt mondjuk, hogy a két mintafél *invariáns* a töltések szempontjából, és továbblépve megkötjük a paraméterek egy újabb csoportját.

Az *invarianciavizsgálat* során a modellek illeszkedését összehasonlítva dönthetünk arról, hogy az adott szint elfogadható-e. Ehhez nem kizárólag a χ^2-különbség-tesztet érdemes használni, mivel az erősen érzékeny a mintanagyságra. Ehelyett gyakran alkalmazzák az illeszkedési mutatók változásának kritériumait: a CFI csökkenése nem haladhatja meg a 0,01 értéket, míg az RMSEA növekedése nem lehet nagyobb 0,015-nél két modell között. Ha ezek a kritériumok teljesülnek, az adott invarianciaszint igazoltnak tekinthető.

Annak, hogy mikor melyik paramétercsoportot kell megkötni, kidolgozott módszertana van. Először a mérési modell invarianciáját kell ellenőrizni, vagyis azt, hogy a mérőeszköz ugyanazt a konstrukciót méri-e minden csoportban. Ez több szinten vizsgálható. Az első a *konfigurális (configural) invariancia*, amikor minden paraméter szabadon becsülhető a két mintafélen, azaz csak a faktorstruktúra azonos. Ez a szint lesz a referencia a további megkötésekhez. A második a *metrikus (metric) invariancia*, amelynél a töltéseknek kell azonosnak lenniük a két mintafélen. A harmadik a *szkaláris (scalar) invariancia*, ahol az item-intercepteknek is meg kell egyezni, valamint negyedik a *szigorú (strict) invariancia*,

amikor az itemek reziduálisainak is azonosnak kell lenni. Minél magasabb invarianciaszint igazolható, annál hasonlóbbnak tekinthető a két csoport.

Miután a mérési invarianciát elfogadhatónak találtuk, lehetőség nyílik a strukturális modell paramétereinek összehasonlítására is. Ekkor vizsgálható, hogy a látens változók közötti kapcsolatok, a regressziós útvonalak, kovarianciák eltérnek-e a csoportok között. Itt sokszor már tényleg paraméterenként (például egyszerre csak egyetlen útvonal) történik a megkötés. Így a multigroupelemzés nemcsak a mérési modell megbízhatóságát ellenőrzi, hanem betekintést nyújt abba is, hogy a vizsgált elméleti összefüggések mennyiben általánosíthatók különböző populációkra. Ezáltal a módszer a CB-SEM egyik leghasznosabb eszköze, amely biztosítja az összehasonlíthatóságot, és feltárja a csoportok közötti rejtett különbségeket.

6. Összefoglalás

Összességében a strukturális egyenletmodellezés, legyen szó akár a kovarianciaalapú vagy a varianciaalapú megközelítésről, olyan összetett statisztikai keretet kínál, amely egyszerre teszi lehetővé a mérési modellek érvényességének ellenőrzését és az elméleti összefüggések tesztelését. A módszertan legnagyobb ereje abban rejlik, hogy a látens változókat közvetlenül beemeli az elemzésbe, így a kutatók egyszerre tudják kezelni a mérési hibát, valamint feltárni a feltételezett kapcsolatok struktúráját. Mindkét eljárás különböző előfeltételek mellett alkalmazható, de közös bennük, hogy a társadalomtudományi és gazdasági kutatásokban egyaránt hozzájárulnak az elméletek mélyebb megértéséhez, empirikus alátámasztásához és a modellek általánosíthatóságának vizsgálatához.

7. Önellenőrző kérdések

1. kérdés: Miben különbözik alapvetően a CB-SEM és a PLS-SEM megközelítése?

Válasz: A CB-SEM célja a modelllilleszkedés és az elmélet tesztelése, míg a PLS-SEM inkább prediktív orientációjú, és kevésbé támaszkodik erős statisztikai feltételezésekre.

2. kérdés: Miért fontos a mérési modell invarianciájának vizsgálata a CB-SEM-ben?

Válasz: Az invariancia biztosítja, hogy a látens változókat minden csoportban azonos módon mérjük, így az összehasonlítások érvényesek és értelmezhetők.

3. kérdés: Mikor érdemes a PLS-SEM-et alkalmazni a CB-SEM helyett?

Válasz: Akkor, ha a kutatás prediktív jellegű, kisebb mintával dolgozik, vagy a modell komplex, és nem teljesíti a normalitási feltételeket.

4. kérdés: Milyen mutatókat használhatunk a PLS-SEM-ben a mérési modell érvényességének vizsgálatára?

Válasz: Az itemek vizsgálatához a külső töltésértékeket. A konstruktum vizsgálatához a reflektív modelleknél a megbízhatóságot (például ρ_A és ρ_c), a konvergens validitást (AVE) és a diszkriminánsvaliditást (HTMT) értékelhetjük. A formatív modellnél a multikollinearitást (VIF) ellenőrizzük.

5. kérdés: Milyen mutatókat használhatunk a PLS-SEM-ben a strukturális modell érvényességének vizsgálatára?

Válasz: A strukturális modell értékelésére gyakran használt mutatók az R^2, az f^2 hatásméret, a Q^2 prediktív relevancia, valamint az útvonal-együtthatók értéke és szignifikanciája bootstrapping segítségével.

Ajánlott irodalom

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cronbach, L. J. Coefficient alpha and the internal structure of tests. (1951) Psychometrika 16, 297–334. https://doi.org/10.1007/BF02310555

Dijkstra, T. K., & Henseler, J. (2015). Consistent Partial Least Squares Path Modeling1. *MIS Quarterly, 39*(2), 297–316. https://doi.org/10.25300/MISQ/2015/39.2.02

Dybro Liengaard, B. (2024). Measurement invariance testing in partial least squares structural equation modeling. *Journal of Business Research, 177*, 114581. https://doi.org/10.1016/j.jbusres.2024.114581

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50. https://doi.org/10.2307/3151312

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)*. SAGE.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. The Guilford Press.

McDonald, R. P. (1999). Test theory: A unified treatment. Lawrence Erlbaum Associates Publishers.

McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods, 28*(1), 61–88. https://doi.org/10.1037/met0000425

Ringle, C. M., Sarstedt, M., Sinkovics, N., & Sinkovics, R. R. (2023). A perspective on using partial least squares structural equation modelling in data articles. *Data in Brief, 48*, 109074. https://doi.org/10.1016/j.dib.2023.109074

Roemer, E., Schuberth, F., & Henseler, J. (2021). HTMT2–an improved criterion for assessing discriminant validity in structural equation modeling. *Industrial Management & Data Systems, 121*(12), 2637–2650. https://doi.org/10.1108/IMDS-02-2021-0082

Rosseel, Y. (2012). *lavaan : An R Package for Structural Equation Modeling*. *Journal of Statistical Software, 48*(2). https://doi.org/10.18637/jss.v048.i02

Sarstedt, M., Adler, S. J., Ringle, C. M., Cho, G., Diamantopoulos, A., Hwang, H., & Liengaard, B. D. (2024). Same model, same data, but different outcomes: Evaluating the impact of method choices in structural equation modeling. *Journal of Product Innovation Management, 41*(6), 1100–1117. https://doi.org/10.1111/jpim.12738

Sharma, P. N., Liengaard, B. D., Hair, J. F., Sarstedt, M., & Ringle, C. M. (2023). Predictive model assessment and selection in composite-based modeling using PLS-SEM: extensions and guidelines for using CVPAT. *European Journal of Marketing, 57*(6), 1662–1677. https://doi.org/10.1108/EJM-08-2020-0636

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *European Journal of Marketing, 53*(11), 2322–2347. https://doi.org/10.1108/EJM-02-2019-0189

Wolf, M. G., & McNeish, D. (2023). *Dynamic : An R Package for Deriving Dynamic Fit Index Cutoffs for Factor Analysis*. *Multivariate Behavioral Research, 58*(1), 189–194. https://doi.org/10.1080/00273171.2022.2163476

XI. BEÁGYAZOTT ANOVA (NESTED ANOVA)

SOLTÉSZ-VÁRHELYI KLÁRA

A fejezet a következő kérdéseket válaszolja meg:

a) Milyen elméleti alapokon nyugszik a beágyazott ANOVA?
b) Milyen adatstruktúrák elemzésére alkalmas a beágyazott ANOVA?
c) Mik a beágyazott ANOVA gyakorlati lépései, hogyan kell az elemzést helyesen beállítani, az eredményeket helyesen értelmezni?
d) Milyen minőségi ismérvek alapján különböztethető meg egy kiváló tanulmány egy átlagostól?

1. Bevezetés

A könyvben most következő fejezet némileg rendhagyó. Egy olyan statisztikai módszertannal fogunk megismerkedni, amely bár tudásában jóval túlmutat a legtöbb egyetemi statisztikaképzés tananyagán, a modern adatelemzés világában mégsem számít csúcstechnológiának – sőt sok szempontból kifejezetten túlhaladottnak tekinthető.

Felmerülhet a jogos kérdés, hogy ha van újabb, jobb és rugalmasabb, akkor miért szentelünk egy egész fejezetet egy látszólag elavult eljárásnak, a *beágyazott ANOVA*-nak (nested ANOVA). A válasz többrétegű, és messze túlmutat a statisztikatörténeti érdekességeken.

Először is, ez a módszer makacsul tartja magát a tudományos gyakorlatban. Régebbi cikkekben lépten-nyomon találkozhatunk vele, de még ma is előkerül, sokszor pontatlanul, félig-meddig tévesen alkalmazva. Ahhoz, hogy ezeket a hibákat felismerjük, és kritikusan tudjuk olvasni a szakirodalmat, ismernünk kell a módszer logikáját, korlátait és buktatóit. Célunk tehát, hogy tisztán meghatározzuk, milyen adatstruktúrára és milyen elemzési céllal használható ez az eszköz, és melyek azok a félreértések, amelyeket ideje egyszer és mindenkorra eloszlatni.

Másodszor, és talán ez a legfontosabb, a beágyazott ANOVA kiváló felvezető a modernebb eljárásokhoz. Olyan alapvető, kulcsfontosságú statisztikai koncepciók megértéséhez biztosít tökéletes terepet, mint a hierarchikus, egymásba ágyazott adatstruktúrák, a fix és véletlen hatások közötti különbség vagy a kiegyensúlyozott és kiegyensúlyozatlan kísérleti elrendezések problémája.

Ezeknek a fogalmaknak a magabiztos ismerete nélkül a modern eljárások útvesztőnek tűnhetnek. Ez a fejezet tehát egyfajta elméleti alapozás, itt fektetjük le azokat a fogalmi építőköveket, amelyekre a következő, jóval komplexebb és hatékonyabb módszertan, a lineáris kevert modellek (Linear Mixed Models) módszere épül. Úgy is fogalmazhatunk, hogy megtanulunk biciklizni, mielőtt motorra ülnénk, azaz egy stabil, megbízható és nélkülözhetetlen tudás megszerzése a cél, ami a jövőbeli magabiztos elemzések alapját képezi.

Ez és a következő fejezet közös célja, hogy rendet teremtsen az egymással átfedő és ezért gyakran tévesen összekevert módszertanok között azáltal, hogy összegzi, mi a hasonlóság és különbség a beágyazott ANOVA (nested ANOVA), az ismételt méréses ANOVA (Repeated Measures ANOVA), a vegyes ANOVA

(mixed ANOVA), a hierarchikusan lineáris modellek (HLM), a többszintű modellek (MLM) és a lineáris kevert modellek (LMM) között.

2. Elméleti háttér

Ahhoz, hogy eljussunk a beágyazott ANOVA modelljéhez, először több a témához kapcsolódó, az ANOVA oktatása során hallgatólagosan használt, de nyíltan ritkán kimondott statisztikai koncepciót is tisztáznunk kell.

2.1. A függetlenség feltétele

A legtöbb statisztikai kurzus alapvető részét képezi a varianciaanalízis (ANOVA) két alaptípusának megkülönböztetése. Megtanuljuk, hogy a független mintás ANOVA-t akkor használjuk, ha a kísérleti személyek és az általuk alkotott csoportok is teljesen függetlenek egymástól (például egy kísérleti és egy kontrollcsoport különböző személyekből áll). Ezzel szemben az ismételt méréses ANOVA akkor a megfelelő eszköz, ha ugyanazokat a személyeket mérjük le több különböző időpontban vagy körülmény között. De miért ez a szigorú megkülönböztetés? A válasz a statisztikai modellek legfontosabb, mégis leggyakrabban megsértett alaptételében, *a mérések függetlenségének feltételében* rejlik.

Képzeljük el, hogy egy ismételt méréses elrendezést – ahol minden személytől van egy első és egy második mérésünk – hibásan, független mintásként szeretnénk elemezni. Ehhez az adatokat „egymás alá pakolnánk", mintha az első mérések az egyik, a második mérések pedig egy másik, teljesen független csoportból származnának. Ezzel súlyos hibát követnénk el. Az ugyanattól a személytől származó két mérés nem független, mivel egy magas képességű személy valószínűleg mindkét időpontban átlag felett, míg egy alacsonyabb képességű személy mindkétszer átlag alatt fog teljesíteni. A független mintás modell ezekről a személyen belüli korrelációkról nem tud, azokat nem tudja kezelni, és tévesen azt hiszi, hogy kétszer annyi független információval rendelkezik, mint valójában. Ennek következtében drasztikusan alábecsüli a valódi véletlen hibát (zajt), ami az F-statisztika (amely a hatásnak és zajnak a hányadosa) nevezőjében szerepel. A mesterségesen lecsökkentett zaj miatt

az F-érték irreálisan nagynak tűnik, a mesterségesen megnövelt F-érték viszont kisebb szignifikanciaértékeket eredményez majd (azaz a kelleténél szignifikánsabb eredményeket), ez pedig az *elsőfajú hiba* valószínűségének drámai megnövekedéséhez vezet.

A függetlenség feltételének biztosítása miatt van szükség a helyes eljárás megválasztására. Azokat a modelleket, melyek csak *független mintás szempontokat* (between subject factors) tartalmaznak, *független mintás ANOVA*-nak nevezzük. Erre példa, ha azt vizsgáljuk, hogy az autizmus (érintett vagy egészséges) és a nem (férfi vagy nő) hatással van-e az érzelemfelismerési képességre. Azokat a modelleket, amelyek csak *összefüggő mintás szempontokat* (within subject factors) tartalmaznak, *ismételt méréses (Repeated Measures) ANOVA*-nak nevezzük, és jó példa rá, ha azt vizsgáljuk, hogy a tej színezése (ami lehet kék és piros) és hőmérséklete (ami lehet hideg vagy meleg) befolyásolja-e azt, hogy mennyire érezzük finomnak, és minden kísérleti személy kipróbál minden kombinációt (hideg-kék, meleg-kék, hideg-piros, meleg-piros). Azokat a modelleket, melyek tartalmaznak független és összefüggő mintás szempontot is, *vegyes (mixed) ANOVA*-nak nevezzük, és jó példa rá, ha azt nézzük, hogy az „A" típusú vagy „B" típusú terápia növeli-e jobban az énhatékonyságot, ahol mindenkit kétszer mérünk le, a terápia előtt és után (tehát az idő lesz az ismételt méréses szempont), de mindenki csak vagy az „A", vagy a „B" terápiában vesz részt (ez lesz a független mintás szempont).

A korábbi példa az ismételt mérések egymás alá helyezésével azért volt hibás, mert a sorok nem voltak egymástól függetlenek. A megfigyelések függetlensége pedig a hagyományos lineáris modellek alapja. Vagyis az egész statisztikai logika arra épül, hogy minden egyes adatsor (megfigyelés) egy új, független információmorzsát ad hozzá a nagy egészhez. Ha ez a feltétel sérül, az egész modell érvényét veszti, a kapott p-értékek pedig megbízhatatlanná válnak.

A függetlenség feltétele azonban nemcsak ilyen egyértelmű adatkezelési hibákkal sérülhet, hanem sokkal finomabb, rejtettebb módon is. Tegyük fel, hogy két új tanítási technikát szeretnénk kipróbálni az egyetemen. A hat gyakorlati csoportból három az „A" technikát, a másik három pedig a „B" technikát alkalmazza, és félév végén tervezem lemérni az oktatással való elégedettségüket. Erre nem elég egy egyszerű független mintás t-próbát végezni, mert a két technika ugyan valóban független csoportosító tényező, hiszen egy hallgató csak az egyikben vehetett részt, azonban maguk a hallgatók, a kísérleti személyek, már nem teljesen függetlenek egymástól. Azok a diákok, akik egyazon gyakorlati csoportba jártak, sok szempontból „klaszteresődnek", mivel ugyanaz a gyakorlatvezetőjük, ugyanabban az időpontban és teremben volt az órájuk, és a csoporton belüli interakciók, a közös csoportdinamika is hatással volt a

tapasztalataikra. Ezt a hasonlóságot az azonos csoportba járó hallgatók között nevezzük *osztályon belüli korrelációnak* (intraclass correlation, ICC).

Ez a rejtett csoportosulás pontosan ugyanahhoz a problémához vezet, mint az ismételt mérések helytelen kezelése. A modell, ha nem vesszük figyelembe az ICC-t, tévesen azt feltételezi, hogy minden egyes diák egy független adatpont. Ezzel ismét alábecsüli a valódi hibát, és megnöveli az elsőfajú hiba kockázatát, hiszen a valódi független egységek száma a technikák összehasonlításában nem a diákok száma, hanem a gyakorlati csoportoké. És ezzel meg is érkeztünk a beágyazott ANOVA archetípusához, az olyan helyzetekhez, ahol a minket érdeklő (fix) csoportosító tényezőn túl létezik egy másik, „zavaró" csoportosító tényező is, amely egymásba ágyazott adatstruktúrát hoz létre. Ezt a rejtett struktúrát a modellben kötelező figyelembe venni, még akkor is, ha maga a klaszterező tényező (például a konkrét gyakorlati csoport) nem áll a kutatásunk fókuszában.

2.2. Keresztezett és beágyazott kísérleti elrendezések

Persze egy vállrándítással mondhatnánk, hogy ha a gyakorlati csoportokba való tartozás sérti a függetlenség elvét, akkor egyszerűen tegyük be a gyakorlati csoportok azonosítóját is a modellbe egy második szempontként, és ezzel megoldanánk a problémát. Miért nem ezt tesszük? Röviden azért, mert a hagyományos ANOVA-modell matematikailag nem tud lefutni egy ilyen adatstruktúrával. Gondoljuk végig, mi kellene ahhoz, hogy egy két szempontos ANOVA működjön? Az, hogy a két független változó minden lehetséges kombinációjában legyenek méréseink.

Emlékezzünk vissza a korábbi példára, ahol azt vizsgáltuk, hogy az autizmusspektrum-zavar és a nem hatással van-e az érzelemfelismerési pontszámra. Itt a modell tökéletesen működik, mert a két szempont (autizmus és nem) minden szintje kombinálódik. Képzeljük el ezt egy 2x2-es táblázatban (24. ábra A panelje), a mintánk tartalmaz autista fiúkat, autista lányokat, kontrollfiúkat és kontroll-lányokat is. Minden cella fel van töltve adatokkal. Ezt a felépítést, ahol minden szempont minden szintje kombinálódik a másik szempont minden szintjével, *keresztezett elrendezésnek* (crossed design) nevezzük.

De mi a helyzet az oktatási technikás példánkkal? Itt a két szempont az oktatási technika (A vagy B) és a gyakorlati csoport (1-től 6-ig). Ha megpróbálnánk felrajzolni a cellákat, azonnal falba ütköznénk. Az 1-es, 2-es és 3-as gyakorlati csoportok tagjai kizárólag az A technikával találkoztak. Az ő celláik a B technika alatt teljesen üresek. Fordítva, a 4-es, 5-ös és 6-os csoportoknál

az A technika alatti cellák üresek. A modell nem tudja megbecsülni a hiányzó kombinációk hatását. Ezt a struktúrát, ahol az egyik szempont (a gyakorlati csoport) szintjei csakis a másik szempont (az oktatási technika) egyetlen szintjén belül léteznek, *beágyazott elrendezésnek* (nested design) nevezzük.

Felmerülhet a gondolat, hogy egy okos kísérleti elrendezéssel feltölthetnénk a hiányzó cellákat. Ehhez az kellene, hogy minden gyakorlati csoport találkozzon mindkét oktatási technikával. Így létrehoznánk egy vegyes elrendezésű ANOVA-t, ahol az oktatási technika egy ismételt méréses, csoporton belüli szempont lenne. Bár ez bizonyos esetekben, például a kognitív pszichológiai kísérletekben (ahol ugyanaz a személy több különböző ingerrel is találkozik) viszonylag egyszerűen megoldható, a mi oktatási példánkban szinte lehetetlen. Nem taníthatom ugyanazt az anyagot kétszer, két különböző technikával ugyanannak a csoportnak, mert a második alkalommal már nem a technika hatását mérném, hanem az ismétlését. Ha pedig két különböző tananyagot tanítanék a két technikával, az eredmények összehasonlíthatatlanná válnának a tananyagok eltérősége miatt.

Láthatjuk tehát, hogy a kísérleti elrendezés alapvetően meghatározza, milyen statisztikai modellt alkalmazhatunk. Keresztezett elrendezést akkor használunk, ha a szempontjaink szintjei szabadon kombinálódnak, és a fő kérdésünk a szempontok önálló hatására irányul (valamint esetleg a két szempont interakciójára). Az egyetemi képzésekben leggyakrabban tanult ANOVA-eljárások mind keresztezettek, még ha ez legtöbbször nincs is kimondva. Ezzel szemben beágyazott elrendezést akkor kell alkalmaznunk, ha az egyik szempont szintjei hierarchikusan a másik szempont szintjei alá vannak rendelve. Ilyenkor a fő kérdés a felsőbb szintű szempont hatása, miközben kontrolláljuk az alacsonyabb szintű csoportosulásból fakadó torzítást.

A kísérleti elrendezés megalkotásakor alaposan át kell gondolni, melyik szempont van beágyazva melyikbe. A mi példánkban a gyakorlati csoportok vannak az oktatási technikák alá beágyazva, mert egy csoport csak egy oktatási technikával találkozott. Fordítva ez nem igaz, az oktatási technika nincs beágyazva a gyakorlati csoportok alá, mert egy technikával több csoport is találkozott.

Ez a koncepcionális különbség a modellek felépítésében matematikailag is megmutatkozik, leglátványosabban az F-statisztika képletének megváltozásában. A keresztezett ANOVA-k világában a szabály egyszerű és következetes, csak a lényeget kiragadva: egy szempont hatását úgy teszteljük, hogy a hatásból származó varianciát, azaz az összehasonlítandó minták közötti eltérést (matematikailag az úgynevezett hatás szórásnégyzetét, $MS_{hatás}$) elosztjuk a kísérleti személyek közötti, megmagyarázatlan véletlen varianciával, más néven a hibával vagy zajjal (matematikailag a hiba szórásnégyzetével, MS_{hiba}), azaz

$$F = \frac{MS_{hatás}}{MS_{hiba}}$$

Egy nagyon egyszerű példával élve, a férfiak és nők magassága közötti különbség keresésénél a férfi és női átlag közötti eltérést osztjuk a férfiak és nők csoportján belül a személyek közötti eltéréssel.

$$F_{nem} = \frac{MS_{nem}}{MS_{hiba}}$$

A beágyazott modellekben azonban ez a logika egy láncolattá alakul. A legalsó szintű, beágyazott szempont (a mi esetünkben a gyakorlati csoport) hatását még a hagyományos módon teszteljük, az egyéni hibához viszonyítva.

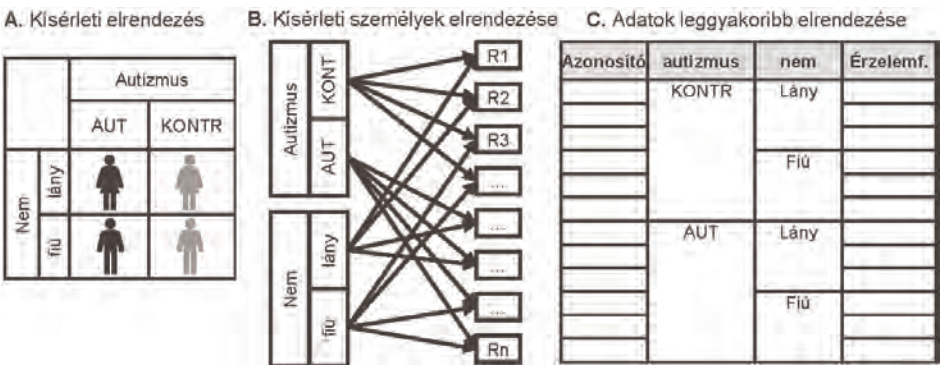
$$F_{gy.csoport} = \frac{MS_{gy.csoport}}{MS_{hiba}}$$

Azonban a minket valóban érdeklő, felsőbb szintű szempont (jelen esetben az oktatási technika) tesztelésénél a nevezőbe már nem az egyéni hiba, hanem a beágyazott szempont szórásnégyzete, a gyakorlati csoportok között tapasztalható változatosság kerül. A képletek tehát így alakulnak:

$$F_{technika} = \frac{MS_{technika}}{MS_{gy.csoport}}$$

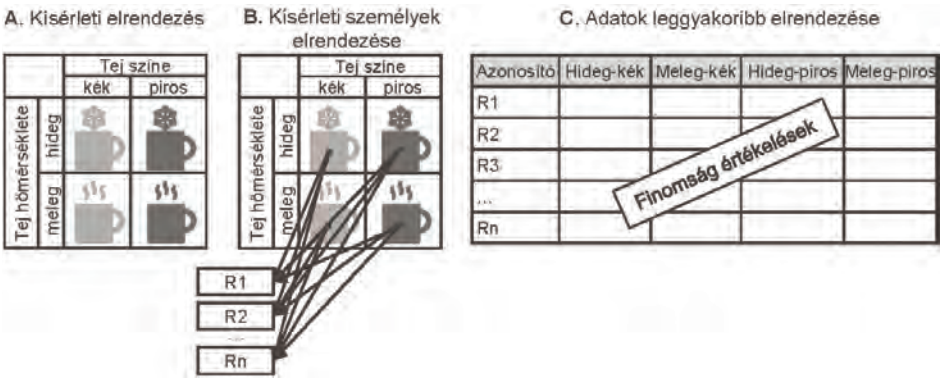
Ennek logikája mélyen összefonódik a függetlenség megsértésének problémájával. Mivel a diákok csoportokba tömörülnek, a technika hatásának valódi zaját már nem az egyes diákok közötti véletlen különbségek jelentik, hanem az egész gyakorlati csoportok közötti véletlenszerű eltérések. A modell bölcsen felismeri, hogy a technika hatásának le kell győznie a csoportokból fakadó természetes varianciát ahhoz, hogy valóban szignifikánsnak tekinthessük. Ez a nevezőcsere az a matematikai lépés, amellyel a modell kontrollálja a beágyazottságból fakadó klaszteredést.

Nézzük át az eddig látott kísérleti elrendezéseket. A 24. ábra a tipikus független mintás ANOVA elrendezését mutatja, és az autizmus és a nem hatását vizsgálja az érzelemfelismerésre. Ez egy két szempontos, keresztezett elrendezés. Két független szempont van, az Autizmus (autista/kontroll) és a Nem (fiú/lány), és e két szempont minden kombinációja megjelenik a kutatásban. A B panelben látható, hogy minden kombinációban van is résztvevő (R-ekkel jelölve). A C panel az adatok helyes elrendezését mutatja, ahol a különböző minták egymás alatt helyezkednek el, hiszen ezek egymástól független csoportok.



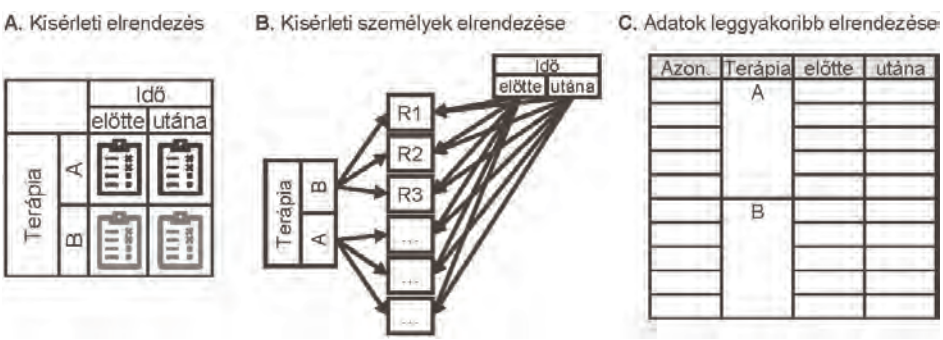
24. ábra: Két szempontos, független mintás ANOVA modellje

A 25. ábra egy tipikus ismételt méréses elrendezés. Itt is két szempont van, a tej színe (kék/piros) és hőmérséklete (hideg/meleg), és ezek hatását vizsgáljuk az észlelt finomságra. Ez is keresztezett elrendezés, hiszen minden lehetséges kombináció megjelenik, és mindkét szempont összefüggő mintás, mert minden kísérleti személy kóstol hideget és meleget is, kéket és pirosat is, összesen négyféle tejet. A C panelen látható, hogy a mintákat egymás melletti változókban rögzítettük, helyesen, hiszen ezek összefüggő mérések.



25. ábra: Két szempontos, ismételt méréses ANOVA modellje

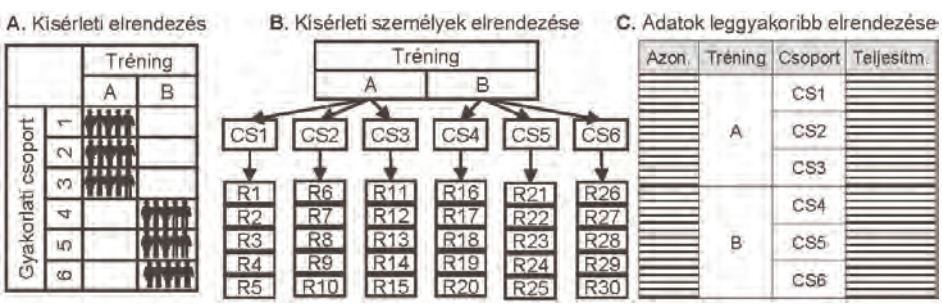
A 26. ábrán egy vegyes ANOVA modellje látható. Azt vizsgáljuk, hogy A vagy B terápia hatására csökken-e jobban a szorongás. A modellben van egy független mintás szempont (A vagy B terápia), és van egy ismételt méréses szempont, a mérés ideje (terápia előtti/utáni mérés). Ez az elrendezés is teljes egészében keresztezett, mindenkit megmértünk a terápia előtt és után is, függetlenül attól, hogy az A vagy a B terápiában vett részt.



26. ábra: Két szempontos, vegyes ANOVA modellje

A 27. ábra jeleníti meg a beágyazott ANOVA modelljét. A vizsgálatban a két oktatási technikával való elégedettség különbségét vizsgálom a hat gyakorlati csoportban. Több dolgot is érdemes észrevenni. Egyrészről itt is két szempont van, a technika (ami lehet A vagy B) és a gyakorlati csoportok (1-től 6-ig). Ez utóbbi hatása ugyan nem érdekes számunkra, de az ICC miatt mégis figyelembe kell vennünk a modellben. Ez a modell már nem keresztezett, mivel a két változó kereszttáblájában vannak üres cellák (27. ábra A panel), minden gyakorlati

csoport csak az egyik technikával tanul, a másikban nincsen hozzá adat. A szétválasztás szigorú, nincs egyetlen csoport se, de még csak egyetlen hallgató sem, akinek mindkét technikánál lenne adata. Ez a klasszikus beágyazott szerkezet. A gyakorlati csoport van beágyazva az oktatási technikába, mert egy gyakorlati csoport csak egy technika alatt fordulhat elő, míg fordítva ez nem igaz, egy oktatási technikát több csoport is kipróbált. Ebben az esetben az oktatási technika a fő szempont, amelynek a hatását vizsgáljuk, és ezt nevezzük *szülőnek*. A gyakorlati csoportok az egyéb hatás, melyre kontrollálnunk kell, és ezt nevezzük *gyereknek*. A statisztikai szoftverek szintaxisában ez úgy jelenik meg, hogy *gyerek(szülő)*, azaz a gyerek tényező van beágyazva a szülő tényező alá. Vigyázzunk, nehogy felcseréljük a szerepeket! Antropomorfizálva a képletet, logikusabbnak tűnhet, hogy a szülő öleli át a gyermeket, és ezért a kezdő kutatók gyakran elrontják a felírást. Gondoljunk inkább rá úgy, hogy a gyerek szempontját adjuk éppen meg, amihez zárójelben megadjuk azt a tulajdonságát, hogy be van ágyazva a szülő alá!



27. ábra: Beágyazott ANOVA modellje

2.3. A modell kiegyensúlyozottsága

Van még egy szembeötlő tulajdonsága a 27. ábra modelljének. Az, hogy teljesen kiegyensúlyozott, azaz mindkét technika alá ugyanannyi csoport (3-3) van rendelve, és minden csoport ugyanannyi hallgatót tartalmaz. Ez nem véletlen, mivel a beágyazott ANOVA alapvetően olyan kísérleti elrendezésekre lett kialakítva, melyeknél az adatfelvétel kiegyensúlyozott, és ezzel el is érkeztünk az ANOVA-vizsgálatok következő, nyíltan viszonylag ritkán tárgyalt kérdéséhez, a kiegyensúlyozottsághoz.

De mit is jelent a kiegyensúlyozottság?

Egy kísérleti elrendezést akkor nevezünk *kiegyensúlyozottnak* (balanced), ha a független változók hierarchiájának minden szintjén egyenlő számú alacsonyabb szintű egység található.

Egy keresztezett ANOVA-nál ez azt jelenti, hogy a szempontok minden lehetséges kombinációjában (cellájában) azonos az elemszám, az előző példára visszautalva a vizsgálatban pontosan ugyanannyi autista fiú, autista lány, kontrollfiú és kontroll-lány teljesítményét mérték meg. Egy beágyazott modell esetében ez egy láncolatot feltételez, hiszen itt a kombinációk többsége eleve nem is létezik. Az oktatási technika példáját használva, a tökéletes kiegyensúlyozottsághoz két feltételnek kell egyszerre teljesülnie: 1) Minden oktatási technika (szülő szempont) alatt pontosan ugyanannyi gyakorlati csoport (gyerek szempont) kell hogy legyen (például 3-3 csoport az A és a B technikánál). 2) Minden egyes gyakorlati csoportban pontosan ugyanannyi hallgató (mérés) kell hogy legyen (például minden csoportban 15 diák).

Ha ezen feltételek bármelyike sérül, akár azáltan, hogy az egyik technikát csak három, míg a másikat négy csoporton teszteljük, vagy úgy, hogy a csoportokba eltérő számú diák jár, az elrendezés máris *kiegyensúlyozatlanná* (unbalanced) válik. Ez a szigorú, többszintű követelmény az egyik oka annak, hogy a valós kutatási adatok ritkán tökéletesen kiegyensúlyozottak.

A kiegyensúlyozatlanság mind a keresztezett, mind a beágyazott modellek számára probléma. Megtöri a szempontok közötti *ortogonalitást*, vagyis a hatásaik „összefolynak", és nehézzé válik eldönteni, hogy a függő változó varianciájának egy adott részét melyik szempont magyarázza. Az autizmussal kapcsolatos példán bemutatva: ha az autista csoportban sokkal nagyobb a férfiak aránya, mint a kontrollcsoportban, akkor az autizmus és nem hatása az érzelemfelismerésre részben átfedésbe kerül, nem tudjuk egyértelműen szétválasztani, hogy a csoportok közötti különbség az autizmusnak vagy a nemek aránytalanságának tudható-e be. (Azért teljesít rosszabbul az autista csoport, mert gyengébb az érzelemfelismerő képességük, vagy azért, mert több a csoportban a férfi, és a férfiaknak gyengébb ez a képessége? Esetleg mindkettő, de akkor is kérdés, milyen arányban?)

A legtöbb egyetemi statisztikakönyv és kurzus ezen a ponton elegánsan továbblép, hiszen a modern statisztikai szoftverek a háttérben automatikusan kezelik a problémát. A legelterjedtebb megoldás a Type III típusú négyzetösszegek (Sum of Squares) használata. Ez a módszer minden egyes szempont hatását úgy teszteli, hogy közben statisztikailag kontrollálja az összes többi hatást a modellben, így egy torzítatlan, megbízható becslést ad minden főhatásra és interakcióra, függetlenül az elemszámok egyenlőtlenségéből fakadó átfedésektől.

Azonban míg a keresztezett ANOVA-nál a kiegyensúlyozatlanság egy könnyen megoldható korrekciós probléma, a beágyazott elrendezéseknél sokkal mélyebbre hatol, és magát a hipotézistesztelés logikáját kérdőjelezi meg. Ahogy korábban láttuk, egy beágyazott modellben a szülő szempont hatásának teszteléséhez a nevezőben az alá ágyazott gyerek szempont szórásnégyzete (MS_{gyerek}) szerepel. Kiegyensúlyozatlan esetben ez az MS-érték már nem tiszta becslése a gyerek szempont varianciájának, így az F-statisztika képlete érvényét veszti. A probléma tehát nem csupán a becslés pontosságában rejlik, hanem abban, hogy maga az F-statisztika helyes felépítése válik kétértelművé.

Ezzel pedig elértünk a beágyazott ANOVA egyik legfontosabb korlátjához. A torzítás kezelésére léteznek ugyan korrekciós eljárások, mint például a rendkívül bonyolult Satterthwaite-féle közelítés, amely különböző szórásnégyzetek kombinálásával próbál meg egy használható, szintetikus nevezőt létrehozni az F-teszthez, de ezeknél a matematikai háttér nehézkessé válik, a becslés pontossága megkérdőjelezhető, és a modell statisztikai ereje jelentősen csökken.

Egy sokkal elegánsabb és hatékonyabb megoldást ismerünk majd meg a következő fejezetben. A megoldás az lesz, hogy teljesen elvetjük a nehézkessé vált F-statisztikán alapuló logikát, és áttérünk egy másik becslési eljárásra, a Maximum Likelihood (ML) módszerére. Ez a valószínűségi alapú megközelítés közvetlenül a varianciakomponenseket becsüli meg minden szinten, kikerülve ezzel az F-arányok felépítésének buktatóit. Ez azt jelenti, hogy a tesztválasztás során is hátrahagyjuk az ANOVA-eljárásokat (és ezzel a beágyazott ANOVA-t is), és áttérünk a modernebb lineáris kevert modellezésre (LMM), mely egyértelműen a pontosabb, nagyobb statisztikai erejű és koncepcionálisan is tisztább megoldást kínálja.

2.4. Fix és random hatások

Ha meg is hoztuk az olvasó kedvét, hogy átugorjon a modernebb eljárásokra, időzzünk még el egy kicsit ennél az egyszerűbb elemzési módnál!

Ahhoz, hogy a beágyazott ANOVA és lineáris kevert modellezés logikáját teljes mértékben megértsük, elengedhetetlen, hogy megismerkedjünk a statisztikai modellezés egyik legfontosabb megkülönböztetésével, a fix és random hatások közötti különbséggel.

Egy szempontot *fix hatásnak* (fixed effect) tekintjük, ha a vizsgálatban szereplő szintjei (csoportjai) önmagukban érdekesek számunkra, és a kutatásunk kizárólag ezekre a konkrét szintekre vonatkozik. Más szóval, a szempont összes lehetséges, számunkra érdekes szintjét bevontuk a vizsgálatba, és nem célunk

az eredményeket más, be nem vont szintekre általánosítani. Tipikusan ilyenek a kísérleti manipulációk (például kísérleti és kontrollcsoport) vagy az olyan demográfiai változók, amelyeknek minden releváns kategóriáját vizsgáljuk (például férfi és nő).

Ezzel szemben egy szempontot random hatásnak (random effect) tekintünk, ha a vizsgálatba bevont szintjei csupán egy véletlenszerű mintát képviselnek egy sokkal nagyobb, lehetséges szintekből álló populációból. Ilyenkor nem a konkrét, bevont szintek (például a konkrét gyakorlati csoportok vagy a mérést végző konkrét orvosok) közötti különbség érdekel minket. A célunk csupán az, hogy az általuk képviselt varianciát, a különbözőségeikből származó zajt megbecsüljük és statisztikailag kontrolláljuk, hogy a minket valóban érdeklő fix hatásokra vonatkozó következtetéseink általánosíthatóbbak és pontosabbak legyenek.

A beágyazott elrendezésekben a hierarchia felsőbb szintjén álló, minket érdeklő szempont szinte mindig fix, míg a beágyazott, klaszterződést okozó szempont jellemzően random. Nézzünk erre néhány példát!

1. A klasszikus eset: mikroműanyagok a vizekben. Tegyük fel, hogy arra vagyunk kíváncsiak, eltér-e a mikroműanyag-szennyezettség mértéke Magyarország álló- és folyóvizeiben. Ehhez egyenként húsz-húsz mintát veszünk három tóból (Balaton, Velencei-tó és Fertő-tó) és három folyóból (Duna, Tisza és Dráva). Ekkor a fix hatás a víztípus (tó vagy folyó). Ez a kutatásunk központi kérdése. A két szint (állóvíz, folyóvíz) kimeríti az érdeklődési körünket, az eredményekből erre a két kategóriára vonatkozóan akarunk következtetéseket levonni. A random hatás a konkrét víztest (Duna, Tisza stb.). A hat kiválasztott víztest csupán egy minta Magyarország összes tava és folyója közül. Nem az a kérdés, hogy a Duna szennyezettebb-e a Tiszánál, hanem az, hogy általában a tavak és folyók különböznek-e, miközben figyelembe vesszük azt a természetes varianciát, amely az egyes folyók vagy tavak között egyébként is fennáll.

2. A komplexebb eset: tolerancia az óvodákban. Azt vizsgáljuk, hogy az integrált, fogyatékos gyerekeket is fogadó óvodai csoportokban magasabb-e az egészséges gyerekek toleranciája, mint a nem integráló csoportokban. Itt a fix hatás a csoport típusa (integráló vagy nem integráló). Ez a kutatási hipotézisünk magja. A random hatás az óvodai csoport. A vizsgálatba bevont konkrét csoportok (például a Katica és Süni csoport) csupán egy mintát képviselnek az összes lehetséges óvodai csoport populációjából. A célunk a csoportokból fakadó zaj (például az óvónő személyisége, a csoportdinamika) kontrollálása. A helyzetet tovább bonyolítja, hogy a valóságban a csoportok óvodákba vannak ágyazva, így egy háromszintű hierarchiánk van, hiszen az gyerekek óvodai csoportokhoz tartoznak, a csoportok pedig óvodákhoz, és mindennek a tetején kellene a minket érdeklő fix szempontnak (integráló/nem integráló) lenni. A beágyazott ANOVA

használatához mindezeknek teljes mértékben beágyazottnak kellene lenniük kereszteződések nélkül, de a valóság természetesen nem így rendeződik, mert egy óvoda nem lesz teljes egészében, minden csoportjával integráló vagy nem integráló. Egy ilyen modell elemzése már erősen feszegeti a beágyazott ANOVA kereteit. Egy ilyen helyzetben a kutatónak át kellene térnie a lineáris kevert modellezésre, vagy ha meg akar maradni a beágyazott ANOVA-k világában, akkor egy szakmailag gyakran megengedhető egyszerűsítéssel el kellene tekintenie az óvoda szintjétől, feltételezve, hogy a csoportok közötti különbségek sokkal nagyobb hatásúak, mint az egyes intézmények közötti eltérések.

3. A kutatói döntés esete: oktatási technikák. Végül térjünk vissza az oktatási technikás példánkhoz! Itt a fix hatás az oktatási technika („A" vagy „B"). Ez egyértelmű, hiszen a két konkrét technika hatékonyságát hasonlítjuk össze, de mi a helyzet a gyakorlati csoport szempontjával? Mondhatnánk, hogy mivel összesen hat gyakorlati csoportom van, és mindet bevontam a vizsgálatba, ez egy fix hatás. De mi a kutatási kérdésünk? Ha az érdekelne, hogy a kedd reggeli 3-as csoport és a csütörtök déli 5-ös csoport között van-e különbség, akkor a szempont valóban fix lenne. A kutatás célja azonban szinte soha nem ez. A cél annak megállapítása, hogy az „A" technika általánosságban jobb-e, mint a „B", kontrollálva a gyakorlati csoportok különbözőségeit. Ekkor a gyakorlati csoport hatása nem érdekel minket, csak az a zaj, amelyet a rendszerbe visz. Ezért sokkal helyesebb random szempontként kezelni, mert így az oktatási technika hatását egy általánosabb, valósabb kontextusban tudjuk értékelni.

Láthatjuk tehát, hogy a beágyazott szempontok legtöbbször random hatásként kezelendők. Amikor egy beágyazott ANOVA-elemzés során egy szempontot randomként jelölünk meg, akkor a szoftver a klasszikus szórásnégyzeteken alapuló logikán belül maradva, a korábban bemutatott láncolat szerint, a megfelelő nevezőket használja az F-statisztikák kiszámításához. Ez egy okos, de viszonylag merev rendszer, melynek korlátaival a későbbiekben még szembesülni fogunk.

24. doboz: A beágyazott és hierarchikus kísérleti elrendezések statisztikai alapfogalmai

Függetlenség feltételezése (Assumption of independence): az az alapfeltételezés, hogy a mintában szereplő minden egyes megfigyelés (adatsor) teljesen független az összes többitől, és ezek nem befolyásolják egymást.
Pszeudoreplikáció (Pseudoreplication): az a súlyos statisztikai hiba, amikor nem független adatpontokat (például egyazon csoportba tartozó diákokat) független adatként kezelünk, ami az elsőfajú hiba esélyének drámai növekedéséhez vezet.
Keresztezett elrendezés (Crossed Design): olyan kísérleti elrendezés, amelyben az egyik szempont minden szintje kombinálódik a másik szempont minden szintjével. Beágyazott elrendezés (Nested Design): olyan hierarchikus kísérleti elrendezés, ahol az egyik szempont („gyerek") szintjei kizárólag a másik szempont („szülő") egyetlen szintjén belül léteznek.
Osztályon belüli korreláció (Intraclass Correlation, ICC): egy 0 és 1 közötti mutató, amely számszerűsíti, hogy a modell fix hatásai által meg nem magyarázott variancia mekkora hányada fakad az adatok csoportokba való tömörüléséből (klasztereződéséből).
Szülő szempont (Parent Factor): a beágyazott hierarchia felsőbb szintjén álló, jellemzően fix hatás (például A és B beavatkozás), amely tartalmazza a gyerek szempont szintjeit.
Gyerek szempont (Child Factor): a beágyazott hierarchia alacsonyabb szintjén álló, jellemzően random hatás, amely a szülő szempontba van beágyazva (osztály, gyakorlati csoport).
Kiegyensúlyozottság (Balance): az az ideális állapot egy beágyazott modellben, amikor minden szülő szempont alatt azonos számú gyerek szempont van, és minden gyerek szemponton belül azonos a megfigyelések száma. (Keresztezett elrendezésekben a kiegyensúlyozottság a szempontok kombinációiban előforduló elemek számának azonosságát jelenti.)
Type III négyzetösszegek (Type III Sum of Squares): kiegyensúlyozatlan modelleknél használt számítási módszer, amely egy szempont hatását annak egyedi, más szempontoktól független variancia-hozzájárulásán keresztül teszteli. Keresztezett modelleknél elegendő korrekció, de gyakran beágyazott modelleknél is csak ez elérhető.
Satterthwaite-féle közelítés (Satterthwaite Approximation): statisztikai eljárás, amely kifejezetten beágyazott, random hatásokat tartalmazó, kiegyensúlyozatlan ANOVA-modellekben használatos, amikor a Type III korrekció már nem elég. A helyes F-teszt elvégzéséhez egy szintetikus hibaszórásnégyzetet hoz létre több, a modellben szereplő varianciaforrás kombinálásával.

Fix hatás (Fixed Effect): olyan szempont, amelynek konkrét szintjei érdekelnek minket, és a vizsgálatba bevont szintek kimerítik az összes olyan szintet, amelyre az eredményeket általánosítani szeretnénk.
Random hatás (Random Effect): olyan szempont, amelynek szintjei egy nagyobb populációból vett véletlenszerű mintát képviselnek, és a célunk nem a szintek összehasonlítása, hanem az általuk képviselt variancia kontrollálása.
F-teszt nevezőcseréje (Change of F-test Denominator): a beágyazott ANOVA azon matematikai sajátossága, hogy a felsőbb szintű (szülő) szempont szignifikanciájának teszteléséhez a nevezőben nem az egyéni hiba, hanem az alá ágyazott (gyerek) szempont szórásnégyzete szerepel.
Lineáris Kevert Modellek (Linear Mixed Models, LMM): a beágyazott ANOVA-t felváltó modern, rugalmas eljárás, amely valószínűségi becslést használ a fix és random hatások modellezésére, és kiválóan kezeli a kiegyensúlyozatlan vagy komplex hierarchikus adatokat.

3. Kutatási folyamat

A beágyazott ANOVA többszintű gondolkodást igényel, mivel a modell hierarchikus természetét már a kutatás tervezésének legelső fázisában figyelembe kell venni.

3.1. Mintaválasztás

Elsőként a kísérleti elrendezés hierarchiáját kell pontosan feltérképezni. Ez határozza meg, melyik szempont a minket érdeklő fix hatás (például az oktatási technika), és melyik a klaszterződést okozó, kontrollálandó random hatás (például a gyakorlati csoport). A modell érvényessége ezen a helyes besoroláson múlik.

A struktúra azonosítását követően kiemelten fontos a kiegyensúlyozott kísérleti elrendezésre való törekvés. Bár a tökéletes kiegyensúlyozottság, azaz az azonos számú csoport és a csoporton belüli azonos elemszám, ritkán érhető el, a tervezés során törekedni kell rá, mivel a beágyazott ANOVA-modell erre a legérzékenyebb.

A statisztikai erő elsősorban a felsőbb szintű egységek, azaz a csoportok számától függ. Minél magasabb az osztályon belüli korreláció (ICC), vagyis minél

jobban hasonlítanak egymásra az egy csoportba tartozó tagok, annál több független csoportra van szükség a megbízható eredményhez. Ennek fényében a hangsúlyt inkább a csoportok számára, mintsem a csoportokon belüli elemszám növelésére érdemes helyezni, azaz jobb stratégia tíz csoportot bevonni tíz-tíz fővel, mint öt csoportot húsz-húsz fővel, holott a teljes elemszám mindkét esetben ugyanannyi. Ha az ICC nagyon magas, vagyis az osztályok közötti különbségek dominálnak, érdemes megfontolni a kísérleti elrendezés módosítását, hogy csökkentsük a klaszterződést (például egységesítsd a tanárok képzését). A kutatás tervezésekor ajánlott elemszámbecslést (power analysis) végezni az ICC, a várt hatásméret és a csoportok számának figyelembevételével. Általában legalább négy-hat csoport ajánlott szintenként a random hatás varianciájának megbízható becsléséhez.

Végül, a random hatás nem véletlenül kapta a nevét. Lényege, hogy a modellbe bevont szintek (a konkrét gyakorlati csoportok, osztályok vagy folyók) egy nagyobb populációból vett reprezentatív mintát alkossanak. Ha a mikroműanyag-szennyezettséget vizsgálva csak a legnagyobb folyókat (Duna, Tisza) választjuk ki, a kapott eredmény nem lesz általánosítható a kisebb folyókra. A modell általánosíthatóságának kulcsa tehát, hogy a beágyazott csoportokat lehetőség szerint véletlenszerűen válasszuk ki a populációból, és elegendő számban vonjuk be őket a variancia megbízható becsléséhez.

3.2. Adatgyűjtés, az adatok előkészítése

A beágyazott elrendezés sikeres elemzésének alapja a következetes adatgyűjtés, amely során a hierarchikus struktúra minden elemét pontosan dokumentáljuk. A függő változó rögzítése mellett elengedhetetlen, hogy minden egyes megfigyeléshez (például minden egyes hallgatóhoz) egyértelműen hozzárendeljük a hierarchia összes szintjét, a felsőbb szintű, szülő szempont (például az oktatási technika) szintjét és az alacsonyabb szintű, gyerek szempont (például a gyakorlati csoport) azonosítóját is. A legkritikusabb pont, és egyben a kezdő kutatók leggyakoribb hibája, a beágyazott (gyerek) szempont azonosítóinak egyedisége. Létfontosságú, hogy például az A technika alatti 1-es csoport és a B technika alatti 1-es csoport ne ugyanazt az azonosítót kapja. A szoftver számára a „csoport 1” az „csoport 1”, bárhol is fordul elő. Ezért aranyszabály, hogy a beágyazott szempont szintjeit globálisan egyedi azonosítókkal lássuk el (például 1–3. csoport az A technikánál, 4–6. csoport a B technikánál). Elemzés előtt egy egyszerű kereszttáblával mindig érdemes ellenőrizni, hogy a gyerek szempont egy szintje valóban csak egyetlen szülő szempont szint alatt jelenik-e meg.

Miután az adatokat összegyűjtöttük, a szoftver számára értelmezhető formába, úgynevezett *hosszú formátumba* (long format) kell rendeznünk azokat, ahol minden egyes sor egyetlen megfigyelési egységet (például egy diákot) képvisel, az oszlopok pedig a változókat tartalmazzák. A beágyazott ANOVA esetében a helyes adatszerkezet a 20. táblázatban látható módon néz ki.

Hallgató ID	Technika	Gyakorlati csoport	Elégedettség
1	A	1	19
2	A	1	21
3	A	1	22
...	...	...	...
29	A	3	32
30	A	3	14
31	B	4	34
...	...	....	...
44	B	5	38
45	B	5	29
...	...	...	...
60	B	6	17

20. táblázat: Példa a beágyazott ANOVA adatszerkezetére

Ez a struktúra egyértelműen és félreérthetetlenül tartalmazza a modell számára szükséges összes információt. Minden diákhoz tartozik egy függő változó érték, és minden diák egyértelműen hozzá van rendelve egyetlen gyakorlati csoporthoz és egyetlen oktatási technikához. Az így előkészített adatbázissal a szoftver már képes lesz helyesen értelmezni a beágyazott hierarchiát és lefuttatni a megfelelő elemzést.

3.3. Az elemzés beállításai

Bár a beágyazott ANOVA egy régebbi módszer, a legtöbb komoly statisztikai programcsomag ma is képes lefuttatni. Azonban az, hogy pontosan hol találjuk meg, és milyen logikával működik, sokat elárul a szoftver filozófiájáról és arról a statisztikai paradigmáról, amelyben gyökerezik.

25. doboz: A beágyazott ANOVA futtatása és beállításai SPSS és R környezetben

A nagy kereskedelmi szoftverek (például SPSS, SAS, Stata) a hagyományos és korszerű eljárások szinte teljes eszköztárát kínálják, így a modernebb LMM-en kívül meghagyták a hagyományos beágyazott ANOVA-t is mint a General Linear Model (GLM) családon belüli elemzési lehetőséget (például SPSS: Analyze > GLM > Univariate...), de a modell specifikálásához a kutatónak kell megadnia a beágyazott szerkezetet. A kutató szabadon dönthet, hogy a hagyományos keretrendszeren belül oldja meg a problémát, vagy a modernebb alternatívát választja.

Az R mint a legrugalmasabb statisztikai környezet szintén több lehetőséget kínál. Hagyományos megoldásként a beépített aov() funkció képes kezelni a beágyazott elrendezéseket. A modell specifikációja itt egy speciális Error() taggal történik, ami jelzi a hierarchiát. Ez a megközelítés feleltethető meg leginkább a beágyazott ANOVA logikájának. Mindezek ellenére az R-közösség elsöprő többsége ma már az lme4 (vagy nlme) csomagot és annak lmer() (Linear Mixed-Effects in R) funkcióját használja az ilyen adatok elemzésére. Bár ez már egy lineáris kevert modell, fontos látni, hogy az R-felhasználók számára ma már ez a természetes és alapértelmezett megközelítés a beágyazott adatok kezelésére.

Említésre méltó még a JASP és Jamovi. Ezek a modern, felhasználóbarát, menüvezérelt szoftverek érdekes helyzetet teremtenek. A hagyományos ANOVA-menüpontjukban nem találunk opciót a beágyazott random hatások specifikálására. Ezek a szoftverek a filozófiájukkal egyértelműen a modern megközelítés felé terelik a felhasználót. Nem kínálnak egy „lebutított" vagy korlátozott ANOVA-eljárást a beágyazott adatokra. Helyette a felhasználónak mindenképp a Mixed Model modult kell használnia, ahol a beágyazott szempontot Random Effectként specifikálhatja.

A beágyazott modell futtatásához két kulcsfontosságú lépést kell tennünk: 1) meg kell jelölnünk, mely szempontot tekintjük fixnek (ez lesz a szülő szempont), melyet randomnak (ez lesz a gyerek), és 2) fel kell építenünk a helyes modellstruktúrát. A fix szülő szempont önmagában kerül a modellbe, míg a gyerek szempont a szülőbe beágyazva, azaz a modellstruktúra két tagból áll: szülő és gyerek(szülő). Újdonság még, hogy a klasztereződés mértékét jellemző hatásnagyságot, az ICC értékét jellemzően nekünk kell kiszámolnunk a modell kimenetében szereplő szórásnégyzetekből (MS-értékekből). Minden más beállítás megegyezik az ANOVA-elemzéseknél megszokottal, kikérhetjük a leíró statisztikákat, a feltételteszteléshez a normalitás- és szóráshomogenitás-vizsgálatokat és a fix hatás hatásnagyságának értékét (például a parciális éta-négyzetet, η^2_p). A háttérben futó motor beállításait a szoftverek részben automatikusan végzik,

azaz kiegyensúlyozatlan adatok esetén automatikusan a robusztusabb Type III becslést alkalmazzák. Azonban a beágyazott modellek F-tesztjeinek ennél is bonyolultabb, Satterthwaite-típusú korrekcióit már nem végzik el automatikusan, ami pontosan rámutat a hagyományos ANOVA-keretrendszer korlátaira.

3.4. Az elemzés értelmezése

A beágyazott ANOVA lefuttatása után a kimeneti táblázatban jellemzően két fő F-statisztikát fogunk találni, amelyek a hierarchia különböző szintjeiről adnak felvilágosítást. Az első és legfontosabb a fix (szülő) szempont szignifikanciatesztje. Ha ez az F-érték szignifikáns, az azt jelenti, hogy a kutatásunk központi hipotézisét igazoltuk, azaz a felsőbb szintű csoportok (például az A és B oktatási technika) között valós különbség van, még azután is, hogy a modell statisztikailag kontrollálta az alacsonyabb szintű csoportosulás (a gyakorlati csoportok) hatását. Amennyiben a fix szempontnak kettőnél több szintje van (például A, B és C technika), az ANOVA-ban megszokott módon post hoc teszteket kell végezni, hogy kiderüljön, pontosan mely szintek között van a különbség.

A második F-statisztika a random (gyerek) szempont hatását teszteli. Ennek az eredménynek az értelmezése eltér a hagyományostól. Ha a gyakorlati csoport szempontja szignifikáns, az nem azt jelenti, hogy az 1-es csoport jobb, mint a 2-es. Ehelyett azt jelzi, hogy a csoportok között szignifikáns mértékű a variancia, vagyis a csoportok valóban különböznek egymástól annyira, hogy a teljesítményben megfigyelhető szóródás egy jelentős részét magyarázzák. Ez egy rendkívül fontos diagnosztikai eredmény, megerősíti, hogy a klasztereződés valós, és a beágyazott modell használata indokolt volt. Ha ez a hatás nem lenne szignifikáns, az azt sugallaná, hogy a csoportok valójában nem sokban különböznek egymástól, és a diákok közötti különbségek a fix szempont hatásán felül szinte teljes egészében egyéni szinten dőlnek el.

A szignifikanciateszten túl a random hatás legfontosabb mutatója az osztályon belüli korrelációs együttható (ICC). Míg a random szempont szignifikanciája csak azt mondja meg, hogy létezik-e a csoportosulás, az ICC számszerűsíti annak mértékét. Az ICC egy 0 és 1 közötti érték, ami azt fejezi ki, hogy a modell fix hatásai által meg nem magyarázott, fennmaradó (reziduális) variancia mekkora hányada (hány százaléka) magyarázható a csoportszintű különbségekkel. Például egy ICC = 0,22 érték azt jelenti, hogy miután kontrolláltunk az oktatási technikák hatására, az elégedettségben megfigyelhető maradék szóródás 22%-a a gyakorlati csoportok közötti eltérésekből fakad, míg a maradék 78% az egyéni, diákok közötti különbségekre vezethető vissza. Egy magas ICC-érték

tehát nemcsak a beágyazott modell választását igazolja, hanem rávilágít a csoportszintű tényezők (például a gyakorlatvezető személye, a csoportdinamika) kiemelt fontosságára is a vizsgált jelenségben.

4. Mikor használjuk és mikor ne használjuk a módszert?

Az összetettebb adatstruktúrák elemzésének legfontosabb, de egyben legnehezebb része, hogy eldöntsük, mikor milyen eljárás megfelelő hozzájuk. Mikor kell például ismételt méréses ANOVA-t használni, mikor beágyazott ANOVA-t, mikor többszintes modellezést vagy lineáris kevert modellezést (ez utóbbi kettőt a következő fejezet tartalmazza). Azonkívül, hogy tudatosítjuk magunkban a különböző módszerek egyértelmű strukturális feltételeit (egy összefoglaló táblát tartalmaz majd a következő fejezet), a legjobb tanulási mód a különböző példák átgondolása. Lássunk tehát ilyeneket!

Első példa: Egy agrárkutatásban három különböző műtrágya hatását vizsgálják a terméshozamra. Mindhárom műtrágyát 4–4, véletlenszerűen kiválasztott, független parcellán tesztelik, minden parcelláról 20-20 földmintát vételezve. Ez a beágyazott ANOVA klasszikus példája. A módszer jól működik, mert a parcellák be vannak ágyazva a műtrágyába, a felépítés kiegyensúlyozott, a fő szempont fix (három trágya), a beágyazott szempont random (a 4-4 parcella).

Második példa: Ugyanez a kutatás, de a B műtrágyát használó egyik parcellát elveri a jég, így onnan nincsenek adatok, egy másik parcellán pedig a tervezettnél kevesebb mintavételi pontról tudnak csak adatot gyűjteni. A tervezett struktúra ugyan megfelel a beágyazott ANOVA-nak, de az elrendezés kiegyensúlyozatlanná vált a megvalósítás során, ami a beágyazott ANOVA-t problémássá teszi, helyette a helyes választás az LMM.

Harmadik példa: Egy longitudinális gyógyszerkísérletben a pácienseket több időponton keresztül követjük, de a páciensek különböző orvosokhoz járnak, és az adatgyűjtés során többen lemorzsolódnak. A probléma a kiegyensúlyozatlan ismételt mérés és a keresztezett orvoshatás, a megoldás pedig itt is az LMM.

Negyedik példa: Vegyük a beágyazott ANOVA-ra tanult példát az oktatási technikákkal (gyakorlati csoport beágyazva a technikába). Tegyük fel, hogy a bevont diákok között vannak államilag finanszírozott és költségtérítéses

hallgatók is. A finanszírozási forma egy, a diákhoz kötődő tulajdonság, amely keresztezi a gyakorlati csoport szempontját, hiszen vélhetően minden csoportban lehetnek állami ösztöndíjas és költségtérítéses hallgatók is. A beágyazott ANOVA ezt az összetett (beágyazott és keresztezett is) struktúrát nem tudja kezelni, a megoldás itt is az LMM.

Ötödik példa: Egy kutató azt szeretné összehasonlítani, hogy az állami és a magániskolák diákjai között van-e különbség a természettudományos ismeretekben. Ehhez több állami és több magániskolából is gyűjt adatokat, és minden iskolán belül több osztályból is. A beágyazott ANOVA jó választás, de a kezdő kutató felcseréli ezt, és az Iskola(Osztály) struktúrát próbálná megadni a szoftvernek, míg a valóságban az osztályok vannak beágyazva az iskolákba, tehát a helyes modellben az Osztály(Iskola) beágyazottságot kell specifikálnunk. Bár a szoftver valószínűleg hibaüzenet nélkül lefuttatja a rosszul specifikált modellt, az eredményei teljesen értelmetlenek lesznek. A modell felépítésekor tehát mindig tegyük fel a kritikus kérdést: „Melyik egység tartalmazza a másikat?" A helyes válasz adja meg a szülő-gyerek viszonyt, amely a beágyazott modell alapja.

Hatodik példa: Azt vizsgáljuk, hogy a családi elégedettség eltér-e az apák, anyák és gyerekek körében. Bár van egyfajta beágyazottság, ez nem beágyazott ANOVA, mert a hierarchia fordított, mint amit ez a módszer elemezni tud. Ebben a vizsgálatban a szerep (apa, anya, gyerek) a fix szempont, míg a család a random szempont, de a beágyazott ANOVA-hoz szükségessel ellentétben nem a család van beágyazva a szerepbe (hiszen egy család nem tartozhat csak az apák vagy csak az anyák csoportjába), hanem a szerep mérései tartoznak egy-egy családhoz. A család a független egység, míg a szerep az egy családon belüli ismételt méréses szempont. Ha családonként egy gyerek van, akkor ez egy ismételt méréses ANOVA, ahol a család lesz a független adategység (az adatbázisban egy sor), a szerep pedig az adategységen belüli ismételt mérés. Az adatokat széles formátumban rögzítjük, ahol egy sor egy család, és külön oszlopban van az apa, az anya és a gyermek elégedettsége. Ha több, családonként változó számú gyerek is van, akkor a modell bonyolódik. Itt már az ismételt méréses ANOVA nem működik, mivel családonként változó számú gyerekoszlopra lenne szükség. A helyes megoldás ismét az LMM, amely képes egyszerre kezelni az ismételt mérést (szerep) és a beágyazottságot (szülők és gyerekek a családban).

Hetedik példa: Kíváncsiak vagyunk, hogy a kórházi osztályok típusa befolyásolja-e az orvosok kiégését. Három kórházból gyűjtünk adatokat, de a kórházakban eltérő osztályok működnek (például az egyikben van szülészet és ortopédia, a másikban szülészet, belgyógyászat és aneszteziológia stb.). Ez több okból sem beágyazott ANOVA. Először is, a hierarchia itt is fordított lenne, a valóságban az osztályok vannak beágyazva a kórházakba, nem pedig fordítva. Másodszor,

a szempontok valójában keresztezettek, mivel ugyanaz az osztálytípus (például szülészet) több kórházban is előfordul. Harmadszor, az elrendezés súlyosan kiegyensúlyozatlan, hiszen nincs minden kórházban minden osztálytípus. A beágyazott ANOVA-modell itt teljesen használhatatlan. Ez a helyzet az LMM-ek egyik legtipikusabb alkalmazási területe.

Nyolcadik példa: Két egyetemi büfé elégedettségét hasonlítjuk össze. Minden diák egy hétig az A büfében, egy másik hétig a B büfében ebédel, és minden nap értékel. A hagyományos megoldás itt az, hogy átlagoljuk minden diáknál a 7–7 napi értékelést. Így minden diákhoz két számított adatpont tartozik majd (egy A-átlag és egy B-átlag), amelyre összefüggő mintás t-próbát (vagy egy szempontos ismételt méréses ANOVA-t) futtathatunk. Ez egy helyes, de információt vesztő megoldás. A probléma ezzel a megoldással, hogy az átlagolással elveszítjük a napok közötti varianciát. Egy másik megoldás az lenne, ha két szempontos ismételt méréses ANOVA-t használnánk. Ekkor a diákok lesznek a sorok, és 14 oszlop lesz, sorban az A büfé 1, 2, 3… 7 napja, majd a B büfé 1, 2, 3… 7. napja. A két szempont a büfé és a nap száma. Ennek akkor van értelme, ha valami egyértelmű mintázatot várunk a napok között, például arra számítunk, hogy minden büfé az újdonság varázsa miatt az első napokban jó értékelést kap, de a hét végére unalmassá válik. Harmadik megoldás az LMM, amelyet akkor érdemes választani, ha ugyan figyelembe akarjuk venni a napok változatosságában lévő információt, de nem feltételezünk egyértelmű mintázatot, vagy az nem érdekes számunkra. Az LMM képes az összes adatpontot (7+7) egyszerre kezelni anélkül, hogy átlagolna, helyette a változatosságot a modell szerves részévé teszi. Fix hatásként modellezi a büfé hatását. Ez megadja a választ a fő kérdésre: van-e átlagos különbség A és B között? Random hatásként kezeli a diákokat, mindenkihez egyedi azonosítót rendelve. Ez kontrollálja azt a tényt, hogy egyes diákok eleve szigorúbbak vagy elégedettebbek (ez a diákok közötti variancia). Ami pedig a legfontosabb, a modellen belüli reziduális hiba a napról napra történő, diákon belüli ingadozást fogja megragadni. A modell „látni” fogja, hogy Béla reziduális hibája sokkal nagyobb, mint Annáé. Ennek eredményeként az LMM a büfé hatását egy sokkal gazdagabb információs környezetben becsüli meg. Nemcsak azt a kérdést teszi fel, hogy van-e különbség az átlagok között, hanem egy sokkal kifinomultabbat: „Miután figyelembe vettük, hogy a diákok eleve különböznek egymástól, és a véleményük napról napra ingadozik, van-e megbízható, szisztematikus különbség a két büfé között?” Ezzel egy statisztikailag erősebb és a valósághoz közelebb álló eredményt kapunk.

Kilencedik példa: Down-szindrómás és kontrollszemélyek oldanak meg matematikai feladatokat, hol vizuális segítséggel, hol anélkül. A résztvevők eltérő számú feladatot tudnak megoldani a rendelkezésre álló idő alatt. A hagyományos

megoldás az, hogy átlagoljuk a személyenkénti teljesítményt minden kondícióban. Ez azért problémás, mert ez a számítás az egyetlen feladatot megoldó személy átlagát (amely egyetlen pontból áll) ugyanolyan megbízhatónak tekinti, mint egy tíz feladatból számolt átlagot. A jobb megoldás megint csak az LMM, amely nem igényel átlagolást. Az összes megoldott feladatot bevonja az elemzésbe, és a személyt random hatásként kezelve természetes módon kezeli az egyénenként eltérő számú mérést, így a becslései sokkal megbízhatóbbak és pontosabbak lesznek.

5. A minőség biztosítása

Egy statisztikai elemzés közlésekor a célunk, hogy az olvasó számára ne csak az eredményt, hanem az egész gondolatmenetet és eljárást átláthatóvá és megismételhetővé tegyük. A tiszta közlés különösen fontos a beágyazott ANOVA esetében, amelynek módszertanával kapcsolatban gyakoriak a félreértések, és ennek megfelelően a szakirodalom hemzseg a kisebb-nagyobb mértékben félreértelmezett vagy félreérthetően közölt vizsgálatoktól. Az alábbi lista segít abban, hogy a beágyazott ANOVA eredményeiről szóló beszámoló teljes körű és szakszerű legyen, és felismerjük a hibásan elvégzett vagy pontatlanul közölt vizsgálatokat.

Szükséges közlés (minimum elvárás):

a) Egyértelműen közölje, hogy beágyazott ANOVA-t használt!
b) Nevezze meg a függő változót! Egyértelműen közölje, hogy melyik szempontot kezelte fix (szülő), és melyiket random (gyerek) hatásként, és indokolja a döntést!
c) Adja meg a fix szempont szintjeinek (a fő csoportoknak) átlagait és szórásait!
d) Mind a fix, mind a random (beágyazott) szempontra vonatkozóan közölje a legfontosabb statisztikai mutatókat, tehát az F-értéket, a két szabadságfokot és a p-értéket!
e) A szignifikáns fix hatásokat szükség esetén kövesse utóvizsgálattal!

f) Szignifikáns random hatásnál jelezze, hogy a klasztereződés jelentős varianciát ad a modellhez, de ne kövesse utóvizsgálattal a random csoportok különbségeit!
g) Fogalmazza meg egyszerű nyelven, hogy a statisztikai teszt eredménye mit jelent a kutatási kérdés szempontjából (például „Az oktatási technikák között szignifikáns különbség volt kimutatható…")!

Ajánlott közlés (jó gyakorlat) a szükséges közlésen felül:
a) Röviden magyarázza el, miért volt szükség beágyazott ANOVA-ra (például az adatok hierarchikus struktúrája, a klasztereződés miatt)!
b) A fix hatás esetében számoljon be valamilyen hatásnagyság-mutatóról (például parciális éta-négyzet, η^2_p)!
c) A random hatás esetében közölje az ICC értékét!
d) Említse meg, hogy az ANOVA alapfeltételeit (például reziduálisok normalitása, szóráshomogenitás) ellenőrizte, és azok teljesültek-e!

Kiváló közlés az ajánlott közlésen felül:
a) Egyértelműen közölje, hogy az elrendezés kiegyensúlyozott vagy kiegyensúlyozatlan volt-e! Kiegyensúlyozatlan esetben specifikálja a használt korrekciós eljárást (például Type III típusú négyzetösszegek)!
b) Ne csak közölje az ICC értékét, hanem értelmezze is azt a kutatás kontextusában (például: „Az ICC = 0,22 érték azt mutatja, hogy a teljesítményben megfigyelhető, fix hatások után maradó reziduális variancia 22%-a a gyakorlati csoportok közötti különbségekből fakad, ami jelentős csoportszintű hatásra utal.")!
c) Készítsen olyan ábrát, amely a beágyazott struktúrát is szemlélteti (például dobozdiagram vagy hegedűdiagram, ahol a csoportok külön-külön láthatók a fix szempont szintjein belül), ne csak összegző jellegű oszlopdiagramokat jelenítsen meg az átlagokról!
d) Egy kiváló elemzésben a szerző kitér a beágyazott ANOVA korlátaira (különösen kiegyensúlyozatlan esetben), és indokolja a választását (szemben például a robusztusabb LMM-mel). Ezzel bizonyítja, hogy a módszertani döntése tudatos volt.

6. A módszer változatai és haladó alkalmazásai

Láthattuk tehát, hogy a beágyazott ANOVA egy elegáns logikai keretrendszert kínál a hierarchikus adatokban rejlő függőségek kezelésére. A hagyományos, ANOVA-alapú megvalósítása azonban rendkívül merev, és leginkább a tökéletesen kiegyensúlyozott, egyszerű elrendezésekben működik megbízhatóan. Amint a valós kutatási helyzetek komplexitása megjelenik – legyen az kiegyensúlyozatlan adat, kettőnél több hierarchiaszint, beágyazott és keresztezett szempontok keveredése vagy akár ismételt mérések a beágyazott struktúrán belül –, a számítás nehézkessé válik, erejét veszti, és bonyolult matematikai korrekciókra szorul.

Ez az a pont, ahol a beágyazott ANOVA eléri a határait, és ahol a modern adatelemzés már más utat javasol. Ahelyett, hogy a régi keretrendszert próbálnánk meg „toldozni-foltozni", hogy valahogy megbirkózzon ezekkel a kihívásokkal, egyszerűen áttérünk egy olyan módszertanra, amelyet eleve az ilyen összetett, zajos adatok kezelésére terveztek. Ez a következő fejezetben részletesen tárgyalt, sokkal rugalmasabb és hatékonyabb keretrendszer, a lineáris kevert modellek világa.

7. Összefoglalás

A beágyazott ANOVA egy klasszikus statisztikai eljárás, amelyet kifejezetten a hierarchikus, egymásba ágyazott adatstruktúrák elemzésére fejlesztettek ki. Fő célja, hogy egy felsőbb szintű, minket érdeklő fix hatás (például egy oktatási módszer) hatását tesztelje, miközben statisztikailag kontrollálja az alsóbb szintű, random hatásként kezelt klaszterező tényező (például egy gyakorlati csoport) által okozott adatfüggőséget. Legfontosabb matematikai sajátossága, hogy a fix hatás F-statisztikájának kiszámításához nevezőként nem az egyéni hibát, hanem a beágyazott random szempont szórásnégyzetét használja, ezzel a hatást a releváns csoportszintű zajhoz méri. Bár az elgondolás elegáns,

a megvalósítása rendkívül merev, mivel érzékeny a kiegyensúlyozatlan adatokra, és nehezen birkózik meg a valós kutatások összetettségével. Éppen ezért a modern adatelemzésben leginkább már csak egy fontos elméleti lépcsőfoknak tekintjük, amely megalapozza a sokkal rugalmasabb és hatékonyabb lineáris kevert modellek megértését.

8. Önellenőrző kérdések

1. kérdés: Mi a legalapvetőbb különbség egy keresztezett és egy beágyazott kísérleti elrendezés között?

Válasz: A keresztezett elrendezésben a szempontok szintjei szabadon kombinálódnak, mint egy szorzótáblában. Ha például a nem (férfi/nő) és a kezesség (jobb/bal) hatását vizsgáljuk, a modellben mind a négy lehetséges kombináció (férfi-jobb, férfi-bal, nő-jobb, nő-bal) megjelenik. Ezzel szemben a beágyazott elrendezés hierarchikus, az egyik szempont szintjei csakis a másik szempont egyetlen szintjén belül léteznek. Például a 4. A osztály csak az Arany János Általános Iskolán belül létezik, egy másik iskola alatt nem, így az osztály be van ágyazva az iskolába.

2. kérdés: Mikor kell egy szempontot fixnek, és mikor randomnak tekinteni?

Válasz: Egy szempontot fix hatásnak tekintünk, ha a vizsgálatban szereplő szintjei önmagukban érdekesek, és a kutatásunk kizárólag ezekre a konkrét szintekre vonatkozik (például „A” vagy „B” oktatási technika). Ezzel szemben egy szempontot random hatásnak kezelünk, ha a szintjei csupán egy véletlenszerű mintát képviselnek egy nagyobb populációból, és nem a konkrét szintek közti különbség, hanem az általuk képviselt általános variancia érdekel minket (például a vizsgálatba bevont gyakorlati csoportok, amelyek az összes lehetséges csoportot reprezentálják).

3. kérdés: A beágyazott ANOVA-ban mit jelent a szülő és gyerek szempont, és hogyan dönthető el, hogy melyik melyik?

Válasz: A szülő és gyerek egyszerű elnevezése a modell hierarchikus szintjeinek. A szülő a felsőbb szintű, általában fix hatás, amelynek a hatására kíváncsiak vagyunk (például oktatási technika). A gyerek az alacsonyabb szintű, beágyazott, általában random hatás, amely a klasztereződést okozza (például gyakorlati csoport). A döntéshez mindig tegyük fel a kérdést: „Melyik egység tartalmazza a másikat?" Az egyes osztályok mindig csak egy oktatási technikához tartoznak, tehát a technika a szülő, az osztály pedig a gyerek. A helyes szintaxis mindig gyerek(szülő).

4. kérdés: Mit jelent pontosan a kiegyensúlyozottság egy beágyazott modellben, és miért olyan fontos?

Válasz: Egy beágyazott modell akkor kiegyensúlyozott, ha a hierarchia minden szintjén egyenlő az elemszám. Ez két dolgot jelent egyszerre: 1) minden szülő csoport alatt pontosan ugyanannyi gyerek csoport van (például minden technikánál 4-4 gyakorlati csoport), ÉS 2) minden gyerek csoportban pontosan ugyanannyi megfigyelés van (például minden csoportban 15 diák). A beágyazott ANOVA rendkívül érzékeny erre, mert a helyes F-statisztika kiszámítása a kiegyensúlyozott arányokon alapul.

5. kérdés: Hogyan sértheti a függetlenség feltételét egy látszólag független csoportos elrendezés, és miért vezet ez hibás következtetéshez?

Válasz: Ez a beágyazott modellek létezésének legfőbb oka. Egy elrendezésben a fő csoportok lehetnek függetlenek (például a diákok vagy az A, vagy a B technikával kapnak oktatást), de a mérések mégsem függetlenek, ha a diákok kisebb csoportokba, klaszterekbe tömörülnek (például gyakorlati csoportokba járnak). Az egy csoportba tartozó diákok a közös élmények (például tanár, csoportdinamika) miatt jobban hasonlítanak egymásra, mint más csoportokba tartozó társaikra. Ezt a rejtett függőséget mérjük az osztályon belüli korrelációval (ICC). A t-próba vagy ANOVA nem veszi figyelembe az ICC-t, ezért alábecsüli a valódi hibát. Ez mesterségesen megnöveli az F-értéket, ami az elsőfajú hiba kockázatának drámai növekedéséhez vezet.

Ajánlott irodalom

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and Other Stories*. Cambridge University Press. https://doi.org/10.1017/9781139161879

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2022). *Multilevel and Longitudinal Modeling with IBM SPSS*. Routledge. https://doi.org/10.4324/9780367824273

Hove, D. T., Jorgensen, T. D., & van der Ark, L. A. (2024). Updated guidelines on selecting an intraclass correlation coefficient for interrater reliability, with applications to incomplete observational designs. *Psychological Methods, 29*(5), 967–979. https://doi.org/10.1037/met0000516

Hurlbert, S. H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs, 54*(2), 187–211. https://doi.org/10.2307/1942661

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2024). *Designing experiments and analyzing data: A model comparison perspective*. Routledge Taylor & Francis Group.

McElreath, R. (2020). *Statistical Rethinking*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429029608

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer. https://doi.org/10.1007/b98882

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Stroup, W. W., Ptukhina, M., & Garai, J. (2024). *Generalized Linear Mixed Models*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429092060

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics*. Pearson.

West, B. T., Welch, K. B., Gałecki, A. T., & Gillespie, B. W. (2022). *Linear Mixed Models*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003181064

XII. LINEÁRIS KEVERT MODELLEK (LMM)

SOLTÉSZ-VÁRHELYI KLÁRA

A fejezet a következő kérdéseket válaszolja meg (feltételezve a beágyazott ANOVA fejezetének feldolgozását):

a) Milyen elméleti alapokon nyugszik a lineáris kevert modellek módszertana?
b) Milyen adatstruktúrák elemzésére alkalmas a módszer?
c) Hogyan kell az elemzést helyesen beállítani, az eredményeket helyesen értelmezni?
d) Milyen minőségi ismérvek alapján különböztethető meg egy kiváló tanulmány egy átlagostól?

1. Elméleti háttér

A lineáris kevert modellek (Linear Mixed Models, LMM) világa egy olyan modern és rendkívül rugalmas keretrendszert kínál, amely nem riad vissza a valós kutatási adatok kuszaságától. Legyen szó akár hiányos, kiegyensúlyozatlan elrendezésről, időben ismételt mérésekről vagy egymást keresztező hatásokról, az LMM a hagyományos eljárásokon messze túlmutató rugalmasságot nyújt azáltal, hogy modellezi az adatpontok közötti finom, rejtett függőségi hálózatot, és ezzel a modern adatelemzés egyik legfontosabb módszertanává válik.

Arra buzdítjuk az olvasót, hogy úgy kezdjen hozzá az LMM tudásanyagának feldolgozásához, hogy előtte átolvasta az előző, beágyazott ANOVA-ról szóló fejezetet, mivel ott részletesen ismertettük azokat a módszertani, kísérleti elrendezést érintő problémákat, amelyekre az LMM megoldásként született. Ennek ellenére, mielőtt rátérnénk az LMM működésére, mivel az mindenképpen szükséges az LMM működésmódjának megértéséhez, ebben a fejezetben újra, de most röviden áttekintjük a hagyományos kísérleti elrendezéseket és az azokkal kapcsolatos leggyakrabban előforduló problémákat.

1.1. A hagyományos kísérleti elrendezések

A hagyományos ANOVA-modellekkel kapcsolatban általában azt szokás megtanítani, hogy a kísérleti elrendezéseket kétféle alapvető *mérési szempont* (vagy faktor, angolul factor) mentén lehet felépíteni. Az első a *független mintás* (between-subject) szempont, ahol az összehasonlítás személyek között történik, és a vizsgált szempont független csoportokat tartalmaz. Erre láttunk példát az előző fejezetben, amikor azt vizsgáltuk, hogy az autizmusspektrum-zavar (autista és kontrollcsoport) és a nem (férfi és nő) hogyan hat az érzelemfelismerésre. Ebben a kutatásban mind az autizmus, mind a nem független szempontok voltak. (Vigyázzunk, ne keverjük a független mintás kifejezést a függetlenség általános feltételével!) A második az *ismételt méréses* vagy összefüggő mintás (within-subject) szempont, mely a személyeken belüli megváltozást méri, azaz ugyanazok a személyek vesznek részt minden mérésben vagy kondícióban. A modellt az előző fejezet hideg/meleg, piros/kék tejjel végzett kísérlete illusztrálta, ahol minden résztvevő kapott hideg és meleg tejet is, pirosat és kéket is. Természetesen elképzelhető olyan kísérleti elrendezés is, ahol független mintás

és ismételt méréses szempontok keverednek (ezt neveztük vegyes ANOVA-nak). Erre az A és B típusú terápiák hatékonyságának vizsgálata volt a példa, ahol a terápia típusa független, míg a mérés időpontja (terápia előtt és után) ismételt méréses szempont volt.

Bár az előbb bemutatott kísérleti modellek látszólag különbözőek, közös bennük az az alapvető szerkezeti tulajdonság, hogy mindegyik *keresztezett elrendezésre* (crossed design) épül. Keresztezett elrendezésről akkor beszélünk, ha a modellben szereplő egyik szempont minden szintje kombinálódik a többi szempont összes szintjével. A nem és autizmus példájánál ez azt jelenti, hogy a kísérleti tervben létezik autista férfi, autista nő, kontrollférfi és kontrollnő is, vagyis a 2x2-es táblázat minden cellája logikailag létezik, és ideális esetben adatokat is tartalmaz. A hagyományos ANOVA-modellek túlnyomó többsége ezt a teljes, szisztematikus keresztezettséget feltételezi.

Ebből a keretrendszerből jelentett egyfajta kilépést a *beágyazott ANOVA* (nested ANOVA), amelyben a szempontok nem keresztezik egymást, hanem hierarchikusan egymás alá vannak rendelve. Jó példa lehet egy iskolások teljesítményét vizsgáló kutatás, ahol bizonyos osztályok egyfajta fejlesztést kaptak, más osztályok pedig másik fajta fejlesztést. Ez a vizsgálat tipikus példa a beágyazott elrendezésre, mert a modellnek nincsenek keresztezett cellái (egy osztály nem tartozhat több fejlesztéstípushoz is).

Összegezve, egy kísérleti elrendezést jellemezhetünk a benne szereplő szempontok természete (független mintás vagy ismételt méréses) és egymáshoz való viszonya (keresztezett vagy beágyazott) alapján. E szigorú megkülönböztetésnek alapvető oka a mérések között fennálló, de eltérő természetű *függőségi viszonyok* helyes modellezésének szükségessége.

Az ismételt méréses elrendezésben a függőség forrása maga a személy.[1] Az egyazon egyéntől származó mérések azért korrelálnak, mert egy közös, rejtett tényező (az egyén stabil képességei, tulajdonságai) köti őket össze. A beágyazott elrendezésben pedig a függőséget egy külső, közös csoporttagság okozza. Az egy osztályba tartozó diákok mérései nem tekinthetők függetlennek, hiszen a közös tanár, a csoportdinamika és a közös élmények miatt teljesítményük klasztereződik, azaz az egy csoportba tartozó diákok hasonlítani fognak egymásra. A csoporton belüli hasonlóságot az *osztályon belüli korreláció* (intraclass correlation, ICC) számszerűsíti, amely megadja, hogy a teljes varianciából mekkora arányt magyaráz a csoportba (például iskolai osztályba) tartozás. Ha e

[1] Itt csak röviden említjük meg, hogy bár az egyszerűség kedvéért gyakran személyekre hivatkozunk, az ismételt mérés alapegysége bármilyen entitás lehet (pl. egy vállalat, egy ország vagy egy biológiai minta), amelyen időben egymást követő méréseket végzünk. Az ilyen típusú adatok specifikumaira a későbbi fejezetekben térünk vissza.

függőségeket nem vesszük figyelembe, akkor a *pszeudoreplikáció* hibájába esünk, ami drasztikusan növeli az elsőfajú hiba valószínűségét. A hagyományos eljárások a különböző kísérleti elrendezések függőségeit csak rendkívül szűkös keretek között, nem optimálisan kezelik.

Első problémaként elvárják a szigorú keresztezettséget (azt, hogy a változók szintjeinek minden kombinációjára legyen adat) vagy a szigorú beágyazottságot (egy csoport szigorúan csak egy fölötte lévő kategóriában jelenhet meg).

Második problémaként érzékenyek a *kiegyensúlyozottság* (balanced design) feltételére. Egy kísérleti elrendezést akkor nevezünk kiegyensúlyozottnak, ha a vizsgálat minden csoportjában vagy feltételében azonos számú megfigyelés található. Egy keresztezett elrendezés akkor kiegyensúlyozott, ha a szempontok minden lehetséges kombinációjában pontosan ugyanannyi adatpont van. Például az autista férfiak, autista nők, kontrollférfiak, kontrollnők csoportjában is azonos az elemszám. Egy beágyazott elrendezés kiegyensúlyozottsága azt jelenti, hogy a hierarchia minden szintjén azonos számú egység van (például mindkét oktatási technika alatt pontosan három-három osztály van, és minden osztályban pontosan tizenöt diák). Amint az osztályok száma vagy elemszáma eltérővé válik – ami a valós kutatásokban szinte elkerülhetetlen –, a hagyományos ANOVA-modellek megbízhatósága jelentősen csökken. Keresztezett elrendezéseknél ez a probléma legtöbbször néhány egyszerű korrekcióval orvosolható (ilyen például a beágyazott ANOVA fejezetben tárgyalt Type III számítás), de a beágyazott adatok elemzését a kiegyensúlyozatlanság akár lehetetlenné is teheti.

A harmadik, komoly korlát a *statisztikai erő pazarlása*. Tegyük fel, hogy felismerjük az ismételt mérésekből vagy a beágyazottságból fakadó függőségeket, és megpróbáljuk őket kezelni azzal, hogy a függőséget okozó tényezőt (például személy azonosítója, osztályt jelölő változó) beemeljük a modellbe prediktorként. Bár ez bizonyos esetekben matematikailag lehetséges, egyáltalán nem hatékony, mert a hagyományos modellekben minden egyes prediktor bevonásának statisztikai költsége van. *Szabadságfokokat* (degrees of freedom, df) használ fel, ami a modell becslési kapacitását és erejét jelenti. Egy k szintű csoportosító változó prediktorként való modellbe vonása k-1 szabadságfokot köt le. Ha például tíz különböző ZH van egy félév során, az időpontok hatásának modellezése kilenc szabadságfokot emészt fel. Ha tizenöt különböző iskolai osztályból gyűjtünk adatot, az osztályok hatásának figyelembevétele tizennégy szabadságfokba kerül. Ez hatalmas pazarlás, mivel gyakran egyáltalán nem vagyunk kíváncsiak arra, hogy a második ZH szignifikánsan nehezebb volt-e, mint a nyolcadik, vagy hogy a B osztály szignifikánsan jobban teljesített-e, mint a G. Célunk csupán a *klaszterződésből fakadó függőség kontrollálása*. A szempont prediktorként

való kezelésével a modell rengeteg szabadságfokot pazarol el olyan, számunkra érdektelen összehasonlítások becslésére, amelyek nem is részei a kutatási kérdésünknek. Feleslegesen terheljük a modellt, csökkentve ezzel a statisztikai erejét a minket valóban érdeklő hipotézisek tesztelésére.

Csak hab a tortán, hogy a valós kutatási helyzetek ritkán tiszták. Mi történik, ha egy elrendezés egyszerre tartalmaz beágyazott és keresztezett szempontokat? Vagy ha egy longitudinális vizsgálatban a résztvevők lemorzsolódnak, és így személyenként eltérő számú mérésünk van? Mi a teendő, ha a csoportok elemszáma drasztikusan eltér, vagy ha bizonyos csoportkombinációkból egyáltalán nincsenek adataink? Hogyan kezeljük azt a helyzetet, amikor egy személyhez nem egyetlen adatpont, hanem több tucatnyi egyedi mérés (például reakcióidő) tartozik, és ezt az információt nem akarjuk egyetlen átlagba sűrítve elveszíteni? Hogyan elemezzünk egy ismételt méréses elrendezést, ha a mérések nem fix időpontokban (például előtte és utána), hanem minden személynél egyedi, folytonos időskálán történtek? Mit tegyünk, ha nemcsak a kísérleti személyek különböznek egymástól, hanem a kísérletben használt ingerek is eltérő nehézségűek, és erre az ingerhatásra is szeretnénk kontrollálni? Végül, hogyan vizsgálhatjuk meg, hogy egy kezelés vagy beavatkozás hatása (például egy készség fejlődésének üteme) egyénenként eltérő-e, ahelyett, hogy csak egy átlagos hatást becsülnénk az egész mintára? Az ANOVA-módszerek ezekre a kérdésekre csak nehézkes korrekciókkal vagy egyáltalán nem tudnak választ adni.

Egyértelművé vált tehát, hogy szükség van egy olyan átfogó eljárásra, amely egyszerre képes kezelni a keresztezett és beágyazott hatásokat, valamint a független és ismételt méréses szempontokat anélkül, hogy a kiegyensúlyozatlanság, a hiányzó adatok vagy a rendszertelen mérések falába ütközne, és mindezt képes hatékonyan, a statisztikai erő pazarlása nélkül megvalósítani. Az LMM megjelenése pontosan ezt a paradigmaváltást hozta el. Itt már nem az ANOVA-modellek folyamatos javítgatásáról és bővítgetéséről van szó. Ehelyett elhagyjuk a klasszikus keretrendszert, és egy új, valószínűségi alapú megközelítést alkalmazunk, amely sokkal nagyobb szabadságot és pontosságot kínál a modern adatelemzés összetett kihívásainak kezelésére.[2]

Mindezzel azonban az LMM egy újfajta kihívás elé állította a *tankönyvírókat*, mivel rugalmassága miatt nehéz egységes választ adni az olyan kérdésekre, mint hogy „mit csinál?" és „hogyan működik?", mert a tényleges megvalósulása nagymértékben függ az adatok jellegzetességeitől és a kutató szándékaitól is.

[2] Bár a kutatói gyakorlatban gyakran versenyző alternatívaként kezeljük őket, matematikailag pontosabb úgy fogalmazni, hogy az LMM a lineáris modelleknek (így az ANOVA-nak és a regressziónak) az általánosított és kiterjesztett formája. Az ANOVA valójában az LMM egy speciális esete, egy olyan LMM, amelyben nincsenek random hatások, csupán a fix hatások és a reziduális hiba.

A jelen tankönyv írója ezt a kihívást úgy próbálja leküzdeni, hogy tucatnyi példa felsorolásán keresztül szemlélteti az LMM működését, remélve, hogy a végére sikerül kialakítani egy egységes képet róla.

1.2. Fix és random hatások

Az LMM a hagyományos eljárások kapcsán felsorolt nehézségek gordiuszi csomóját egyetlen konceptuális váltással vágja át, és a prediktorokat két, alapvetően eltérő szerepkörre osztja, elkülönítve a fix és random (véletlenszerű) hatásokat.

A *fix hatások* a kutatási kérdés szempontjából fontos tényezők. A „Mi a különbség A és B minta között?" típusú kérdésekre válaszolnak, és olyan hipotéziseket tesztelnek, amelyek konkrét, előre meghatározott csoportok összehasonlítására irányulnak. Hatékonyabb-e az A terápia a B terápiánál? Különbözik-e a férfiak és a nők problémamegoldó képessége? Jobban teljesítenek-e a diákok a kísérleti csoportban, mint a kontrollcsoportban? Alacsonyabb-e a szorongás a terápiát követően, mint a terápiát megelőzően volt? Ennek értelmében a fix hatások a hagyományos regressziós módszertan prediktorainak felelnek meg. A modell ezen tényezők hatását közvetlenül értelmezhető paraméterek (*regressziós együtthatók, β*) formájában becsüli meg, pontosan úgy, ahogy azt a korábbi eljárásoknál megszoktuk. A kimeneten egy konkrét értéket kapunk minden egyes csoportra vagy összehasonlításra, amely megmutatja a hatás nagyságát és irányát (például a kísérleti csoport átlagosan 5.2 ponttal teljesített jobban). Az eredményeket kizárólag a vizsgálatba bevont szintekre általánosíthatjuk, tehát ha azt találjuk, hogy az A terápia jobb, mint a B, ez a következtetés csak erre a két konkrét terápiára érvényes, egy C terápiára már nem. A fix hatások számára az LMM megengedi, hogy továbbra is felvegyék a becslésükhöz szükséges szabadságfokokat.

A *random hatások* közé kerülnek azok a klaszterező tényezők, amelyek a függőségi struktúrát okozzák, de amelyek szintjeinek egyedi összehasonlítása nem célja a kutatásnak. Ilyenek az ismételt méréseknél a személyek azonosítói vagy a klaszteredést okozó osztályhatások. A random hatások a „Mekkora a variabilitás…?", „Mekkora a klaszteredés mértéke…?" típusú kérdésekre adnak választ. Nem az egyes szintek összehasonlítása a cél, hanem az általuk képviselt alapsokaság általános szóródásának megértése. Mekkora az átlagos különbség az iskolák között a diákok átlagteljesítményében? Mennyire tér el az egyes páciensek gyógyulási üteme? Mekkora varianciát visz a modellbe a mérést végző asszisztensek személye?

Egy szempontot akkor kezelünk random hatásként, ha a vizsgálatba bevont szintjei csupán egy véletlenszerű mintát képviselnek egy sokkal nagyobb, lehetséges szintekből álló populációból. Ilyenkor nem a konkrét, bevont szintek (például a tizenöt konkrét osztály, a húsz konkrét kísérleti személy, a tíz konkrét orvos) közötti különbség érdekel minket. A célunk csupán az, hogy az általuk képviselt varianciát, a különbözőségeikből származó zajt vagy adatstruktúrát megbecsüljük és modellezzük. Ez két dolgot is eredményez. Egyrészt az eredményeket a szintek teljes populációjára általánosítjuk. Amikor megbecsüljük az osztályok közötti eltérést, az eredmény nemcsak a mintánkban szereplő tizenöt osztályra, hanem ideális esetben az ország összes osztályára vonatkozó következtetést enged meg. Másrészt a modell nem pazarol el erőforrásokat arra, hogy minden egyes személy vagy csoport hatását külön paraméterként megbecsülje. Ehelyett egy sokkal hatékonyabb megoldást választ, és csak az egyedi, torzító hatások *varianciáját (σ^2, szigma-négyzet)* becsüli meg. Nem azt vizsgálja, hogy A, B, C és a többi osztály különbözik-e egymástól, csak azt, hogy összességében mekkora a változatosság az osztályok teljesítményében. Ez a megközelítés rendkívül hatékony a szabadságfokokkal. Ahelyett, hogy tizenöt osztály tizennégy szabadságfokot emésztene fel a modellből, a változó random hatásként való kezelésével csupán egyetlen paramétert (a csoportátlagok varianciáját) kell megbecsülnünk. Így a modell képes figyelembe venni az osztályból fakadó klaszterződést anélkül, hogy a minket valóban érdeklő hipotézisek tesztelésétől vonná el a statisztikai erőt.

Ez a kettősség adja a kevert modellek erejét. Azt, hogy képesek vagyunk a fő hipotéziseinket (fix hatások) tesztelni, miközben pontosan és rugalmasan modellezzük az adatokban rejlő függőségi struktúrákat (random hatások), legyen az a személyek, iskolák, ingerek vagy bármilyen más klaszterező tényező miatti variancia. Mindezeknek az az eredménye, hogy az LMM egyszerre kezeli a függőségeket és mégis maximalizálja a statisztikai erőt. A két hatás összehasonlítását a 21. táblázatban összegeztük.

Tulajdonság	Fix Hatás	Random Hatás
Cél	Konkrét szintek összehasonlítása.	A szintek populációja varianciájának becslése.
Érdeklődés	Az átlagok közötti különbség.	A szintek közötti variancia.
Modell kimenete	Minden szinthez egy együttható (β).	Az eloszlás varianciája (σ^2).
Költsége	A fix hatás minden szintje k-1 szabadságfokba kerül.	A random hatás modellbe emelése csak néhány (legtöbbször egy) szabadságfokot használ fel a random hatás szintjeitől függetlenül.
A szintek	A vizsgálat kimeríti az összes releváns szintet.	A szintek egy nagyobb populációból vett minta.
Általánosítás	Csak a vizsgált, konkrét szintekre.	A szintek teljes populációjára.

21. táblázat: Fix és random hatások tulajdonságainak összehasonlítása

1.3. Az adatok strukturálása

Képzeljük el, hogy egyetemi tanárként arra vagyunk kíváncsiak, hogy a délutáni kurzusokra járó hallgatók rosszabbul teljesítenek-e, mint a reggeliek. A vizsgálatba tíz kurzust vonunk be, ötöt-ötöt a Pedagógia és a Pszichológia Intézetből, a csoportlétszám 14-16 fő között változik, összesen 150 fő, és a hallgatók minden héten zárthelyi dolgozatot (ZH-t) írnak, a félév során összesen tizenegyet.

A LMM működésének megértését segíti, ha megfigyeljük, milyen formátumban várja az adatokat. A hagyományos elemzéseknél a *széles formátum* (wide format) a megszokott, ahol egy személy minden adata ugyanabban a sorban helyezkedik el, és az ismételt mérések külön változókba vannak rendezve (például a ZH1, ZH2, ..., ZH11 külön oszlopban helyezkedne el). Ezzel szemben az LMM az úgynevezett *hosszú formátumot* (long format) várja (22. táblázat). Ebben a formátumban minden egyes mérés külön sort kap, azaz az ismételt mérések is egymás alatt helyezkednek el. A hosszú formátumba rendezett adatoknál egyetlen változó (példánkban a Teljesítmény) lesz csak kimeneti változó szerepben, amely egyetlen oszlopot jelent az adatbázisban (szemben az ismételt méréses ANOVA elrendezésével, ahol minden ZH-eredmény külön oszlopba került volna). Az adatok minden strukturális jellemzőjét (például az ismételt mérések összekapcsolását) változók felvételével kell megoldanunk.

Az egy személyhez tartozó, összefüggő mérések jelöléséhez felveszünk egy hallgatói azonosítót (Hallgató ID), és mivel tizenegy ZH-t írt mindenki, ugyanaz az azonosító tizenegy sorban fordul elő. A ZH-k beazonosításához felvételre kerül egy ZH-sorszám változó 1-től 11-ig menő értékkel. Van információ arról is, hogy a ZH javítását a tanári karból ki végezte el (Pontozó ID 1-től 5-ig menő értékekkel). A hallgatókkal kapcsolatos információként van egy változó annak jelölésére, hogy a személy melyik kurzusra jár (1-től 10-ig menő Kurzus ID), melyik intézethez tartozik (Pedagógia vagy Pszichológia) és olyan egyéb adatokat is kódolunk, mint a hallgató neme, kora vagy a felkészüléssel töltött idő stb. Mindezeken túl kutatásunk fő kérdésének megválaszolásához egy változó kódolja azt is, hogy reggeli vagy délutáni kurzusra járt-e a hallgató (Napszak). A kezdő adatelemző itt elbizonytalanodhat, hogy nem probléma-e, hogy ugyanazt az adatot (például a személy neme) ebben az elrendezésben tizenegyszer is megadja. A válasz természetesen az, hogy nem probléma, sőt szükséges is, hogy így legyen.

Hallgató ID	ZH-sorszám	Pontozó ID	Intézet ID	Kurzus ID	Nem	Kor	Felkész	Napszak	Teljesítmény
1	1	1	Pszi	1	Nő	23	2	Reggel	85
1	2	4	Pszi	1	Nő	23	4	Reggel	88
…	…	…	…	…	…	…	…	…	…
1	11	2	Pszi	1	Nő	23	1	Reggel	82
2	1	1	Ped	6	Férfi	21	6	Délután	90
2	2	4	Ped	6	Férfi	21	5	Délután	76
…	…	…	…	…	…	…	…	…	…
2	11	3	Ped	6	Férfi	21	3	Délután	75
3	1	1	Pszi	2	Nő	24	6	Reggel	55
…	…	…	…	…	…	…	…	…	…

22. táblázat: Hosszú formátumú adattábla

1.4. A varianciák modellezése

A hagyományos varianciafelbontáson alapuló eljárások (lineáris regresszió, ANOVA) egyenlete szerint a kimeneti változó értékét a prediktorokból becsült érték és a hiba összege adja. Ennek megfelelően az egyenlet:

$$Y = X\beta + \varepsilon$$

A képletben Y a kimeneti változó tényleges értékeit tartalmazó vektor (a példánkban a teljesítményértékek). Az X a prediktor változók értékeinek mátrixa (az egyszerűség kedvéért tegyük fel, hogy csak egy prediktor van, a napszak, melynek értékei lehetnek a reggel és a délután). A β a regressziós együtthatók vektora, így az Xβ a kimenethez becsült értékeket tartalmazó vektor lesz, ahol a reggeli kurzusokra járó hallgatókhoz a reggeli kurzusok átlagos teljesítményértékét, a délutános hallgatókhoz a délutáni kurzusok átlagos teljesítményét becsüljük. Végül az ε (epszilon) pedig a becslés hibáját tartalmazó vektor, azaz a becsült és mért értékek eltérése.

Ennek megfelelően a regresszióban a *kimeneti változó teljes varianciája* két részre bomlik: a *prediktorok által megmagyarázott varianciára* (a mi esetünkben a ZH-teljesítmények közötti változatosság egy része abból adódik, hogy a reggeliek magasabb teljesítményt nyújtanak, mint a délutániak) és a *maradékvarianciára*, más néven a *reziduumra* vagy *reziduális varianciára* (tehát arra, hogy a reggeli kurzuson írt ZH-k között is vannak jobban sikerültek és kevésbé jók, és hasonló a helyzet a délután írt ZH-knál is). A hagyományos eljárások itt meg is állnak, mivel ezt a maradékot egyetlen, struktúra nélküli zajként kezelik, amelynek elemeit egymástól függetlennek tekinti (a reziduális hibák függetlensége a lineáris regresszió alapfeltevése).

Ha a mérések tényleg függetlenek egymástól (és ez sok kutatásban így van), akkor ez a megközelítés elegendő. A példában szereplő adatokhoz regressziót választani azonban hiba, mert itt a reziduálisok valójában nem függetlenek egymástól, hiszen egy adott hallgató ZH-teljesítményei vagy az egy kurzusba járó hallgatók eredményei összetartoznak, és ha ezt nem vesszük figyelembe, akkor a pszeudoreplikáció hibájába esünk, ami az elsőfajú hiba valószínűségét jelentősen növeli. Felmerülhet megoldásként, hogy akkor tegyük be prediktorként az egyéb csoportosító tényezőket is, de ez a legtöbb esetben teljességgel lehetetlen. Nem tehetjük be prediktornak a hallgatói azonosítót 150 hallgatóval! De még ha olyan szerencsések is vagyunk, hogy elméletileg megoldható az új prediktor bevitele (például a ZH-kat pontozó tanár személye talán beemelhető lehetne prediktorként), akkor is feleslegesen terheljük a modellt (növeljük

a modell komplexitását), hiszen a tanárok közötti eltérés valójában nem fókusza a kutatásunknak. Itt jön az LMM újítása. A random hatások (például a Hallgató ID, a Pontozó ID vagy a Kurzus ID) bevezetése pontosan ezt a problémát oldja meg. Az LMM a nem független megfigyelésekből fakadó varianciát és kovarianciát expliciten modellezi, így figyelembe veszi a függőségeket, de mindezt a fix hatásoktól elkülönítve teszi, hiszen a random hatások (legtöbbször) nem érdekesek a kutatási kérdéseink szempontjából.

Lássuk, hogyan éri ezt el! Az LMM egyenlete egy árnyalattal komplexebbé válik a regresszióhoz képest, mivel egy új taggal bővül:

$$Y = X\beta + Zu + \varepsilon$$

Az egyenletben az Xβ továbbra is a fix hatások által becsült értékek vektora, a Zu a random hatások hozzájárulása a varianciákhoz és kovarianciákhoz, az ε pedig a valódi, megmagyarázhatatlan maradékhiba.

Ezzel együtt a variancia kezelése is megváltozik, és ez adja a LMM másik konceptuális váltását a hagyományos eljárásokhoz képest. Ahelyett, hogy a teljes varianciát egyszerűen kettéosztanánk a fix prediktorok és a reziduálisok összegére, az LMM logikája a következő:

A fix hatások esetében a csoportok átlagai és azok eltérése érdekel minket, ezért a fix hatások az adatok *átlagstruktúráját* (mean structure) modellezik. Azt írják le, hogyan változik a kimeneti változó *várható értéke* a különböző csoportokban (például eltérő átlagok becslése a reggeli és délutáni csoportoknál). Ezt az Xβ tartalmazza, és a becslés hasonlít arra, amit a regresszióban megszoktunk. Emellett az LMM-modellben is definiálható minden mérésre a becsült érték és a ténylegesen mért érték eltérése (a regresszióban ez lenne a reziduális értéke, és ennek szóródása lenne a reziduális variancia), és a random hatások vizsgálatakor a varianciák és kovarianciák becslése már nem a nyers ZH-pontszámokat, hanem logikailag ezt a fix hatásokra kontrollált változatot veszi alapul, ennek megfelelően *a random hatások elemzése az adatok varianciájának fix hatásokkal csökkentett, megmagyarázatlan részén történik*. Ez azt is jelenti, hogy a fix hatások által magyarázott variancia nem jelenik meg a varianciák modellezésének képleteiben (hanem a modell illeszkedési mutatóiban, például a marginális R^2-ben fog majd tükröződni, lásd később). Ez segíti az LMM-et abban, hogy a fix és random hatásokat szétválasztva modellezze.

Itt vezetjük be a *V-mátrixot*, amely ezeknek a fix hatások figyelembevétele mellett megmaradt varianciáknak és kovarianciáknak a feltételezett struktúráját írja le. A V-mátrix két részből épül fel, a G- és R-mátrixból. A *G-mátrix* a random hatások (például a hallgatók közötti különbségek) feltételezett varianciáját

és kovarianciáit tartalmazza, és az egyenletben a Zu-taghoz kötődik. Az *R-mátrix* pedig a valódi, megmagyarázhatatlan maradékhiba feltételezett struktúráját írja le. Az egyenletben az ε-taghoz kötődik, és a G-mátrix által leírt struktúrán túl megmaradó, de jellemzően már egyszerűbb függőségi mintázatokat modellezi.

Ezen a ponton tisztáznunk kell három gyakori félreértést.

Az első a fix és random hatások sorrendiségének kérdését érinti. Fontos hangsúlyozni, hogy a fix és random hatások becslése *egyidejűleg* (szimultán), egyetlen modell keretein belül történik. A kapcsolat kölcsönös, mert ahogy a random hatások varianciáját a fix hatásokra kontrollálva becsüljük, úgy a fix hatások becslése is a random hatások figyelembevételével zajlik. A különbség a modellált jellemzőben van. Míg a fix hatások az átlagstruktúrát (pl. csoportátlagokat és azok eltéréseit), addig a random hatások a variancia-kovariancia struktúrát (az adatok csoportosulásából eredő változatosságot és összefüggéseket) írják le.

A második félreértés a G-, R- és V-mátrix szerepével kapcsolatos. Ezek a mátrixok nem a valós adatokat tartalmazzák, hanem azt az elméleti struktúrát, amelyet a random hatásokhoz, a maradékhibához és az ezek összegéből kialakuló fix változókra kontrollált varianciarész összefüggés-mintázatához *feltételezünk*. A modell majd ezt a feltételezett variancia-kovariancia mátrixot illeszti az adatokhoz, a mátrix elemeinek értékét úgy beállítva, hogy azok a lehető legjobban illeszkedjenek a mért adatok összefüggéseihez.

Az utolsó félreértés a fogalmak félrevezető átfedésével kapcsolatos. A regresszióban minden prediktort fix hatásként kezeltünk, és a reziduális variancia vagy reziduum kifejezést arra a változatosságrészre használtuk, amelyet ezek a prediktorok nem tudtak megmagyarázni. Az LMM-ben is a kutatás szempontjából fontos változókat fix hatásként modellezzük, de azokat külön elemezzük. A V-mátrix az a variancia-kovariancia mátrix, amelyet a fix hatásokra való kontrollálás mellett feltételezünk, és (most jön a félreérthetőség forrása) ez írja le a regresszió reziduumának megfelelő teljes varianciastruktúra feltételezett mintázatát. De az LMM munkája még csak itt kezdődik, amikor a V-mátrix által leírt varianciát és kovarianciát tovább strukturálja a random változókkal (ezt tartalmazza a G-mátrix), és azt, ami ezek után marad, ami jó esetben már tényleg csak a mérési hibát tartalmazó rész (melynek feltételezett struktúrája az R-mátrix), nevezi az LMM maradékhibának, valódi reziduumnak vagy hanyagul egyszerűen csak reziduumnak.

Összegezzünk!

A fix hatásokra fókuszálnak a kísérleti kérdéseink, és itt az átlagstruktúra modellezésén van a hangsúly, tehát itt a csoportátlagok és az azok közötti eltérés a fontos. A változatosság egy részét a fix prediktorok meg tudják magyarázni, egy részét viszont nem. A fix hatásokra való kontrollálás után maradt, azok által meg nem magyarázott változatosság elméleti mintázatát tartalmazza a V-mátrix, amely tehát azt írja le, hogy az egyéni megfigyelések feltevésünk szerint hogyan ingadoznak és függenek össze egymással az átlagos trendek körül.

Ennek az összefüggésrendszernek a leírásában van szerepe a random hatásoknak, amelyekhez a G-mátrix kötődik. Itt az elemzés fókuszában már nem a csoportátlagok vannak, csak azt szeretnénk felmérni, hogy a random hatások mekkora változatosságot hoznak létre az adatokban. Ez kevesebb szabadságfokot igényel, ezáltal növelve a vizsgálat statisztikai erejét.

Végül annak a változatosságnak a feltételezett mintázatát, amelyet se a fix hatások, se a random hatások nem tudtak megmagyarázni, az R-mátrix írja le, és azt feltételezzük, hogy ez a maradék variancia már tényleg csak mérési hiba és zaj.

1.5. Döntés a fix és random hatásokról

A modell felépítése során nekünk kell eldöntenünk, hogy az egyes prediktorokat fix vagy random hatásként kezeljük. A döntést a kutatásunk fókusza és a változók tulajdonságai alapján kell meghozni. Azokat a változókat, melyeknek minden, a kutatás szempontjából releváns szintjét lemértünk, és melyek értékeit egymással össze akarjuk hasonlítani, fix hatásként tesszük a modellbe. Azokat a változókat, amelyek olyan csoportokat jelölnek, ahol csak néhányat mértünk le a lehetséges csoportok populációjából, és amely változóknál a csoportok közötti változatosság érdekel minket, vagy talán még az sem, hanem csak a csoportosulásból származó függőségekre akarunk kontrollálni, random változóként kell a modellbe helyezni. Folytonosként kezelt változó elsődlegesen csak fix hatásként kerülhet a modellbe. Végül ha egy csoportosító tényezőnek az adatbázisban ötnél kevesebb szintje szerepel – akár azért, mert a valóságban is csak ennyi létezik, akár azért, mert a lehetséges sok közül mi csak ennyit

mértünk le –, azt is mindenképp fix hatásként kell bevonni a modellbe,[3] így a kevés szint miatt bizonytalanná váló variancia helyett a csoportok közötti konkrét különbséget becsüljük. A gyakorlati tapasztalat az (mely csupán empirikus ajánlás), hogy tíz-húsz szint alatt is érdemes a fix hatásként való modellezést választani. Ha mégis random hatásként tartanánk meg, akkor minden esetben ajánlott a modell illeszkedésének és a varianciakomponensek becslésének stabilitását ellenőrizni (például AIC-értékkel). Ha a modell instabilnak bizonyul, fontolóra kell venni a kevés szinttel rendelkező random hatás elhagyását vagy fix hatásként való kezelését.

A Napszak (reggel kontra délután) egyértelműen fix hatás, mert erre a változóra vonatkozik a fő kutatási kérdésünk, és a két konkrét csoportot akarjuk összehasonlítani.

A Hallgató ID egyértelműen random hatás, hiszen nem az egyes hallgatókat akarjuk összehasonlítani egymással.

A Kurzus ID (a gyakorlati csoportok egytől tízig) tipikusan random hatás. Nem az első és ötödik gyakorlati csoport különbsége érdekel minket, hanem az, hogy általában a csoportok eltérései (ami mögött állhat a gyakorlatvezető, a csoportdinamika stb.) mekkora varianciát visznek a rendszerbe, másként megfogalmazva, mennyire teszik hasonlóvá az egy csoportba tartozó hallgatók teljesítményét.

Az Intézet ID (Pedagógia és Pszichológia) az elemzés fókusza szempontjából random hatás lenne, azonban egy random hatás használata, ha annak csak két szintje van, komoly statisztikai problémákat vet fel, mivel a modell nem képes megbízhatóan megbecsülni a csoportok közötti valódi varianciát ilyen kevés adatból, és ez instabil eredményekhez vezethet. Ezért az Intézet ID-t vagy fix hatásként érdemes kezelni, vagy érdemes elhagyni a modellből.

A pontozó tanár ID-je szintén random hatás.

A ZH-k sorszáma érdekesebb kérdés, mert a modellben két szerepet is betölthet. Kezelhetjük fix hatásként, ha elsősorban a ZH-k különbségeire, vagyis a tanulási görbére vagyunk kíváncsiak. Példa az ilyen kutatási kérdésre, ha vizsgálni szeretnénk, hogy volt-e kifejezetten rosszul sikerült ZH (ZH-sorszám főhatásának vizsgálata és post hoc tesztekkel való követése), vagy ha az a hipotézisünk, hogy a félév során egyre jobbá válik a tanulók teljesítménye (ZH-sorszám főhatásának trendként való elemzése), vagy ha azt vizsgálnánk, hogy nő-e a reggeli és délutáni csoportok közötti különbség a félév során

[3] A minimálisan szükséges csoportszám pontos értékéről eltérő vélemények vannak a szakirodalomban. Van, aki erős elméleti indokoltság mellett akár három csoport esetén is megengedné a változó random hatásként való kezelését óvatos értelmezés mellett, de van, aki szerint a minimálisan szükséges csoportszám akár tíznél van.

(Napszak és ZH-sorszám interakciója). De dönthetünk úgy is, hogy a ZH-sorszámra mérési elemként (item) tekintünk, és a random hatás struktúráján keresztül kezeljük. Ebben az esetben nem érdekel minket a konkrét ZH-k különbsége, de szeretnénk figyelembe venni, hogy a ZH-k nehézségében van változatosság. Döntésünkkor azt a gyakorlati tanácsot is érdemes figyelembe venni, hogy a kevés szinttel rendelkező változókat inkább fix hatásnak érdemes tekinteni, míg a sok (több mint 5-6) szinttel rendelkezőket inkább random hatásként kezelni. Tehát ha csak három ZH lett volna a félév során, azt inkább fix hatásként helyeznénk a modellbe, de ha tizenegy volt, és a ZH-k összehasonlítása nincs a kutatás fókuszában, akkor észszerűbb a random hatásként történő kezelése.

Hasonló okokból a nem is fix hatásként kerül a modellbe. Az életkor és a felkészülés mértéke folytonos kovariánsok, így azok is fix hatásként kerülnek a modellbe.

1.6. A fix hatások modellezése

A fix hatások becslése az LMM-ben sok szempontból hasonlít a hagyományos regressziós és ANOVA-eljárásokhoz, bár a háttérben működő mechanizmusok eltérőek. A felszínen a modell a fix hatások elemzésekor a kimeneti változó (a Teljesítmény) és a kutató által kijelölt fő prediktorok (például a Napszak, Nem, Kor) közötti kapcsolatot vizsgálja, jellemzően regressziós együtthatók (β) formájában. Ez lehetővé teszi a főhatások és interakciók tesztelését, valamint post hoc összehasonlítások elvégzését is, ahogy azt a hagyományos modellekben megszoktuk.

A lényegi eltérés az, hogy az LMM-nek a fix hatások becslése során nem kell foglalkoznia az adatokban rejlő függőségi struktúrák elemzésével (például ugyanazon hallgató többszöri mérése vagy kurzusok közötti csoportosulás), mert azokat a random hatásokon keresztül kezeli a modell. Ez szemléletesen abban is látható, ahogy a hosszú adatformátum miatt nincs különbség a független minták és az ismételt mérések között, a ZH-sorszám (ha úgy döntünk, hogy fix hatásként emeljük a modellbe) ugyanolyan prediktor lesz, mint a Napszak. Ez azt jelenti, hogy nem a fix hatások elemzését terheli az, hogy a hallgatók több ZH-t is írtak, vagy hogy kurzusokba vannak csoportosítva, mert mindezt a random hatások modellezik, például az ismételt mérés jelölése a Hallgatói ID feladata, a kurzusokba tömörülés pedig a Kurzus ID-é.

Fontos azonban hangsúlyozni, hogy a fix hatások becslése nem független a random hatásoktól. Az LMM egyszerre, egyetlen integrált folyamatban becsüli

a modell összes komponensét. A Maximum Likelihood (ML) vagy Restricted Maximum Likelihood (REML) eljárások a random hatások által leírt varianciakovariancia struktúrát (a G- és R-mátrixokat) is figyelembe veszik, amikor a fix hatások együtthatóit és standard hibáit számolják. Ennek eredményeként a fix hatások becslései már korrigálva vannak a függőségekre. Például a Napszak hatásának becslése már arra a kérdésre válaszol, hogy mekkora a különbség a reggeli és délutáni csoportok között, miután statisztikailag kontrolláltuk a kurzusok és a hallgatók közötti varianciát.

A fix hatásokra vonatkozó statisztikai következtetéseket jellemzően t- vagy F-tesztekkel és a hozzájuk tartozó p-értékekkel, illetve konfidenciaintervallumokkal alapozzuk meg. A kategorikus változók szintjei közötti specifikus különbségek post hoc tesztekkel vagy kontrasztokkal vizsgálhatók, hasonlóan az ANOVA-ban megszokotthoz. A modell továbbá lehetővé teszi az interakciók (például Napszak × Nem) tesztelését is, amelyek azt vizsgálják, hogy egy prediktor hatása függ-e egy másik prediktor szintjeitől.

1.7. A random hatások modellezése

A random hatások modellezésének logikáját a legjobban úgy érthetjük meg, ha nem a kész, komplex modellt próbáljuk azonnal megemészteni, hanem azt lépésről lépésre, nulláról építjük fel. Induljunk ki a lehető legegyszerűbb, de emellett lehető legrosszabb modellből, hogy pontosan lássuk, milyen problémákat kell megoldanunk, és az egyes új komponensek hogyan teszik egyre korrektebbé és valósághűbbé az elemzésünket. A modellt fokozatosan tesszük összetetté. A fejezet során megmaradunk a reggeli és délutáni csoportok teljesítményének vizsgálata mellett, de ahol egy egyszerűbb modell szemléltetéséhez az szükséges, a korábbi alfejezetben bemutatott kísérleti elrendezést egyszerűbbre módosítjuk majd.

A különböző vizsgálati elrendezésekben lépésről lépésre találkozunk a különböző kísérleti elrendezésekhez tartozó mátrixok alakjainak nevével, megtanuljuk, hogy néz ki egy Scaled Identity, egy Variance Components, egy Block-Diagonal vagy egy AR(1) mátrix. Azt is áttekintjük, hogy melyik elrendezés milyen függőségeket feltételez az adatokban, és milyen kísérleti elrendezésnél melyiket érdemes választani. Érdemes erre figyelmet fordítani, mert az LMM beállításairól való kommunikáció során a kutatók ezen kulcsszavak segítségével jellemzik a modelljüket. Fontos azt is értenünk, hogy ezek a kulcsszavak a variancia-kovariancia mátrixok alakjának általános leírására szolgálnak, így lehet például Scaled Identity alakja az R-mátrixnak, ahol a maradékhibák

függetlenségét feltételezzük, de akár a G-mátrixnak is, ha egy random hatás teljes függetlenségét szeretnénk leírni.

Minden modellhez megadjuk annak szintaxisát is. A *modell szintaxisa* az a speciális, képletszerű jelölésrendszer, amellyel az elméleti modellünket egyértelmű és végrehajtható paranccsá alakítjuk a statisztikai szoftver számára. Bár a szintaxis tényleges megvalósulása eltér a különböző szoftverekben, felépítésének logikája alapvetően mindenhol azonos. Ebben a tankönyvben az LMM-et futtató leggyakoribb elemzési környezetek nyelvezetét választottuk példának, így az R program lme4 és (szükség esetén) nlme funkciójának, illetve az SPSS MIXED funkciójának szintaxisát adjuk meg. Kiemelnénk, hogy a szintaxisból a könyv csak a modell felépítésére vonatkozó részleteket közli, míg a teljes szintaxisok további, modellbeállításokra vonatkozó részeket is tartalmaznak.

Nulladik szint, mi történik, ha a random hatásokat figyelmen kívül hagyom?

Az abszolút nulladik lépés, ha a Teljesítmény kimeneti változót megpróbáljuk megjósolni a minket érdeklő egyetlen fix hatással, a Napszakkal, és semmi egyéb tényezőt nem veszünk figyelembe, azaz nem helyezünk a modellbe random hatásokat. A modell szintaxisa ekkor:

26. doboz: Nulladik szintű (random hatások nélküli) alapmodell

A kategorikus Napszak prediktor változóval:
lme4:
Teljesitmeny ~ Napszak

SPSS:
```
MIXED Teljesitmeny BY Napszak
 /FIXED=Napszak | SSTYPE(3).
```

A kategorikus Napszak és a folytonos Felkészülés prediktor változókkal:
lme4:
Teljesitmeny ~ Napszak + Felkeszules

SPSS:
```
MIXED Teljesitmeny BY Napszak WITH Felkeszules
 /FIXED=Napszak Felkeszules | SSTYPE(3).
```

Az lme4 kifejezését angolul úgy fejezzük ki, hogy a Teljesítmény „is regressed on" Napszak, magyarul pedig azt mondjuk, hogy a Teljesítményt a Napszakkal becsüljük. Ez a modell egy egyszerű lineáris regressziónak (vagy egy független mintás t-próbának/ANOVA-nak) felel meg. Azt a kérdést teszi fel, hogy vajon a Napszak magyarázza-e a Teljesítményt. Kiszámolja a reggeli és délutáni csoportok átlagát, és megnézi, hogy a különbség szignifikáns-e a kettő között. Ebben a modellben nincs random hatás, a variancia két részből áll: amit a Napszak megmagyaráz, és minden más, ami egyetlen, differenciálatlan reziduális hibába kerül.

Az SPSS MIXED funkciója a BY szót használja arra, hogy jelölje, a Teljesítményt a kategorikus Napszakkal magyarázzuk, míg a folytonos prediktorok, azaz a kovariánsok (például a Felkészülés) a WITH szó után kerülnek felsorolásra. A /FIXED beállítás jelöli, hogy a Napszak fix hatás (a Type III korrekciót részletesen a beágyazott ANOVA fejezetben tárgyaltunk).

Miért gyenge, sőt hibás ez a modell? Azért, mert a valós adatgyűjtési helyzet legfontosabb jellemzőit hagyja figyelmen kívül. Nem veszi figyelembe az ismételt méréseket, vagyis a modell nem tudja, hogy egy hallgatótól tizenegy adatpont is származik. Nem veszi figyelembe a hallgatók közötti különbségeket, így nem kezeli azt a tényt, hogy egyes hallgatók eleve jobbak, mint mások. Hasonlóképpen azt sem veszi figyelembe, hogy vannak szigorúbb és nagyvonalúbb pontozó tanárok. Nem veszi figyelembe a mérések korrelációját, így a modell nem számol azzal, hogy aki jobb az első ZH-n, az valószínűleg a másodikon is jobb lesz. Nem veszi figyelembe a kurzusok hatását, vagyis azt, hogy az egy kurzusba járó hallgatók teljesítménye valószínűleg jobban hasonlít egymásra.

Mivel ebben a modellben nincsenek random hatások, ezért nincs G-mátrix, és az összes variancia, amelyet strukturálni kellene (a hallgatókból, tanárokból, kurzusokból, időből fakadó rendezettség), kezeletlenül egy egységes reziduumba kerül, és nem válik szét a random hatásokhoz köthető változatosságra és a valódi reziduális hibára.

Mindezekből együttesen adódik a legsúlyosabb hiba, az, hogy a modell nem tudja, hogy tíz kurzusunk van, kurzusonként 14-16 hallgatóval és hallgatónként tizenegy méréssel, hanem azt hiszi, hogy 10 * 15 * 11 = 1650 teljesen független személytől származó adatunk van. Ha ezek a mérések valóban függetlenek lennének egymástól (mindegyik külön személytől származna, akik egyáltalán nincsenek egymásra hatással), jóval nagyobb változatosság lenne közöttük, de a valóságban ezt a változatosságot lecsökkentik a függőségek (klaszteresődések, ismételt mérések). Ez a pszeudoreplikáció tipikus esete, ami drasztikusan megnöveli az elsőfajú hiba valószínűségét, mert a modell a teljes függetlenséget feltételezi, de egy függőségek által lecsökkentett változatossággal találkozik,

vagy fordítva is megfogalmazhatjuk: egy mesterségesen felduzzasztott mintaméretteI dolgozik.

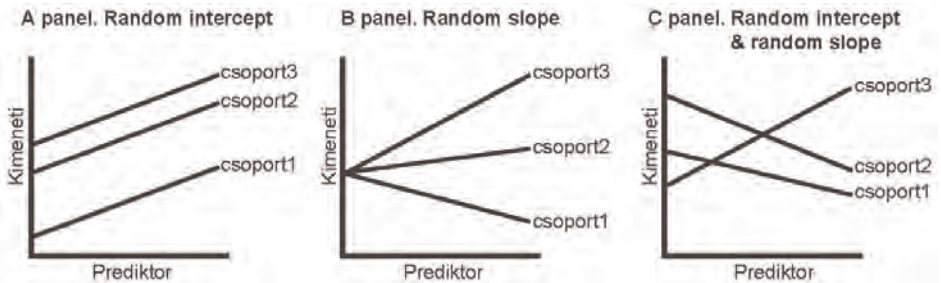
A random intercept és a random slope

Az LMM-ben a random hatások bevezetése lesz a megoldás erre a problémára, amelyek a függőségi struktúrát a modell varianciakomponensein keresztül, és nem a prediktorok szintjén kezelik. Ezen hatások két alapvető típusa a random intercept és a random slope.

A random intercept (magyarul véletlen tengelymetszet lenne, de ez az elnevezés nem elterjedt) modellje az a kiindulópont, ahol a modell elismeri a csoportok közötti alapvető különbségeket. Ez a modell megengedi, hogy minden egyes csoportnak saját, egyéni alapvonala legyen, amely az általános átlagtól véletlenszerűen tér el. Ugyanakkor azt feltételezi, hogy bármely prediktor hatása minden csoportban egyformán érvényesül. Ahogy a 28. ábra A paneljén látható, ez a modell párhuzamos egyeneseket eredményez, tehát a regressziós egyenesek más-más magasságból indulnak, de a meredekségük azonos.

A tiszta random slope (véletlen meredekség) modellje egy ritkán használt, elméleti lehetőség, amely az előzőnek a fordítottját feltételezi. Itt a modell minden csoportot azonos kiindulási ponttal kezel, de megengedi, hogy a prediktor hatásának mértéke, a meredekség csoportonként véletlenszerűen változzon. A 28. ábra B panelje ezt a helyzetet szemlélteti, ahol az összes egyenes egyetlen, közös pontból indul ki, de utána különböző meredekséggel, legyezőszerűen ágazik szét.

A legrugalmasabb és a valóságot leggyakrabban leíró modell a véletlen tengelymetszetet és meredekséget is tartalmazó random intercept and random slope modell. Ez az eljárás ismeri el a csoportok közötti teljes különbséget azáltal, hogy megengedi, hogy a csoportok eltérő szintről induljanak, és azt is, hogy a regressziós egyenesek meredeksége is eltérő legyen. A 28. ábra C paneljén látható kuszaság tökéletesen mutatja ezt a komplexitást, ahol minden egyenesnek saját, egyedi kiindulópontja és meredeksége van, létrehozva az adott csoportra jellemző egyéni fejlődési pályákat.



28. ábra: Grafikon random intercepttel és random slope-pal

1.8. Első szint, a random intercept

Az LMM a fix hatások által meg nem magyarázott varianciát (a hagyományos regressziós modellek reziduumát) tovább strukturálja. Ezt a strukturálást a random hatásokon keresztül valósítjuk meg, melyek a függőségi viszonyokat modellezik. A modellépítés során logikus, hogy a legegyszerűbb függőségi struktúrától haladjunk a komplexebbek felé, így az olvasó számára is követhetővé válik a modellépítés logikája.

A függőség legegyszerűbb és leggyakoribb formája a sima csoportosulás vagy klasztereződés, melyben a mérések nem függetlenek, mert egy közös csoporthoz (például egy személyhez, egy iskolához, egy kurzushoz) tartoznak. Ennek a hierarchikus függőségnek a modellezésére szolgáló alapvető eszköz a random intercept, amellyel a következő alfejezetben részletesen megismerkedünk.

Első példa: egy független mintás hatás random intercepttel

A követhetőség kedvéért egyszerűsítsünk is rögtön a kísérleti elrendezésen, és tegyük fel, hogy minden hallgató csak egy ZH-t ír, viszont tartsuk meg azt, hogy a 150 hallgató tíz, egyenként 14-16 fős gyakorlati csoportba tömörül, melyből öt reggel, öt délután valósul meg. A vizsgálatunk fő kérdése, hogy eltére a reggeli és délutáni kurzusokra járó hallgatók teljesítménye, de szeretnénk azt is figyelembe venni, hogy egy-egy csoporton belül a hallgatók hasonlóbbak, klasztereződnek.

Ez a fajta hierarchikus adatstruktúra, ahol az egyének csoportokba rendeződnek, rendkívül gyakori a kutatási gyakorlatban, és az LMM jelenti a legkorszerűbb elemzési módszert a kezelésére. Gondoljunk csak a gyermekekkel

végzett kutatásokra, ahol az osztályokba vagy iskolákba ágyazott adatok elemzése alapvető kihívás, vagy a klinikai vizsgálatokra, ahol a betegek különböző kórházakban vagy rendelőkben kapják a kezelést.

A modell fix hatásként fogja tartalmazni a Napszakot, és tegyük be a Kurzus ID-t random hatásként, hogy a modell felismerje a csoportok közötti különbségeket. Ezzel a módosítással a modell már képes figyelembe venni, hogy a csoportok lehetnek eltérő mértékben ügyesek, azaz minden csoportnak van egy rá jellemző, átlagos teljesítményszintje. Ez azt is feltételezi, hogy az egy csoporton belüli hallgatók teljesítménye hasonlítani fog egymásra.

Fontos elméleti lépés ez, ezért fordítsunk időt annak megértésére, hogy a csoportok közötti különbségek megengedése miért jelenti a hallgatók csoportonkénti összefüggőségét! Konceptuálisan azt akarjuk, hogy a modell vegye figyelembe, hogy a hallgatói teljesítmények változatosságának egy része abból fakad, hogy melyik csoportba járnak a hallgatók, mert az ügyesebb csoportvezetőt kapó vagy jobb csoportdinamikát kialakító csoportokban mindenki ZH-ja várhatóan egyöntetűen jobb lesz, egy kevésbé ügyes csoportban pedig mindenki ZH-ja várhatóan egyöntetűen rosszabb. Ezt az egyöntetűséget méri a hallgatók csoportokon belüli összefüggése. A modell beállításánál ezt úgy érjük el, hogy a Kurzus ID random hatásnál az alapvonalbeli (intercept) különbségeket engedélyezzük, azaz azt, hogy a csoportok átlaga különbözhessen. A szakzsargonban ez úgy hangzik, hogy a Kurzus ID-nél az intercept szabadon lett engedve. Ekkor a modell szintaxisa:

27. doboz: Random intercept modell felépítése

lme4:

```
Teljesitmeny ~ Napszak + (1 | KurzusID)
```

SPSS:

```
MIXED Teljesitmeny BY Napszak
  /FIXED=Napszak | SSTYPE(3)
  /RANDOM= INTERCEPT | SUBJECT(KurzusID) COVTYPE(VC).
```

Az lme4 teljes képletét angolul úgy mondjuk, hogy a Teljesítmény „is regressed on" Napszak, „with a random intercept for" Kurzus ID. Az új kifejezésrész a random hatást és annak beállításait jelöli. Bontsuk szét a jelölést! A zárójel jelöli, hogy egy random hatásról van szó. A zárójelen belül az 1 konvencionálisan az interceptet, vagyis az alapvonali, átlagos szintet jelöli. A függőleges vonal azt jelenti: „…csoportosítva a következő szerint" vagy „…szintjein belül". Végül

a Kurzus ID az a csoportosító változó, amelynek a szintjein (azaz az egyes gyakorlati csoportokon) belül a random hatást becsülni szeretnénk. A kifejezés együtt tehát azt jelenti: „Becsüljünk egy random interceptet (egyéni alapvonalat) minden egyes gyakorlati csoporthoz!"

Az SPSS-ben a Teljesítményt a Napszak és Kurzus ID változóval magyarázzuk, ahol a napszak fix hatás, a Kurzus ID random hatás. Észrevehetjük, hogy a BY szó után csak a fix hatásokat kell felsorolni, míg a random hatások később a SUBJECT kifejezéssel vannak jelölve. Az INTERCEPT kifejezés az lme4 „1|"-jelölésével azonos jelentésű, tehát azt jelöli, hogy a Kurzus ID-nél a random intercept lett megengedve. A COVTYPE(VC) kifejezés a random hatáshoz tartozó G-mátrix alakjára vonatkozik, melyről később ejtünk szót.

A modellben most már minden csoporthoz eltérő alapteljesítményt becsülünk az adott csoportba járó 14-16 hallgató teljesítménye alapján, és ezeket a csoporthoz kötődő átlagokat az úgynevezett u-vektor tartalmazza. Röviden találkoztunk már az u-vektorral az LMM általános képletének felírásakor. Részletesebben most sem foglalkozunk vele, mert a Kurzus ID egy random hatás, azaz minket nem a csoportok átlagai érdekelnek (nem akarjuk a kettes csoportot az ötöshöz hasonlítani), csak a csoportok közötti változatosság, azaz a csoportok átlagainak varianciája.

A variancia általános statisztikai jele σ^2 (szigma-négyzet), és hagyományosan ezt használjuk a random változók csoportjai közötti variancia jelölésére is, ennek megfelelően a kurzusok közötti eltérésre σ^2_K-ként (K mint kurzus) tudunk hivatkozni.

Ne felejtsük el, hogy a random hatások modellezése mindig a fix hatásokra való kontrollálás mellett szereplő adatokon történik meg, így az u-vektor is a fix hatás csoportátlagaihoz viszonyítja az egyes csoportok teljesítményét. Például ha a reggeli órákra járók teljesítményének átlaga 45 pont, és az egyik reggeli kurzus átlaga pedig 50 pont, akkor az u-vektorban ennél a csoportnál +5 pont fog szerepelni. A σ^2_K is úgy számítja a kurzusok közötti szóródást, hogy az a fix hatásra kontrollálva történik, tehát azt adja meg, hogy a reggel/délután különbözőségén túl milyen további eltérések vannak a kurzusok között. A további gondolatmenetben, ha a fix hatásra való kontrolláltságot nem is írjuk ki mindenhol, az olvasó tartsa azt mindig észben.

Mivel a Kurzus ID egy random hatás, a modellnek megjelenik a G-mátrixa. Emlékezzünk: a G-mátrix a random hatások varianciához való feltételezett hozzájárulását tartalmazza. A G-mátrixot kétféleképp vázolhatjuk fel. Megjeleníthetünk paraméterszinten (ez a gyakoribb, de kevésbé didaktikus), vagy felírhatjuk teljesen kibontva, a random hatás csoportjainak szintjén is.

A paraméterszintű felrajzolás (1. mátrix) során azt jelöljük, hogy a Kurzus ID azzal járul hozzá a megmagyarázatlan variancia strukturálásához, hogy figyelembe veszi a kurzusok átlagos teljesítményének szóródását. Ekkor G-mátrix egy egyszerű 1x1-es mátrix, gyakorlatilag egy skalár (egyetlen érték), amely a kurzusokhoz tartozó interceptek varianciáját, a σ^2_K-t tartalmazza.

$$G = \overset{\text{intercept}}{\underset{\text{intercept}}{\left[\sigma^2_K\right]}}$$

1. mátrix: G-mátrix paraméterszintű felrajzolása

A részletes, adatszintű kibontásban (2. mátrix) a G-mátrix egy *variancia-kovariancia mátrix*, ahol minden kurzus külön sorba és oszlopba kerül, azaz ez egy nxn-es mátrix, ahol n a kurzusok száma. A mi példánkban tíz kurzusról van szó, ezért a G-mátrix mérete 10x10. A mátrix alakja *diagonális*, azaz csak az átlóban tartalmaz értékeket, az átlón kívül mindenhol nulla szerepel benne. Az átlón kívüli terület, ahol egy kurzus sora egy másik kurzus oszlopával találkozik, a két kurzus kovarianciáját jelöli, azaz a két kurzus mérései között feltételezett összefüggést. A most vizsgált kísérleti elrendezésben az átlón kívül minden nulla, aminek az a jelentése, hogy a modell a kurzusokat egymástól függetlennek feltételezi. Mi persze tudjuk, hogy a valóságban a kurzusok nem teljesen függetlenek egymástól, de a napszak miatti függőséget a Napszak magyarázza, a Pszichológia és Pedagógia Intézeteken belüli hasonlóságot pedig nem a Kurzus random hatásának, hanem az Intézet random hatásának kell majd magyaráznia. Az Intézet még nincs a modellbe emelve, így a két intézet által létrejövő hasonlóságot egyelőre csak mérési hibaként, zajként látjuk majd. Ezeken felül azonban a kurzusok elméleti szempontból már tényleg függetlennek tekintendők, hiszen különböző személyek alkotják a csoportokat, és a csoportok nincsenek egymásra hatással.

Az átlóban, ott, ahol egy kurzus sora saját maga oszlopával találkozik, a kurzus önmagával való kovarianciájának becslése szerepel. Bármi önmagával vett kovarianciáját varianciának nevezzük, azaz a mátrix átlójában σ^2_K áll, mely az adott kurzus hallgatói között feltételezett hasonlóság.

A figyelmes olvasónak feltűnhet, hogy a σ^2_K-t kétféleképpen definiáltuk. A paraméterszintű felírásnál azt mondtuk, hogy a σ^2_K a kurzusok átlagos teljesítményének szóródása, azaz a gyakorlati csoportok egymástól való eltérése, ezzel szemben az adatszintű felírásban azt mondtuk, hogy a σ^2_K a csoportokon belül a hallgatók hasonlóságát magyarázza. Hogyan lehet mindkét állítás igaz?

A megoldás abban rejlik, ha visszagondolunk, konceptuálisan mit tanultunk a függőségi viszonyokról. A csoportok közötti eltérések (vannak ügyesebb és kevésbé ügyes gyakorlati csoportok) az egyazon csoporton belüli hallgatók teljesítményének hasonlóságát, azaz azok összefüggését jelentik (egy ügyesebb csoportban vélhetően mindenki ZH-ja jobb lesz).

Az SPSS-ben ezt a diagonális G-mátrixot a random hatás után megadott COVTYPE(VC) beállítással választhatjuk ki. A COVTYPE a G kovarianciamátrix alakjának megadására használatos, és az SPSS-ben kifejezetten sok beállítás közül lehet választani. A VC a *Variance Components* elnevezés rövidítése, mely a random hatások G-mátrixának alapértelmezett beállítása. A VC ebben az egy random változót tartalmazó modellben a fentebb bemutatott diagonális elrendezést jelöli, de teljesebb formáját majd több random változó modellbe emelését követően látjuk csak.[4] Az lme4 automatikusan feltételezi a VC alakot, így ebben a szintaxisban ezt nem kell külön megadni.

Másik kérdésként az merülhet fel, hogy miért ugyanazt a σ^2_K-t használjuk minden kurzuson belüli összefüggés mérésére. A válasz a modellünk beállításaiban rejlik, ahol még csak a Kurzus ID random hatáshoz tartozó interceptet engedtük szabadon, azaz megengedtük, hogy a csoportok átlagteljesítménye eltérjen, ezáltal az egy csoportba járó hallgatók mérései összefüggjenek egymással, de azt még nem engedtük meg, hogy a mérések közötti összefüggés mértéke eltérő legyen a különböző csoportokban.

$$G = \begin{array}{c} \\ \text{Kurzus1} \\ \text{Kurzus2} \\ \text{Kurzus3} \\ \vdots \\ \text{Kurzus9} \\ \text{Kurzus10} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{Kurzus1} & \text{Kurzus2} & \text{Kurzus3} & \ldots & \text{Kurzus9} & \text{Kurzus10} \end{array} \\ \begin{bmatrix} \sigma^2_K & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma^2_K & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sigma^2_K & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2_K & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sigma^2_K \end{bmatrix} \end{array}$$

2. mátrix: G-mátrix adatszintű felbontása

[4] Ez a pont félrevezető lehet: bár a jelenlegi G-mátrix (VC) és a hamarosan bemutatásra kerülő R-mátrix (ID) alakilag teljesen azonosnak tűnik (mindkettő egyszerűen diagonális), ez kizárólag annak köszönhető, hogy a modellben most még csak egyetlen random hatás (intercept) szerepel. Amint a modellbe több random változó kerül, a VC és az egyszerű diagonális vagy identity struktúrák közötti különbség azonnal láthatóvá válik majd.

Lássuk, hol tartunk! A kimeneti értéket (a mért ZH-pontszámokat) részben megmagyaráztuk a napszakkal (ez a fix hatások vizsgálata), a meg nem magyarázott variancia egy részét pedig megmagyaráztuk a kurzusok alapteljesítménye közötti eltéréssel (ez egy random hatás random intercepttel).

A ténylegesen mért adatok és az így becsült értékek közötti különbséget, azaz a maradék változatosságot, az ε (epszilon) maradékhiba-vektor tartalmazza. Az R-mátrix ezen maradékhibák feltételezett variancia-kovariancia struktúráját adja. A mátrix egy-egy sora és oszlopa egy-egy mérést jelenít meg, azaz 150 hallgató esetén ez egy 150x150-es mátrix. Átlójában a maradékhibák feltételezett varianciája, a σ^2_ε (szigma-négyzet epszilon) szerepel, az átlón kívül pedig nullák. Az olyan diagonális mátrixot, melynek átlójában mindenhol ugyanaz az érték áll, *Scaled Identity* (SI vagy ID) mátrixnak szokás nevezni, és ez az LMM alapértelmezett beállítása az R-mátrixok esetében. Az elnevezésben az identity (azonosság) arra utal, hogy ugyanaz a σ^2_ε jelenik meg mindenhol az átlóban, azaz a modell nem engedi meg, hogy a különböző hibatagok eltérőek legyenek. Az átlón kívüli nulla értékek a maradékhibák függetlenségét feltételezik. Tehát ez az R-mátrix a magyarázhatatlan maradékhibákat egymástól függetlennek és azonos varianciájúnak tekinti. A maradékhibák függetlenségének feltételezése ugyanaz, mint amit bármely más elemzésnél is megtanultunk, hogy a végső, meg nem magyarázott változatosság már csak egyszerű zaj, mérési hiba. Az azonos varianciaértékek feltételezése pedig megegyezik a regresszió jól ismert feltételével, a reziduális hibák homogenitásának feltételével, más néven a szóráshomogenitással, amely feltételezi, hogy a becsléseink mindenhol egyformán pontosak, így a becslés hibája mindenhol egyforma mértékű.

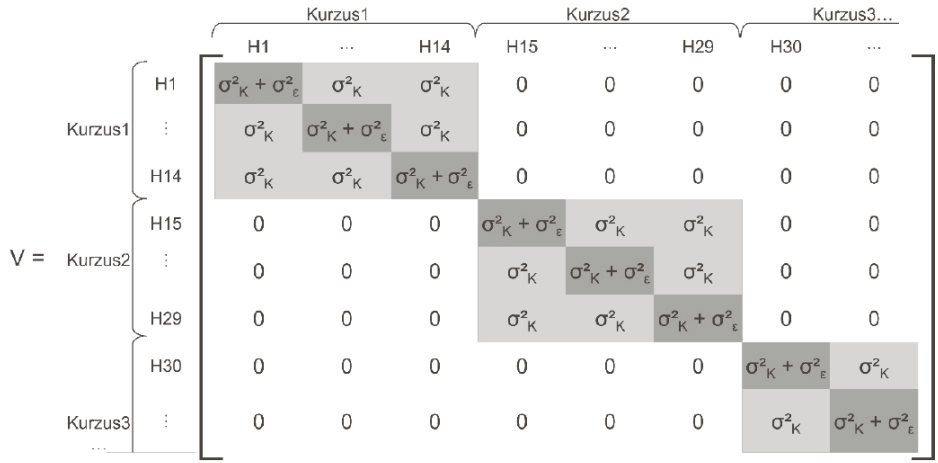
A Scaled Identity alakú R-mátrixra példát a 3. mátrixban láthatunk. (Tisztázásként meg kell említenünk, hogy a 3. és 4. mátrixban a Hallgatók sorszámozása csak a megjelenítést segíti, de a modellnek nem része, mivel a Hallgató ID változót nem tettük a modellbe. Ez azért nem jelent problémát, mert minden hallgatóhoz azonos értékeket feltételez a modell.)

$$R = \begin{array}{c} \\ \text{Hallgató1} \\ \text{Hallgató2} \\ \text{Hallgató3} \\ \vdots \\ \text{Hallgató149} \\ \text{Hallgató150} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{Hallgató1} & \text{Hallgató2} & \text{Hallgató3} & \cdots & \text{Hallgató149} & \text{Hallgató150} \end{array} \\ \begin{bmatrix} \sigma^2_\varepsilon & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma^2_\varepsilon & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sigma^2_\varepsilon & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2_\varepsilon & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sigma^2_\varepsilon \end{bmatrix} \end{array}$$

3. mátrix: Példa az R-mátrixra

Végül a V-mátrix a random hatások és a maradékhibák együttes hatását tartalmazza, amely a kimeneti változó teljes variancia-kovariancia struktúráját írja le a fix hatások kontrollálása után. Fontos megjegyezni, hogy a V-mátrix nem a ténylegesen mért adatok empirikus kovarianciáját írja le, hanem a modell által feltételezett variancia-kovariancia struktúrát adja, amelyet a random hatások és a maradékhibák együttesen határoznak meg, és amely a lehető legjobb közelítése az empirikusan mért mintázatnak. A V-mátrixot a 4. mátrix szemlélteti.

Ebben a példában a V-mátrix alakja jellegzetes, úgynevezett *blokk-diagonális* (Block-Diagonal) formát követ. Ez azt jelenti, hogy a mátrix 150x150-es (150 hallgató miatt), és tíz darab, körülbelül 15x15-ös blokkra osztható, amelyek a tíz gyakorlati csoporthoz tartoznak (14–16 hallgató/csoport, a blokkok pontos mérete a létszám függvénye). Minden blokk a csoporton belüli méréseket írja le, ahol az átlón a teljes variancia ($\sigma^2_K + \sigma^2_\varepsilon$), az átlón kívül pedig a csoporton belüli korreláció (σ^2_K) szerepel, mivel az azonos csoportba tartozó hallgatók mérései korrelálnak a csoport random interceptje miatt. A különböző csoportokhoz tartozó mérések között nulla a korreláció, azaz a modell a csoportokat függetlennek tekinti egymástól, így a blokkokon kívül minden elem nulla, ami a blokk-diagonális szerkezetet adja. Ez a mintázat a független mintás elrendezésből fakad, ahol a gyakorlati csoportok elméletileg függetlenek egymástól a napszak fix hatására való kontrollálás után.



4. mátrix: V-mátrix egy modellhez egy random hatás random interceptjével

Összegezzük, mit tudtunk meg erről a modellelrendezésről! A G-mátrix alakja diagonális, ami azt jelenti, hogy a kurzusok alapteljesítményében a modell megengedi a szóródást, de a kurzusok között nem enged meg összefüggést, azaz függetlennek tekinti őket. Továbbá az átlóban álló értékek azonosak, azaz a kurzusokon belüli szóródás hasonló mértékű minden kurzus esetében, másképpen megfogalmazva: egyforma mértékben hasonló minden kurzus hallgatóinak teljesítménye.

Második példa: elengedhető-e a szóráshomogenitás feltétele?

Az előző random interceptet tartalmazó modell G-mátrixában (2. mátrix) feltűnhetett, hogy az átlóban mindenhol ugyanaz a σ^2_K szerepelt. Azaz megengedtük, hogy az egy kurzuson belüli személyek között hasonlóság legyen, de megkötöttük, hogy ez a hasonlóság egyforma mértékű legyen minden csoportban, vagyis nem lehettek konzisztensebb (erősebb korrelációk a személyek között) és heterogénebb teljesítményt nyújtó (gyengébb korrelációk a személyek között) kurzusok. Ez egyfajta homogenitásfeltételezés. Jogosan merül fel kérdésként, hogy nem lehetne-e az átlóban minden kurzushoz eltérő varianciát rendelni, azaz az egységes σ^2_K helyett lehetne σ^2_{K1}, σ^2_{K2} és így tovább, és ha $\sigma^2_{K1} > \sigma^2_{K2}$, akkor annak az a jelentése, hogy az első kurzusban a hallgatók teljesítménye jobban hasonlít egymásra, mint a másodikban. Az ezt leíró alakot az 5. mátrix szemlélteti.

$$G = \begin{array}{c} \\ \text{Kurzus1} \\ \text{Kurzus2} \\ \text{Kurzus3} \\ \vdots \\ \text{Kurzus9} \\ \text{Kurzus10} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{Kurzus1} & \text{Kurzus2} & \text{Kurzus3} & \cdots & \text{Kurzus9} & \text{Kurzus10} \end{array} \\ \begin{bmatrix} \sigma^2_{K1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma^2_{K2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sigma^2_{K3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2_{K9} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sigma^2_{K10} \end{bmatrix} \end{array}$$

5. mátrix: G-mátrix a szóráshomogenitás elengedése esetén

Az inhomogenitás megengedése technikailag lehetséges az LMM-ben, de a csoportonként eltérő erősségű korrelációk feltételezése mindenféle strukturálás nélkül nagymértékben növelni a modellezés komplexitását, és növeli az elemszámigényt. Kevés szoftver támogatja, és csak a heterogenitás erős elméleti megalapozottsága mellett ajánlott, így a kutatók ritkán szokták választani. A G-mátrix alakja ekkor továbbra is diagonális, de már nem ugyanazokat a varianciaértékeket tartalmazza. Az ilyen alakot szokás *heterogén diagonálisnak* (Heterogeneous Diagonal) vagy egyszerűen *diagonális*nak (Diagonal) nevezni.

Az SPSS-ben a heterogén diagonális elrendezés a COVTYPE(DIAG) beállítással valósítható meg, az lme4-ben nem lehetséges a modellezése, míg az általánosabb nlme funkcióban a weights = varIdent() függvény használatával lehetséges.

28. doboz: Heterogén diagonális struktúra beállítása

SPSS:

```
MIXED Teljesitmeny BY Napszak
 /FIXED=Napszak | SSTYPE(3)
 /RANDOM= INTERCEPT| SUBJECT(KurzusID) COVTYPE(DIAG).
```

Összegezzük, mit tudtunk meg erről a modellelrendezésről! A G-mátrix alakja heterogén diagonális, mely feltételezi, hogy az egy kurzusra járó hallgatók hasonlóbbak legyenek egymáshoz, továbbá megengedi, hogy ez a hasonlóság eltérő legyen a különböző kurzusokon. A kurzusokat egymástól függetlennek tekinti.

Harmadik példa: független mintás, keresztezett vizsgálati elrendezések random intercepttel

A könyv itt szakít azzal az általános tanítási gyakorlattal, mely első szintként a random interceptet mutatja be, majd azonnal áttér a random slope bemutatására, mert ez utóbbi téma megkezdése előtt elidőzünk még a strukturálisan egyszerűbb modelleknél, és megnézzük, mi történik, ha egynél több random hatást emelünk be a modellbe. A döntés mögötti motiváció az, hogy az LMM rendkívül rugalmas keretrendszerében a legnehezebb feladat annak eldöntése, hogy egy aktuális kutatás adatstruktúrájához és kísérleti kérdéseihez melyik beállítások a megfelelőek, és e beállítások milyen következményekkel járnak az elemzésre.

Tegyük fel, hogy a tíz gyakorlati csoport egyszerre írja a ZH-t, és utána összeül több oktató is azok kijavítására. A ZH-kat egy nagy halomba gyűjtjük, és szétosztjuk a javítást végző személyek között, tehát a tanárok véletlenszerűen kapják meg a dolgozatokat. A leírt kísérleti helyzet egy jó példa a keresztezett (crossed) random hatások modelljére. (Bár a valóságban nagy valószínűséggel így van, az elrendezés nem követeli meg, hogy ugyanazok a személyek legyenek a gyakorlatvezetők, mint akik a javítást végzik.)

A modellben a Kurzus ID és a Pontozó tanár ID-ja is random hatásként szerepel. Azért kezeljük őket random hatásként, mert nem az egyes konkrét csoportok vagy tanárok érdekelnek minket, hanem az általuk képviselt populációk (az összes lehetséges kurzus és az összes lehetséges pontozó tanár) varianciájára vagyunk kíváncsiak. Azt szeretnénk megbecsülni, hogy mekkora a teljes varianciából a csoportok közötti tudáskülönbségből, illetve a tanárok javítási szigorának különbségéből fakadó rész.

A két szempont keresztezi egymást, mivel egy tanár különböző csoportokba járó hallgatók dolgozatát is javítja, és egy csoport különböző hallgatóit eltérő tanárok értékelhetik. Nincs hierarchikus beágyazottság (például nincsenek fix kurzus-pontozó párosulások), a kapcsolatrendszer hálózatszerű. Ráadásul ez az elrendezés szinte törvényszerűen kiegyensúlyozatlan (unbalanced). Ez azt jelenti, hogy a kurzus-pontozó párosításokból képzett összes lehetséges cellának csak egy részében lesz tényleges megfigyelés (ZH-eredmény). Előfordulhat, hogy az első kurzus ZH-iból egy sem jut Pontozó1 elé. Befolyásolja-e ez jelentősen

a mátrixok szerkezetét? A konkrét felépítésüket részben igen, de a koncepcionális szerepüket nem. Magyarázatunkban erre is ki fogunk térni.

Az, hogy már két random hatást tartalmaz a modell keresztezett elrendezésben, a szintaxisban a tagok számának bővülésében jelenik meg.

29. doboz: Keresztezett random hatások modellezése

lme4:

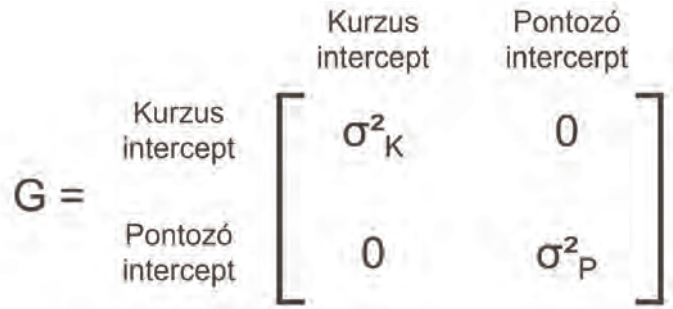
```
Teljesitmeny ~ Napszak + (1 | KurzusID) + (1 | PontozoID).
```

SPSS:

```
MIXED Teljesitmeny BY Napszak
 /FIXED=Napszak | SSTYPE(3)
 /RANDOM=INTERCEPT | SUBJECT(KurzusID) COVTYPE(VC)
 /RANDOM=INTERCEPT | SUBJECT(PontozoID) COVTYPE(VC).
```

Hasonló változás látható a random hatásokat leíró u-vektorban, amely most már a kurzusokhoz tartozó eltérések mellett a pontozó tanárokhoz tartozókat is tartalmazza. Emlékeztetőül, ezek a vektorok nem a kurzusok és tanárok abszolút becsült értékét tárolják, hanem a modell fix hatásai által előre jelzett értékhez viszonyított eltérést. Például ha a modell a reggeli kurzusokhoz 45 pontot jósol, de a reggeli időpontú Kurzus1 hallgatóinak átlaga 50 pont lett, akkor a Kurzus1-hez tartozó érték az u-vektorban +5 lesz. Hasonlóképpen, ha egy tanár az összes általa javított dolgozatot átlagosan 3 ponttal szigorúbban pontozza, mint amit a modell előre jelzett volna, akkor a tanárhoz tartozó random intercept -3 lesz. Ez azonban technikai részlet, mivel a Kurzus ID és Pontozó ID is random hatás, így nem ezekre az egyedi becsült értékekre vagyunk kíváncsiak, csak az általuk megmagyarázott varianciára.

A G-mátrixban annyi változás történik, hogy a Kurzus ID mellett megjelenik a pontozó tanár blokkja is. Konceptuálisan ez a következőképpen jelenik meg (6. mátrix).



6. mátrix: G-mátrix paraméterszintű felbontása két keresztezett random hatással

Az adatszintű felbontásban (7. mátrix) a kurzusok sorai és oszlopai mellett megjelennek a pontozó tanárokhoz tartozó sorok és oszlopok, melyek a kurzus blokkjához hasonló újabb blokkot adnak a G-mátrixhoz. A G-mátrix mérete ezzel (10+5)x(10+5)-re nő. A tíz kurzushoz σ^2_K variancia, az öt pontozó tanárhoz σ^2_P variancia fog tartozni. A kurzusok vagy tanárok blokkján belül az átlón kívül a mátrix nullákat tartalmaz, tehát a kurzusokat egymástól függetlennek feltételezi, és a pontozóknál is ugyanezzel a feltételezéssel él. A kurzusok és pontozók találkozási pontjaiban a cellák értéke szintén nulla, mert a két szempontot egymástól függetlennek feltételezzük azért, mert véletlenszerűen kerül bármely pontozóhoz bármely kurzus hallgatójának ZH-ja. A G-mátrix változása tehát azt jelenti, hogy a modell a teljes meg nem magyarázott varianciából egyre nagyobb részt tud strukturáltan kezelni (a kurzushoz és a pontozó tanárhoz köthető hatásként), így csökkentve a valóban megmagyarázatlan, véletlen hibának tulajdonított variancia arányát, azaz az R-mátrixban feltételezett varianciákat.

Alakját tekintve ez is diagonális mátrix, mely a kurzusokhoz tartozó részen σ^2_K-t, a pontozó tanárok részén σ^2_P-t tartalmaz. Ezt az elrendezést *Variance Components* mátrixnak nevezzük. A Variance Components (VC) kovarianciastruktúra egy szigorúan diagonális G-mátrixot feltételez, ahol 1) minden random hatáshoz külön variancia tartozik, 2) az azonos típusú random hatásokra (például kurzus) azonos varianciát ír elő, 3) és a különböző típusú hatások között nulla kovarianciát feltételez.

7. mátrix: G-mátrix adatszintű felbontása két keresztezett random hatással

Az R-mátrix alakja nem változik, továbbra is Scaled Identity, tehát azonos alakú a 3. mátrixban látottakkal. A különbség csak annyi, hogy az általa leírt σ^2_ε maradékhiba most már remélhetően kisebb, hiszen nemcsak a kurzusba tartozással, hanem a pontozó tanárok közötti eltéréssel is strukturáltunk belőle egy részt.

A V-mátrix továbbra is 150x150-es méretű a százötven hallgatónak megfelelően. Alakja komplex, a keresztezett elrendezés miatt nem tisztán blokk-diagonális, mint az előző példákban. Az átlóban minden hallgatónál szerepel a kurzus által megmagyarázott variancia, a pontozó által megmagyarázott variancia és a maradék variancia ($\sigma^2_K + \sigma^2_P + \sigma^2_\varepsilon$). Az átlón kívül az azonos kurzusra, de eltérő pontozóhoz került hallgatókat csak a kurzus teszi hasonlóvá (σ^2_K), a különböző kurzusra járó, de azonos pontozóhoz került hallgatókat a tanár szigora (σ^2_P), az azonos kurzusra járó és azonos pontozóhoz kerülő hallgatókat pedig mindkettő ($\sigma^2_K + \sigma^2_P$). Végül a különböző kurzusra járó és különböző pontozóhoz kerülő hallgatókat függetlennek tekinti egymástól, a köztük lévő kovarianciát 0-nak feltételezi.

Összegezzük, mit tudtunk meg erről a modellelrendezésről! A G-mátrix alakja Variance Components. Ez a legelterjedtebb több független mintás random hatás esetében. Ez a modell feltételezi, hogy az egyazon kurzusra járó hallgatók ZH-i hasonlóbbak egymáshoz, és azt is, hogy ez a hasonlóság azonos minden kurzus esetében. Emellett feltételezi, hogy az azonos tanár által pontozott ZH-k is hasonlóbbak lesznek egymáshoz, és ez a hasonlóság is azonos mértékű minden tanár esetében. A kurzusokat függetleneknek tekinti egymástól, és a tanárokat is. Továbbá feltételezi, hogy a Kurzus és Pontozótanár szempontok sem függenek össze egymással.

Negyedik példa: független mintás, beágyazott vizsgálati elrendezés random intercepttel

A független mintás, beágyazott elrendezésre jó példa, amikor a hallgatók a gyakorlati kurzusokba tömörülnek, a kurzusok pedig különböző intézetekhez tartoznak. Például az első öt kurzus a Pszichológia Intézethez tartozik, a második öt pedig a Pedagógia Intézethez. Ekkor a Kurzus ID és az Intézet ID egymásba beágyazott. Fontos azonban megjegyezni, hogy a valós adatelemzési gyakorlatban az Intézetet nem kezelnénk random hatásként, mivel csupán két szintje van, és ennyi csoportból nem lehet megbízhatóan varianciát becsülni, ezért inkább fix hatásként vonnánk be a modellbe, vagy elhagynánk az elemzésből. A tankönyvi példa kedvéért most eltekintünk ettől a módszertani gyengeségtől, hogy a tisztán beágyazott, többszintű hierarchiát bemutathassuk.

Tegyük fel, hogy a százötven hallgató tíz darab, egyenként 14-16 fős gyakorlati csoportba jár. Az LMM-et a beágyazott ANOVA-val ellentétben nem zavarja, hogy nincs kiegyensúlyozottság, a gyakorlati csoportok létszáma lehet eltérő. A gyakorlati csoportokba tartozást a Kurzus ID bevezetésével közöljük, a pszichológia- vagy pedagógiaképzéshez tartozást az Intézet ID-vel. Eddig a kurzusokat a napszak hatásán felül függetlennek feltételeztük, most viszont a kurzusok különbségeiben is feltételezni fogunk egyfajta strukturáltságot, azt, hogy az egy képzéshez tartozó csoportok is hasonlítani fognak egymáshoz. Ez egy többszintű hierarchia, mert a hallgatók a gyakorlati csoportokba tartoznak, és a gyakorlati csoportok pedig intézetekhez. A beágyazott ANOVA-val ellentétben itt nem követelmény, hogy a kutatás fókuszában álló fix hatás alá legyen beágyazva minden random hatás, és a jelen példában ez nem is teljesül, mivel a pszichológia- és pedagógiaképzésnek is vannak reggeli és délutáni csoportjai, tehát az Intézet ID és Napszak keresztezett elrendezésű. A többszintű hierarchia jelölését a különböző szoftverek eltérően kérik.

30. doboz: Beágyazott (hierarchikus) random hatások modellezése

lme4:
Teljesitmeny ~ Napszak + (1|IntezetID/KurzusID)

lme4 alternatív jelölések:
Teljesitmeny ~ Napszak + (1|IntezetID) + (1|KurzusID) – itt a beágyazottságot az adatokból olvassa ki a szoftver
Teljesitmeny ~ Napszak + (1|IntezetID) + (1|IntezetID: KurzusID)

SPSS:
```
MIXED Teljesitmeny BY Napszak
 /FIXED = Napszak | SSTYPE(3)
 /RANDOM = INTERCEPT | SUBJECT(IntezetID) COVTYPE(VC)
 /RANDOM = INTERCEPT | SUBJECT(KurzusID) COVTYPE(VC).
```

Az lme4 szintaxisában a zárójel továbbra is a random hatásokra utal, és az 1| a random interceptet jelöli. Ez a modell is két random hatást tartalmaz, ez az Intézet és a Kurzus ID, de a keresztezett példával ellentétben ezeket nem egymás után, pluszjellel elválasztva soroljuk fel, hanem jelöljük a beágyazottságot is. A / jel a beágyazott random hatásokat jelzi, és azt fejezi ki, hogy az egyik változóba (felső szint) egy másik változó (alsó szint) van beágyazva. Példánkban az intézetek felső szintjébe vannak a gyakorlati kurzusok beágyazva, így a helyes sorrend intézet/kurzus. Az lme4 logikája az, hogy ez a felépítés egyértelműen sugallja a beágyazottságot, és kijelöli a szinteket.

Sok félreértés övezi a beágyazott struktúrák jelölését, mert több megjelenítési mód is használatban van. A fentebb leírt (1|IntezetID/KurzusID) az lme4 alapvető szintaxisa a beágyazott struktúrák jelölésére. A legtöbb szoftver azonban az adatokból is képes kiolvasni a beágyazottságot, és ezért nem igényli annak explicit kifejezését a szintaxisban. Ha az adatokban teljesen egyértelmű a beágyazottság (ami azt is jelenti, hogy az alacsonyabb szintű változók ID-i egyediek a teljes mintában, nem csak a felettes csoporton belül), akkor úgy is helyesen beágyazott modellt futtat, ha a kutató csak felsorolja a random változókat, még akkor is, ha az szigorú értelemben keresztezett elrendezésre utalna. Ennek megfelelően a (1|IntezetID) + (1|KurzusID) is beágyazott modellt futtat az egyértelműen beágyazott adatokon, de ez csak a szoftver rugalmasságának köszönhető, nem pedig a két elrendezés azonosságát jelöli. Egy kevésbé elterjedt, de szintén működő kifejezésmód a beágyazott elrendezések jelölésére a (1|IntezetID) + (1|IntezetID: KurzusID). Ez a kifejezés is ugyanazt a beágyazott struktúrát modellezi, koncepcionálisan azt emeli ki, hogy a varianciát strukturálják az intézetek, és az intézeteken belüli varianciát pedig a kurzusok.

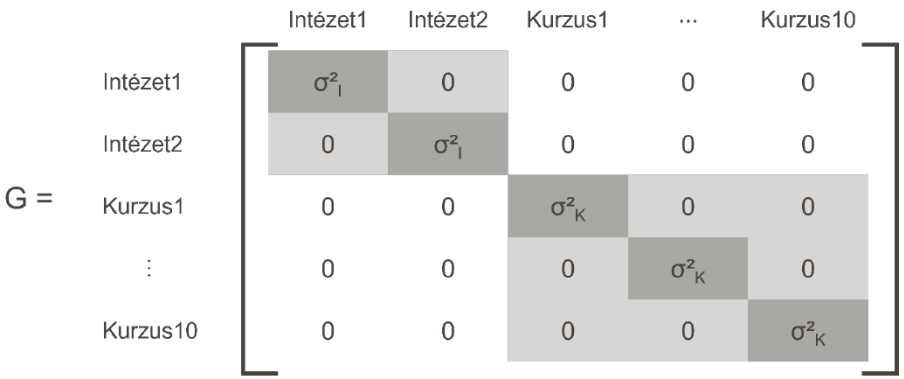
Az SPSS is azon szoftverek közé tartozik, amely ki tudja olvasni a beágyazottságot az adatokból, így a külön random intercepttel megadott Intézet ID és Kurzus ID beágyazott struktúrát fog futtatni, de még egyszer felhívjuk a figyelmet arra, hogy a modell helyes lefutásához az kell, hogy az adatokból tényleg egyértelmű legyen a beágyazott szerkezet.

A beágyazottság az adatstruktúrában jelenik meg, mivel a Pszichológia Intézet alatt csak az első öt, a Pedagógia Intézet alatt csak a második öt kurzusnak van adata. A varianciabontás egy újabb szinttel bővül. A napszak fix hatása

a modellben a teljes populációra vonatkozó átlagos hatást becsüli (például reggel vagy délután), míg az intézet random interceptjei az intézetek átlagos ZH-teljesítményének eltérését modellezik a nagy átlagtól, a kurzus random interceptjei σ^2_K pedig az intézeteken belüli kurzusok átlagos ZH-teljesítményének eltérését modellezik az intézet átlagától. Ez a beágyazottság az u-vektor értékeit befolyásolja. A G-mátrix formailag nem tér el keresztezett és beágyazott struktúrák esetén (ehhez vesd össze a keresztezett 6. és 7. mátrixot a beágyazott 8. és 9. mátrixszal), továbbra is Variance Components alakú, adatszintű megjelenítés esetében méretét tekintve (2+10)x(2+10) méretű. Az első 2x2-es blokk az intézetek hatásának felel meg, és az átlóban az intézetek varianciáját, σ^2_I-t tartalmazza, az átlók kívüli értékek nullák, mivel a napszakra való kontrollálást követően az intézeteket függetleneknek feltételezzük egymástól. A második 10x10-es blokk a kurzusoknak felel meg, ahol az átlóban a σ^2_K szerepel. Az átlón kívül itt is mindenhol nulla szerepel, mivel az intézetek miatti összefüggést az intézet szempont tartalmazza, ettől függetlenül a kurzusokat függetleneknek feltételezzük egymástól. Természetesen azok a cellák is üresek, ahol egy intézet sora egy kurzus sorával találkozik, mert ezeket a szempontokat is függetlennek feltételezzük.

$$G = \begin{array}{c} \\ \text{Intézet intercept} \\ \text{Kurzus intercept} \end{array} \begin{array}{c} \begin{array}{cc} \text{Intézet intercept} & \text{Kurzus intercerpt} \end{array} \\ \begin{bmatrix} \sigma^2_I & 0 \\ 0 & \sigma^2_K \end{bmatrix} \end{array}$$

8. mátrix: G-mátrix paraméterszintű felbontása két beágyazott random hatással



9. mátrix: G-mátrix adatszintű felbontása két beágyazott random hatással

A V-mátrix viszont eltér keresztezett és beágyazott esetben. Továbbra is a 150 hallgatónak megfelelően 150x150 méretű. Az átlóban tartalmazza az intézetek és a kurzusok által magyarázott varianciát, valamint a maradékhibát ($\sigma^2_I + \sigma^2_K + \sigma^2_\varepsilon$). Ahol két olyan hallgató sora és oszlopa találkozik, akik azonos kurzusra járnak, és a beágyazottságnak megfelelően azonos intézethez tartoznak, ott az intézet és kurzus varianciatagja együtt magyarázza a közöttük lévő hasonlóságot ($\sigma^2_I + \sigma^2_K$). Ahol olyan hallgatók sora és oszlopa találkozik, akik azonos intézethez, de eltérő kurzushoz tartoznak, ott az intézet varianciája (σ^2_I) magyarázza a hasonlóságot. A beágyazottság miatt (és itt az eltérés a keresztezett elrendezéstől) olyan eset nem fordul elő, ahol azonos a kurzus, de eltérő az intézet. Minden más esetben a mátrix nullákat tartalmaz.

Összegezve az eddig tanultakat, a klaszterező random változók szintaxisának kulcsszava a *Variance Components* (VC) abban az esetben, ha csak a csoportok átlaga közötti különbséget akarjuk megengedni (random intercept), de a hatások csoportok közötti variálódását (random slope) még nem. A VC a G-mátrix alakjának leírására szolgál, és azt fejezi ki, hogy a csoportok átlaga különböző lehet (ennek a változatosságnak a varianciája van a G-mátrix átlójában), egyébként a csoportokat függetlennek feltételezi egymástól. Több random hatás esetében az átlóban minden változóhoz eltérő variancia tartozhat.

Az R-mátrix alakja *Scaled Identity*, ahol az átlóban a maradékhiba állandó varianciája van, ami a szóráshomogenitás feltételezését jelöli, az átlón kívül pedig mindenhol nulla, ami a hibák függetlenségét feltételezi (azonos a 3. mátrixban látottakkal.)

A V-mátrix alakja egy random változó esetén *blokk-diagonális*, ahol az azonos csoportba tartozó egyedek között feltételezett a kovariancia, az eltérő csoportba

tartozó egyedek között viszont függetlenséget feltételezünk. Több keresztezett random hatás (nevezzük A-nak és B-nek) esetén a V-mátrix alakja *komplex*, az átlóban megjelenik mindkét random hatás varianciája és a maradékhiba is ($\sigma^2_A + \sigma^2_B + \sigma^2_\varepsilon$), azaz egy személy értékét az egyik és másik csoportba tartozás, valamint a mérési bizonytalanság befolyásolja, például egy hallgató ZH-pontszámát befolyásolja az is, hogy melyik kurzusra járt, és az is, hogy ki javította a ZH-t. Az átlón kívül, az A és B szempontjából azonos csoportba tartozó személyeket (azonos kurzus és azonos pontozó) $\sigma^2_A + \sigma^2_B$ köti össze. Az A szempontjából azonos, de B szempontjából eltérő csoporthoz tartozó személyeket (például azonos kurzus, de eltérő pontozó) σ^2_A köti össze, a B szempontjából azonos, de A szempontjából eltérő csoportba tartozó személyeket (azonos pontozó, de eltérő kurzus) pedig a σ^2_B. Az A és B alapján is eltérő csoportokhoz tartozó két személyt (eltérő kurzus és pontozó) pedig függetlennek feltételezi a modell.

A beágyazott elrendezéseknek (ahol a magasabb hatást nevezzük C-nek, az alsóbbat D-nek) ugyan van saját szintaxisa, de a legtöbb szoftver az adatokból is ki tudja a beágyazottságot olvasni, hiszen a beágyazottság az LMM-ben csak annyit jelent, hogy a V-mátrixban csak olyan személyek fognak előfordulni, ahol C és D szerint is azonos a csoport (azonos intézet, azonos kurzus), és a kovarianciát $\sigma^2_C + \sigma^2_D$ adja, vagy C szerint azonos, de D szerint nem az (azonos intézet, eltérő kurzus), és a σ^2_C jelöli a függőséget, de olyan személy nem fog előfordulni, akinél D azonos, de C nem az (azonos kurzus, eltérő intézet).

1.9. Ismételt mérések modellezése

Miután megvizsgáltuk a független mintás adatokat, térjünk át az ismételt méréses elrendezésekre. Induljunk ki egy egyszerű példából, még ne legyen se intézet, se kurzus, csak 150 hallgató, akiknek a fele a reggeli, másik fele a délutáni órára jár. A félév során tizenegy ZH-t ír mindenki. Ez a struktúra leginkább a vegyes ANOVA-ra hasonlít, ahol a Napszak egy független mintás, a ZH-sorszám pedig egy ismételt méréses szempont lett volna. Az LMM megközelítése azonban két ponton is újdonságot hoz.

Az első, technikai eltérés, hogy az LMM továbbra is a hosszú formátumban várja az adatokat. Ez azt jelenti, hogy egyetlen Teljesítmény oszlopunk lesz, és egy hallgatónak mind a tizenegy ZH-ja egymás alá, külön sorokba kerül. Szükségünk lesz egy Hallgató ID változóra, amely azt azonosítja, melyik tizenegy mérés tartozik ugyanahhoz a személyhez, valamint egy ZH-sorszám változóra, amely a méréseket azonosítja és sorrendjüket jelöli.

A második és sokkal fontosabb eltérés az, hogy az LMM-mel lehetséges úgy modellezni a napszak hatását, hogy a ZH-sorszám szempontját akár egyáltalán nem emeljük be a modellbe, csak a Hallgató ID-t, ezzel figyelembe véve, hogy a mérések összefüggésben vannak egymással (kikerülve a pszeudoreplikáció hibáját), de nem pazarolva a szabadságfokot a ZH-k közötti eltérések modellezésére, ha a kutatásunknak ez nem témája. A hagyományos eljárásokkal ez nem lett volna lehetséges. Ott két rossz opció közül kellett volna választanunk: vagy összeátlagoljuk a tizenegy ZH-t, ezzel elveszítve minden időbeli információt, vagy ismételt méréses faktorként, 10 szabadságfokot feláldozva kellett volna megbecsülnünk a ZH-k közötti összes különbséget akkor is, ha az nem érdekel minket. Ezzel szemben az LMM-ben a döntést a kutatási kérdés alapján tudjuk meghozni. Ha csak a napszak hatása érdekel minket, és az ismételt mérésekre csupán kontrollálni akarunk, akkor megtehetjük, hogy a napszak fix hatása mellé csak a Hallgató ID-t emeljük be random hatásként. Ha a ZH-k mintázatát pontosabban akarjuk modellezni, de a ZH-kat nem akarjuk egymással összehasonlítani, ezt is megtehetjük a random hatások szintjén belül a Hallgatói ID és ZH-sorszám változók segítségével. Végül ha a kutatási kérdésünk a napszak mellett a tanulási görbére is kiterjed, mert azt keressük, hogy melyik ZH volt a legnehezebb, vagy hogy nő-e a különbség a reggeli és délutáni csoportok között a félév során, akkor természetesen beemelhetjük a ZH-sorszámot fix hatásként is a modellbe. A lényeg, hogy az LMM-ben ez teljes mértékben a kutató választásától függ, nem pedig a módszer korlátaiból származó kényszer.

Ötödik példa: ismételt mérések a személyazonosító bevonásával

Az első lehetőség tehát az, hogy a Napszak fix hatása mellé csak a Hallgató ID-t helyezzük be random intercepttel azért, hogy az ismételt mérésből származó függőségre kontrolláljuk, de a sorrendiséget jelző ZH-sorszám változó még nem része a modellnek. Ez a választási lehetőségeink közül a leggyengébb megoldás, és várhatóan a legalacsonyabb illeszkedést eredményező, mivel a ZH-k sorrendiségét nem veszi figyelembe, és nem engedi meg, hogy a ZH-k nehézsége változzon. Létezhet azonban olyan ismételt mérés is, ahol ténylegesen nincs szerepe a mérési sorrendnek, például egy személytől egy időben több nyálminta gyűjtése esetén.

31. doboz: Egyszerű ismételt méréses modell felépítése

lme4:
Teljesitmeny ~ Napszak + (1|HallgatoID)

SPSS:
```
MIXED Teljesitmeny BY Napszak
 /FIXED=Napszak | SSTYPE(3)
 /RANDOM= INTERCEPT | SUBJECT(HallgatoID) COVTYPE(VC).
```

Ekkor a G-mátrix konceptuálisan egy skalár lesz σ^2_H értékkel, ahol H a hallgatót jelenti (10. mátrix). Ez a random intercept a hallgatók közötti varianciát vagy a másik szemszögből megfogalmazva az egy hallgató által írt ZH-k hasonlóságát magyarázza.

$$G = \begin{matrix} \text{intercept} \\ \text{intercept} \end{matrix} \begin{bmatrix} \sigma^2_H \end{bmatrix}$$

10. mátrix: Egyszerű ismételt méréses elrendezés G-mátrixa

A G-mátrix adatszintű bontása egy 150-150-es mátrix (11. mátrix), ahol az átlóban a σ^2_H jelenik meg, mivel az egyazon hallgatóhoz tartozó ZH-kat egymáshoz hasonlónak feltételezzük, az átlón kívül viszont minden érték nulla, hiszen a hallgatókat függetlennek tekintjük egymástól. Ezt az alakot már jól ismerjük, a G-mátrix alakja Variance Components.

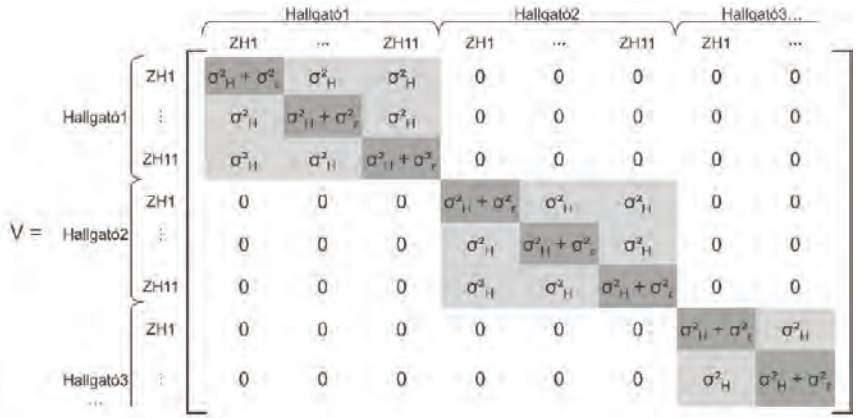
$$G = \begin{array}{c} \\ \text{Hallgató1} \\ \text{Hallgató2} \\ \text{Hallgató3} \\ \vdots \\ \text{Hallgató149} \\ \text{Hallgató150} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{Hallgató1} & \text{Hallgató2} & \text{Hallgató3} & \ldots & \text{Hallgató149} & \text{Hallgató150} \end{array} \\ \begin{bmatrix} \sigma^2_H & 0 & 0 & \cdots & 0 & 0 \\ 0 & \sigma^2_H & 0 & \cdots & 0 & 0 \\ 0 & 0 & \sigma^2_H & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2_H & 0 \\ 0 & 0 & 0 & \cdots & 0 & \sigma^2_H \end{bmatrix} \end{array}$$

11. mátrix: G-mátrix adatszintű felbontása ismételt méréses elrendezésben

Az R-mátrix továbbra is Scaled Identity, tehát a maradékhibát továbbra is független zajnak feltételezzük (a Scaled Identity alakra példa a 3. mátrix). A V-mátrix (12. mátrix) mérete (150*11)x(150*11) a 150 hallgató 11 ZH-jának megfelelően. Az átlóban $\sigma^2_H + \sigma^2_\varepsilon$ szerepel, tehát a modell megengedi, hogy a hallgatók eltérő átlagteljesítményt nyújtsanak, az egy hallgatóhoz tartozó 11 ZH hasonlóságát σ^2_H jelöli, míg a különböző hallgatók ZH-inak találkozásánál a mátrix nullát tartalmaz a hallgatók függetlenségének feltételezése miatt. A V-mátrix egy-egy hallgatóhoz tartozó részének alakját *összetett szimmetriának* (vagy egyenletes korrelációnak, Compound Symmetry, CS) nevezzük. Az összetett szimmetria kifejezetten az ismételt méréses elrendezés leírására szolgáló kovarianciastruktúra, azok közül a legegyszerűbb, amelyben a mérések varianciái, valamint az egy blokkon (egy személyen) belüli bármely két mérés közötti kovarianciák is állandóak. A teljes V-mátrix alakja blokk-diagonális CS, mert egy-egy hallgatón belül CS alakot vesz fel, de az eltérő hallgatók ZH-inak kovarianciája 0. Vigyázzunk, ebben a példában a CS alak a V-mátrixot írja le, míg a G-mátrix alakja továbbra is VC, így a szintaxisban is a VC-beállítást kell választanunk a SUBJECT(HallgatoID)-hez.

A V-mátrix blokk-diagonális CS alakjának megfelelően a modellünk feltételezi, hogy minden megfigyelés (mérés) varianciája azonos (szóráshomogenitás feltételezése), azaz az átlóban végig azonos variancia, $\sigma^2_H + \sigma^2_\varepsilon$ szerepel. A mérések közötti kovarianciát azonosnak feltételezi függetlenül attól, hogy a hallgató melyik két méréséről van szó. A mátrixban ez úgy jelenik meg, hogy egy hallgató blokkján belül az átlót kivéve minden cella a σ^2_H-t tartalmazza.

A V-mátrix a CS-blokkon kívül viszont csak nullát tartalmaz, tehát bármely két eltérő hallgatóhoz tartozó mérést függetlennek feltételezi egymástól.



12. mátrix: V-mátrix egy modellhez egy random hatás random interceptjével

A 12. mátrixnál fontos megemlíteni, hogy a ZH1-11 kiírása a könyvben csak a szemléltetést szolgálja, de a ZH-sorszám változót ez a modell nem ismeri. Ez azért nem okoz problémát, mert a ZH-k között homogén összefüggést feltételezünk.

Ha valaki összehasonlítja a 10–12. mátrixot a 2–4. mátrixszal, észreveheti, hogy az egymásnak megfelelő táblázatok formailag azonosak. Ez nem véletlen, mivel az ismételt méréseknek jelenleg egy annyira egyszerű változatát nézzük, amely még csak azt feltételezi, hogy a mérések összefüggenek egymással, ha ugyanazon hallgatótól származnak, ami azonos jellegű feltételezés azzal, minthogy a hallgatók összefüggenek egymással, ha azonos kurzusra járnak. A modell nem veszi figyelembe az ismételt mérések sorrendiségét. A hagyományos széles formátumot használó elemzési módszereknél az összefüggőség ilyen alapvető formája nem volt megvalósítható, hiszen ott a változók önmaguk adták meg a mérések sorrendjét.

Merre lépjünk tovább? Az LMM ismételt mérésénél a kutató döntése, hogy figyelembe veszi-e a mérések sorrendjét a ZH-sorszám változó akár fix, akár random hatásként való megadásával. Ha a ZH-sorszámot random hatásként emeli a modellbe, azzal azt engedi meg, hogy a modell a ZH-k közötti összefüggést finomabb mintázattal becsülje. Különböző beállításokkal lehetőség van arra, hogy az egyes ZH-k átlaga eltérhessen egymástól, hogy a különböző hallgatóknál eltérő ZH-k legyenek sikeresebbek, vagy hogy az egymáshoz időben közelebb álló

mérések jobban hasonlítsanak egymásra. A ZH-sorszámnak fix hatásként való modellbe emelése pedig a ZH-k közötti különbségek explicit modellezését jelenti majd, mely során összehasonlíthatóvá válnak a ZH-k egymással (vizsgálható az ötödik és harmadik ZH eltérése vagy az, hogy a napszak hatása nő-e a félév során), de emlékezzünk, hogy ennek nagyobb a szabadságfokköltsége.

Hatodik példa: ismételt mérések sorrendisége random interceptként

A modellt megadhatjuk így:

32. doboz: Sorrendiséget kezelő random intercept modell

lme4:

```
Teljesitmeny ~ Napszak + (1|HallgatoID) + (1|ZHsorszam).
```

SPSS:

```
MIXED Teljesitmeny BY Napszak
 /FIXED= Napszak | SSTYPE(3)
 /RANDOM= INTERCEPT | SUBJECT(HallgatoID) COVTYPE(VC)
 /RANDOM= INTERCEPT | SUBJECT(ZH-sorszam) COVTYPE(VC).
```

Az olvasónak feltűnhet, hogy a fenti szintaxis formailag azonos azzal, amit két, egymástól független, keresztmetszeti random hatás (például a Kurzus ID és a Pontozó Tanár ID) modellezése során láttunk. A hasonlóság nem véletlen, hanem a kevert modellek egységes logikájából fakad. A modell matematikai szempontból ugyanis nem tesz különbséget aközött, hogy a mérések közötti hasonlóság egy közös csoporttagságból vagy egy közös forrásból ered. Ahogy a Kurzus ID és a Pontozó Tanár ID két független, keresztezett random tényező volt, úgy ebben az ismételt méréses elrendezésben a Hallgató ID és a ZH-sorszám is két, egymást keresztező random hatásként értelmeződik. A modell így egyszerre veszi figyelembe a variancia két forrását: egyrészt, hogy vannak ügyesebb-ügyetlenebb hallgatók (a Hallgató ID hatása), másrészt, hogy vannak könnyebb-nehezebb ZH-k (a ZH-sorszám hatása).

A modell G-, R- és V-mátrixa is megegyezik azzal, amit a korábbi keresztezett elrendezésben láttunk. A G-mátrix alakja továbbra is Variance Components, az R-mátrix Scaled Identity, míg a V-mátrix a keresztezett elrendezésekre jellemző komplex alakkal rendelkezik.

Sőt, a párhuzamot tovább fűzve, ahogy egy keresztmetszeti modellben a hallgatók be lehetnek ágyazva a kurzusokba és a kurzusok az intézetekbe, és

ezt a hierarchikus struktúrát keresztezheti a pontozó tanár random hatása, úgy egy ismételt méréses modell is lehet hasonlóan komplex. Elképzelhető egy olyan elrendezés, ahol a ZH-kat a hallgatók fűzik csokorba, a hallgatók a kurzusokba vannak ágyazva, és a ZH-sorszám mindkettővel keresztezett viszonyban áll.

Merre lépjünk tovább? A valódi különbség a független mintás klasztereződés és az ismételt méréses elrendezések modellezése között a következő lépésben mutatkozik meg. Míg egy keresztmetszeti csoporttagság hatása állandó, az ismételt méréseknek van egy olyan tulajdonsága, amellyel a keresztmetszeti adatok nem rendelkeznek: a sorrendiség. Ez a mostani példa ezt a sorrendiséget még nem vette figyelembe, hanem a ZH-sorszám változót csak a ZH-k azonosítására használta, azok egymáshoz viszonyított idői pozícióját figyelmen kívül hagyva. Az LMM lehetővé teszi, hogy ezt a sorrendiséget is modellezzük, és megengedjük, hogy az egy hallgatóhoz tartozó ZH-k ne azonos mértékben hasonlítsanak egymásra, és olyan sokkal realisztikusabb modellt is építhetünk, amely figyelembe veszi, hogy a közelebbi mérések vélhetően jobban korrelálnak, mint a távolabbiak. Ennek a fajta, időben változó függőségi struktúrának a modellezésére a tisztán keresztmetszeti példákban nincs analógia.

Hetedik példa: ismételt mérések sorrendiségének modellezése az R-mátrix strukturálásán keresztül

Az ismételt mérések modellezésének egy másik gyakran használt módjával ismerkedünk most meg, mely a maradékhiba-struktúrát módosítja. Az eddigi modellek helyett most egy olyat építünk, amely a ZH-sorszám hatását nem a G-, hanem az R-mátrixban kezeli. Ezt is lépésről lépésre építjük fel, és először egy olyan modellt nézünk meg, amely minden mérés között függetlenséget feltételez, majd erre építve vezetjük be az időbeli függőséget. A szintaxis a következő:

33. doboz: R-mátrix strukturálása – Scaled Identity

lme4:
Nincs elterjedt alternatívája. Ennek az elrendezésnek a futtatásához az ötödik példában bemutatott szintaxist várja.

SPSS:

```
MIXED Teljesitmeny BY Napszak
 /FIXED= Napszak | SSTYPE(3)
 /RANDOM= INTERCEPT | SUBJECT(HallgatoID) COVTYPE(VC)
 /REPEATED=ZHsorszam | SUBJECT(HallgatoID) COVTYPE(ID).
```

A szintaxisban feltűnik egy új beállítás, a /REPEATED sora. Ez a parancs nem egy új random hatás hozzáadására való, tehát nem a G-mátrixot módosítja. Helyette az R-mátrix struktúráját határozza meg.

A /REPEATED sorban a COVTYPE(ID) van beállítva, amely a Scaled Identity rövidítése, és a maradékhibák függetlenségének feltételezését jelenti. Míg az eddigi példákban az R-mátrix alakját mi magunk soha nem adtuk meg, hanem a szoftverek alapbeállítását használtuk a Scaled Identityhez, most ezzel a paranccsal expliciten kimondjuk a Scaled Identity alakot (lásd a 3. mátrixot) és ezzel a függetlenséget. A modell jelentését nem változtatja meg ez a beállítás, továbbra is ugyanaz a modell fut, mint a /REPEATED sor nélkül, ezzel az ötödik példának megfelelő elrendezést hozunk létre (vigyázzunk az összevetésnél, hogy az ötödik és nem pedig a hatodik példával azonos az itt tárgyalt modell). A hallgatók közötti eltérést a Hallgató ID-hez rendelt random intercept modellezi a G-mátrixban, míg a /REPEATED... sor ugyan megadja a ZH-k sorrendjét, de nem engedi meg a maradékhibák összefüggését (mert az R-mátrix az átlón kívül a COVTYPE(ID) miatt csak nullákat tartalmaz). Ez a modell csak annyival elegánsabb, mint az ötödik példa, hogy látja a ZH-k sorrendjét, és egyértelműen kimondja azok függetlenségét, tehát az R-mátrix Scaled Identity alakját. Mivel a G- és R-mátrix nem változik az ötödik példához képest, így természetesen a V-mátrix is azonos az ott látottakkal. Ez a példa csak azt szemlélteti, hogy ugyanaz a modell többféle beállítással is elérhető.

Technikai szempontból rendkívül fontos tudatosítanunk, hogy a hatodik példához képest viszont a (1|ZHsorszam)-nak megfelelő random hatás eltűnt a modellből, mivel redundanciához vezetne, ha a ZH-sorszám szerepét egyszerre modelleznénk a G- és az R-mátrixban is.

Lépjünk tovább oda, ahol az R-mátrix strukturálása már valódi többletinformációt ad!

34. doboz: R-mátrix strukturálása – Autoregresszív (AR1) modell

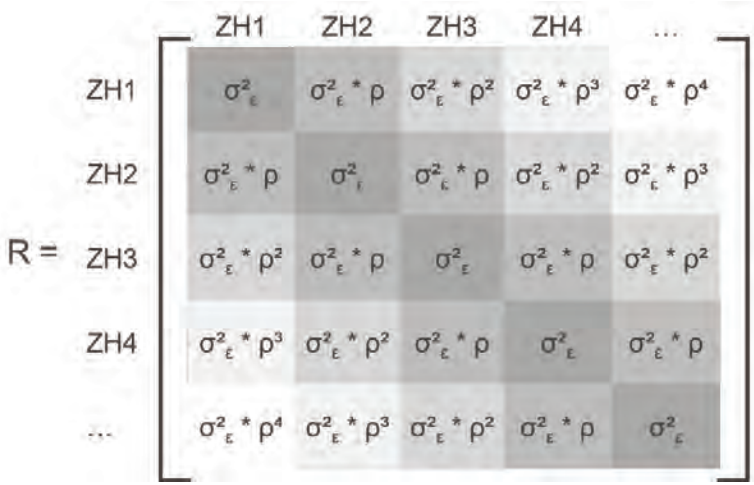
lme4:
Nincs alternatívája, csak egy másik LMM-et megvalósító R-csomagban.

nlme:
Teljesitmeny ~ Napszak, random = ~ 1 | HallgatoID, correlation = corAR1(form = ~ ZHsorszam | HallgatoID)

SPSS:
```
MIXED Teljesitmeny BY Napszak
 /FIXED= Napszak | SSTYPE(3)
 /RANDOM= INTERCEPT | SUBJECT(HallgatoID) COVTYPE(VC)
 /REPEATED=ZHsorszam | SUBJECT(HallgatoID) COVTYPE(AR1).
```

Az AR(1) egy elsőrendű autoregresszív struktúrát jelöl (13. mátrix). Az R-mátrix átlójában a maradékhiba varianciája (σ^2_ε) szerepel. Az átlón kívül azonban már nem nullák állnak, hanem a variancia és egy, a mérések időbeli távolságától (k) függő korrelációs tag (ρ, ró) szorzata: $\sigma^2_\varepsilon * \rho^k$. Ezzel a modell egy rövid távú memóriát ad a hibatagnak. Egy személyen belül a korreláció bármely két mérés hibája között ρ lesz az egymást követő elemekre, ρ^2 a két lépés távolságra lévőkre, ρ^3 a három lépés távolságra lévőkre és így tovább. Mivel a ρ nulla és egy közötti, a hatványozással egyre kisebb értéket kapunk, ennek megfelelően a modell azt feltételezi, hogy minél távolabb esik két ZH egymástól, annál kisebb a hibáik közötti hasonlóság. Mivel ez a számítás a távolságot (k) a mérések sorszáma alapján határozza meg, az AR(1) struktúra alkalmazása csak akkor ad helyes eredményt, ha a mérések között eltelt időintervallumok (lépésközök) azonosak (például minden ZH között egy hét telik el).

Ebben a modellben a Hallgató ID továbbra is random interceptként, a G-mátrixon keresztül modellezi a stabil, személyre jellemző különbségeket, és a G-mátrix alakja megegyezik a 11. mátrixban látottakkal. A ZH-sorszám változó viszont már nem egy független random hatás, hanem az R-mátrix maradékhibáinak idői strukturálására szolgál, amelynek segítségével a modell megtanulja, hogy az egymáshoz közelebbi ZH-k megmagyarázatlan hibái jobban hasonlítanak egymásra, mint a távolabbiaké.



13. mátrix: Példa az AR(1) struktúrára

Láthattuk, hogy a sorrendiség modellezhető a G- és R-mátrixon keresztül is (a hatodik példa keresztezett elrendezése a G-mátrixban és a jelenlegi példa AR(1) alakja az R-mátrixban). Mi a különbség a kettő között? Képzeljük el, hogy városok napi hőmérsékletét modellezzük. A fix hatás az évszakok hatása, ahol vizsgálható, hogy van-e különbség a nyári és téli hőmérséklet között. A random hatás az egyedi területek mikroklímája. Azt mondjuk, hogy vannak általában melegebb és hidegebb városok. Ez egy strukturális magyarázat, ami egy konkrét entitáshoz, a városhoz kötődik. Az AR(1) struktúra ezzel szemben az időjárás lendületének modellezése. Azt mondjuk, hogy függetlenül az évszaktól és a mikroklímától, ha tegnap valamiért megmagyarázhatatlanul meleg volt, akkor ma is valószínűleg melegebb lesz a vártnál. Ez nem egy entitáshoz kötődő magyarázat, hanem a zaj időbeli, sztochasztikus mintázatának leírása.

A mikroklíma és az időjárás lendülete is magyarázza a varianciát, csak teljesen más oldalról közelítik meg a problémát, egyik sem eredendően gyengébb magyarázat, mint a másik. Viszont egymást kizáró lehetőségek, mert egy változó hatását nem tudjuk egyszerre a G- és R-mátrixban is modellezni, mert az redundanciához vezetne.

Mi alapján dönthető el, hogy a ZH-sorszámot random hatásként vagy sztochasztikus mintázatként emeljük a modellbe? A döntés két dologtól függ. Az első az elméleti megfontolás. Mi a valószínűbb mechanizmus az adatok mögött? Stabil, egyéni különbségek, vagy inkább rövid távú, mérésről mérésre átcsorgó hatások? Általában a random hatás egy erősebb, elméletileg jobban

alátámasztható magyarázat az emberi teljesítményre. A másik döntő tényező pedig a modell illeszkedése. Bizonytalanság esetén mindkét alternatív modell lefuttatható és az illeszkedési mutatóik mentén összehasonlítható. Amelyik modellnek jobb az illeszkedése, az a statisztikai értelemben vett jobb modell, mert jobban egyensúlyoz a magyarázóerő és az egyszerűség között.

A /REPEATED... beállításnál az AR(1) mellett további struktúrák is megadhatók, de ezek használata rendkívül ritka, így említést csak kettőről teszünk.

Az AR(1) struktúránál egy komplexebb idősoros mintázatot ír le az ARMA (Autoregressive Moving Average). Ez a modell a maradékhibák struktúráját két komponensből építi fel, az első, az autoregresszív (AR) rész a korábbi hibák emlékezetét modellezi, míg a mozgóátlag (MA) rész a korábbi, véletlenszerű sokkok elhúzódó hatását veszi figyelembe. Bár ez a megközelítés rendkívül rugalmas, pszichológiai vagy társadalomtudományi adatoknál ritkán van rá szükség, mivel a mögöttes elméleti mechanizmusai nehezen értelmezhetők, és a legtöbb függőségi mintázatot a random hatásokkal hatékonyabban lehet modellezni.

A legrugalmasabb, de egyben legkockázatosabb beállítás a strukturálatlan (Unstructured, UN) kovarianciastruktúra. Amikor ezt választjuk, a modell nem tesz semmilyen előzetes feltételezést a hibák mintázatáról, hanem szabad kezet kap, és megpróbál megbecsülni egy egyedi varianciát minden egyes mérésre, és egy egyedi kovarianciát minden egyes méréspárra. Előnye, hogy tökéletesen képes lekövetni bármilyen bonyolult, valós adatstruktúrát. Hátránya azonban, hogy rendkívül adatéhes, mert egy k ismételt méréses modellben k*(k+1)/2 darab paramétert kell becsülnie, ami nagymértékben növeli a túlillesztés veszélyét és a modell instabilitását. Emiatt csak kevés (leginkább maximum három-négy) ismételt mérés és rendkívül nagy minta esetén használható.

Ezen a ponton egy szakmai kitekintést teszünk azért, mert gyakran hallani azt a félinformációt, hogy az ismételt mérések leírására a CS-struktúra szolgál. Ez a kitekintés a tananyag logikai ívének nem része, és esetlegesen ki is hagyható. Következőként állítsuk be így a szintaxist!

35. doboz: R-mátrix strukturálása – Compound Symmetry

lme4:
Nincs elterjedt alternatívája. Ennek az elrendezésnek a futtatásához az előzőekben bemutatott szintaxist várja.

SPSS:
```
MIXED Teljesitmeny BY Napszak
 /FIXED= Napszak | SSTYPE(3)
 /REPEATED=ZHsorszam | SUBJECT(HallgatoID) COVTYPE(CS).
```

Hasonlítsuk össze a szintaxist a 33. dobozban láthatóval! Eltérést jelent, hogy eltűnt a Hallgató ID-hez tartozó /RANDOM sor, valamint a /REPEATED sorral az R-mátrix alakját a Scaled Identityről Compound Symmetryre változtattuk.
A Hallgató ID-hez tartozó random intercept kihagyásával a hallgatók közötti különbséget, azaz az egy hallgatóhoz tartozó ZH-k hasonlóságát teljes egészében áthelyeztük a G-mátrixból az R-mátrixba (14. mátrix). Ennek megfelelően az R-mátrixban jelenik meg a maradékhiba σ^2_{ε} mellett a hallgatók közötti különbséget jelölő σ^2_{H} is. Fontos, hogy ez a kovarianciatag $\sigma^2_{H,}$ és nem σ^2_{ZH}, mert a hallgató az egység (a /REPEATED sorban a SUBJECT(HallgatoID)), a ZH sorszám csak a sorrendiséget adja, és a modell a hallgatókon belüli ZH-k hasonlóságát és ezzel a hallgatók különbözőségét feltételezi, nem pedig a ZH-k különbözőségét! Emellett a G-mátrix üressé vált, hiszen az egyazon hallgatón belüli ZH-k hasonlóságát már tartalmazza az R-mátrix, azt a G-mátrixban is meghatározni redundáns lenne, és a modellezés bizonytalanságához vezetne.[5]

[5] Itt tennünk kell egy megjegyzést a jelölésről. A fenti bekezdés varianciakomponens-szemléletet követ, tehát az R-mátrixban expliciten megkülönbözteti a hallgatóhoz köthető varianciát (σ^2_{H}) és a maradékhiba varianciáját (σ^2_{ε}). A CS-struktúrát egy másik, elterjedt parametrizálásban is megadhatjuk, amely két paramétert használ. A σ^2 a megfigyelés teljes varianciája, míg a ρ a megfigyelések közötti korreláció. Ebben a formálisabb jelölésben az R-mátrix átlója σ^2, az átlón kívüli elemek pedig $\rho \cdot \sigma^2$. A két megközelítés matematikailag teljesen ekvivalens, mert $\sigma^2_{H} = \rho \cdot \sigma^2$ és $\sigma^2_{\varepsilon} = (1-\rho) \cdot \sigma^2$. A varianciakomponens-szemlélet előnye, hogy közvetlen kapcsolatot teremt a random interceptes modellel, és intuitívabban mutatja, mekkora részt magyaráz a hallgatók közötti különbség.

$$R = \begin{array}{c} \\ ZH1 \\ ZH2 \\ \vdots \\ ZH10 \\ ZH11 \end{array} \begin{array}{c} \begin{array}{ccccc} ZH1 & ZH2 & \cdots & ZH10 & ZH11 \end{array} \\ \begin{bmatrix} \sigma^2_H + \sigma^2_\varepsilon & \sigma^2_H & \sigma^2_H & \sigma^2_H & \sigma^2_H \\ \sigma^2_H & \sigma^2_H + \sigma^2_\varepsilon & \sigma^2_H & \sigma^2_H & \sigma^2_H \\ \sigma^2_H & \sigma^2_H & \sigma^2_H + \sigma^2_\varepsilon & \sigma^2_H & \sigma^2_H \\ \sigma^2_H & \sigma^2_H & \sigma^2_H & \sigma^2_H + \sigma^2_\varepsilon & \sigma^2_H \\ \sigma^2_H & \sigma^2_H & \sigma^2_H & \sigma^2_H & \sigma^2_H + \sigma^2_\varepsilon \end{bmatrix} \end{array}$$

14. mátrix: Ismételt mérés megadása a R-mátrix CS alakján keresztül – egy hallgató CS-blokkja

Ha összeadjuk az üres G-mátrixot a CS alakú R-mátrixszal, akkor a V-mátrixhoz megint csak ugyanazt az alakot kapjuk, amelyet a 12. mátrixban láttunk. Ez tökéletesen rávilágít arra, hogy a mérések közötti összefüggést többféleképpen is modellezhetem. Ugyanazt a V-mátrixot háromféleképpen is létrehoztuk, mindhárom esetben azt feltételezve, hogy a hallgatók teljesítménye eltérhet egymástól. Az első megoldás (ötödik példa) tisztán a G-mátrixon keresztül történt, ahol a HallgatóID-nek random intercept lett adva. Ez a modell a ZH-k sorrendjét nem ismeri, ami a legtöbb esetben nem probléma, de pontatlan megjelenítés, mert nem közli egyértelműen, hogy az R-mátrixban a hibák függetlenségét feltételezzük. A második megoldás (hetedik példa első verziója) a hallgatók különbözőségét továbbra is a HallgatóID random interceptjén keresztül adja meg, de emellett a modellbe emeli a ZH-sorszámot is, és az R-mátrix ID-struktúrájával egyértelműen kijelenti, hogy a maradékhibák függetlenségét feltételezzük. A kitérőben bemutatott megoldás elhagyja a G-mátrixból a HallgatóID random interceptjét, helyette a hallgatók különbözőségét (azaz a ZH-k hasonlóságát) teljes egészében az R-mátrixban modellezi CS-struktúrával. Mindezt azért mutattuk meg, hogy szemléltessük, a LMM-ben egy modellnek több alternatív beállítási módja is létezik, ami a módszert rugalmassá teszi, de a kezdő kutató számára rendkívül zavarba ejtő lehet.

Végül a témához kapcsolódóan megjegyezzük azt is, hogy bár a /REPEATED mezőt legtöbbször idői strukturáltság modellezésére használjuk, valójában bármely ismételt mérésre (például térbeli vagy szituatív), és nem csak idői dinamikák modellezésére alkalmas.

Nyolcadik példa: ismételt mérések sorrendiségének modellezése fix hatásként

A ZH-sorszám változót fix hatásként is modellezhetjük, ha a kutatásunk szempontjából az időbeli változás mintázata is érdekes. Ekkor a ZH-sorrend bekerül a modell fix részébe, és a továbbiakban nincs hatással a G- vagy az R-mátrix struktúrájára. A modell a ZH-sorszámot ugyanolyan prediktorként kezeli, mint bármely más fix hatást (például a Napszakot), így ezen a szinten nincs különbség a független mintás és az ismételt méréses elrendezések között. A mérések közötti függőség kezelése ugyanis nem a ZH-sorszám fix hatásának a feladata, hanem azért továbbra is a modellben random hatásként szereplő Hallgató ID a felelős.

Ez az LMM egyik legfontosabb előnye a hagyományos, széles formátumot használó ANOVA-eljárásokkal szemben. Míg ott az ismételt mérések függősége és sorrendisége elválaszthatatlanul összekapcsolódott, addig az LMM szétválasztja a két funkciót, és a függőség kezelését a random hatásokra hagyja, a trend modellezését pedig (opcionálisan) a fix hatásokra bízza.

A ZH-sorszám fix hatásként való kezelésekor azonban felmerül egy gyakorlati probléma. A tizenegy mérés kategorikus változóként való kezelése tíz szabadságfokot is felemésztene, ami csökkentheti a modell statisztikai erejét, különösen, ha nem vagyunk kíváncsiak minden egyes ZH közötti különbségre. Egy sokkal hatékonyabb megoldás lehet, ha a ZH-sorszámot folytonos változóként, azaz kovariánsként kezeljük, ezzel egy lineáris trendet modellezve. Ez a megközelítés kevesebb szabadságfokot használ, és olyan kérdésekre ad választ, mint hogy van-e általánosan, folyamatosan növekvő teljesítmény a félév során (a ZH-sorszám főhatása), vagy a Napszak hatása változik-e az idővel, azaz a délutáni csoportok lemaradása nő-e a félév végére (a ZH-sorszám × Napszak interakció).

Összefoglalva az ismételt mérések során mindig meg kell adnunk egy azonosítóváltozót, amely kijelöli, hogy melyik egységeken belül történt az ismételt mérés. A mi példánkban ez a Hallgató ID volt. Ezenfelül a legtöbb esetben az ismételt méréshez tartozik egy második változó is, mely a mérések sorrendiségét azonosítja. A mi példánkban ez a ZH-sorszám volt. Míg a Hallgató ID egyértelműen random hatás kell hogy legyen a modellben, addig a ZH-sorszámnál van választási lehetőségünk. Beemelhetjük fix hatásként, ha a ZH-k összehasonlítása a kutatási kérdésünk részét képezi. Ha nem képezi részét, akkor dönthetünk aközött, hogy random hatásként a G-mátrix részeként jelenjen meg, vagy ismételt méréses elemként az R-mátrix maradék varianciáját strukturálja.

1.10. Második szint, a random slope

A fix hatások modellezésénél láttuk, hogy azok a hagyományos ANOVA-vizsgálatokhoz hasonlóan működnek. Lehetőségünk van a főhatások vizsgálatára, de a modellezés gyakran a faktorok közötti interakciók elemzésekor válik igazán izgalmassá. Például ha a nem és az autizmus hatását vizsgáljuk az érzelemfelismerésre, a főhatások fontos kérdéseket válaszolnak meg, mint például, hogy eltér-e az autista és a kontrollszemélyek teljesítménye, vagy van-e nemi különbség. Az igazi csemege azonban az interakció, amely azt teszteli, hogy az autizmus hatása vajon eltérő mértékű-e férfiak és nők esetében, például, hogy a férfiaknál kisebb-e a különbség az autista és a kontrollcsoport között, mint a nőknél.

De miért hoztuk most fel a fix hatások közötti interakciókat? Azért, mert a random slope (a véletlen meredekség) koncepcionálisan szintén egy interakció, méghozzá egy fix és egy random hatás közötti keresztezett interakció. A random slope azt a kérdést teszteli, hogy egy prediktor (ami jellemzően fix hatás) hatásának nagysága vajon eltérő-e a random hatás különböző szintjein, tehát a prediktor és kimeneti változó összefüggését leíró regressziós egyenes eltérő meredekségű-e a random hatás különböző csoportjaiban.

Bár elméletileg lehetséges olyan modellt építeni, amely csak random slope-ot tartalmaz random intercept nélkül, ez a modell szinte sohasem indokolható elméletileg. Helyette szinte mindig olyan modellt építünk, mely egyszerre tartalmaz random interceptet és slope-ot. (Tekintsünk vissza a 28. ábra B és C paneljére!)

Az ilyen elrendezést olyan gyakorlati kérdések megválaszolására használjuk, mint hogy eltérően reagálnak-e a munkavállalók a növekvő munkahelyi nyomásra. Itt a fix hatás a munkahelyi stressz mértéke (például a túlórák száma), a random hatás pedig a vizsgálatban részt vevő munkavállalók személye. A random slope bevezetésével azt engedjük meg a modellben, hogy a stressz hatása egyénenként eltérő legyen a teljesítményre. Lesznek olyanok, akiknél a növekvő nyomás drasztikusan rontja a teljesítményt (meredek negatív lejtő), míg másoknál ez a hatás enyhébb, vagy akár serkentőleg is hathat (laposabb vagy pozitív lejtő). Vajon minden iskolában ugyanolyan hatékony egy új digitális oktatási módszer bevezetése? Ebben az esetben a fix hatás az új módszer alkalmazása (vagy az azzal töltött idő), a random hatás az egyes iskolák azonosítója. A random slope azt a kérdést teszteli, hogy a módszer hatékonysága (a regressziós egyenes meredeksége) iskolánként variálódik-e, mert lehetséges, hogy bizonyos iskolákban a módszer látványos javulást eredményez a diákok jegyeiben, míg más intézményekben (például eltérő felszereltség miatt) ugyanez a módszer alig hoz változást. Minden gyermek szókincse azonos ütemben

fejlődik az életkorral? Itt a fix hatás az életkor (mint folytonos változó), a random hatás pedig a gyermekek személye. A random slope segítségével nemcsak azt mondjuk ki, hogy a gyerekek induló szókincse eltérő (ez lenne a random intercept), hanem azt is modellezzük, hogy a fejlődés sebessége (az életkor hatásának meredeksége) egyénenként változik, mert vannak gyorsan tanuló gyerekek meredekebb fejlődési görbével, és lassabban gyarapodó szókincsű társaik laposabb görbével.

Kilencedik példa: példa a random slope-ra

Vegyünk egy olyan példát, ahol a tíz kurzus hallgatói a félév során csak egy ZH-t írtak, ami előtt megkérdeztük, hogy hány órát szántak a felkészülésre. (Azért a felkészülési idővel, és nem a megszokott napszak vizsgálatával kezdünk, mert az felvet majd néhány érdekes részletet, ezért a második példában térünk csak ki rá.) A random interceptet tartalmazó szintaxis így nézne ki:

Teljesitmeny ~ Felkeszules + (1|KurzusID).

Ez a modell vizsgálja a felkészülés hatását a teljesítményre, miközben feltételezi, hogy vannak ügyesebb és kevésbé ügyes kurzusok. Ez felel meg a 28. ábra A paneljének. Tegyük fel, hogy az eddigieken felül azt is feltételezzük, hogy bizonyos kurzusoknál a felkészülés jobban hat a teljesítményre, mint más csoportoknál (28. ábra C panelje). Ez konceptuálisan megfelel a hagyományos eljárások interakciós vizsgálatának. Az LMM-ben a szintaxis a következőképp néz ki:

36. doboz: Random slope modell felépítése

lme4:
Teljesitmeny ~ Felkeszules + (1 + Felkeszules | KurzusID)

SPSS:
```
MIXED Teljesitmeny WITH Felkeszules
 /FIXED = Felkeszules | SSTYPE(3)
 /RANDOM = INTERCEPT Felkeszules | SUBJECT(KurzusID) COVTYPE(UN)
```

A teljesítményt magyarázzuk a Felkészülés fix hatásával, ez megegyezik az eddig tanultakkal. A szintaxis módosulása a Kurzus ID random hatásánál látható, ahol a random intercept jelölője mögé bekerült a Felkészülés is. A /RANDOM...

sort úgy olvashatjuk ki, hogy a modell feltételezi, hogy a kurzusoknak különböző alapteljesítménye van ($\sigma^2_{intercept}$), és a felkészülés hatása is eltérő mértékű a különböző kurzusoknál (σ^2_{slope}). Mindez a G-mátrix alakjában jelenik meg (15. mátrix).

G =	KurzusID INTERCEPT (random intercept)	KurzusID Felkészülés (random slope)
KurzusID INTERCEPT (random intercept)	$\sigma^2_{intercept}$	cov(int, slope)
KurzusID Felkészülés (random slope)	cov(int, slope)	σ^2_{slope}

15. mátrix: Példa G-mátrixra random intercept és random slope esetén

A táblázat átlójában szereplő $\sigma^2_{intercept}$ a kurzusok közötti eltérést modellezi (korábban σ^2_K-val jelöltük), míg σ^2_{slope} a felkészülés hatásának variálódását a különböző kurzusok között. Ezek mellett van a modellnek még egy rendkívül izgalmas és informatív tagja, az intercept és slope közötti kovariancia. A cov(int, slope) tag azt írja le, hogy a csoportok kiindulási szintje (azaz az interceptjük) és a prediktor hatásának meredeksége (azaz a slope-juk) között van-e szisztematikus kapcsolat. Más szavakkal ez a kovariancia azt méri, hogy azok a kurzusok, ahol a hallgatók átlagosan jobban teljesítenek (magasabb intercept), vajon egyben azok is, ahol a felkészülés hatékonyabb (meredekebb slope). A kovariancia (és a belőle számított korreláció) előjele és nagysága több lehetséges forgatókönyvet is felvázol, melyet a 29. ábrán szemléltetünk, az egyszerűség kedvéért csak három kurzus regressziós vonalát felrajzolva. Természetesen a következőkben ismertetett a kimenetelek nem egyszerre jelentkeznek, hanem az alábbi felsorolás azokat a lehetséges variációkat mutatja be, amelyek közül a saját adataink elemzésekor várhatóan egyetlen, domináns mintázatot fogunk majd azonosítani.

A leggyakrabban talán negatív kovarianciával találkozunk, ami egyfajta kompenzációs hatást, felzárkózást vagy plafonhatást jelez (29. ábra A panelje). A negatív kovariancia jelentése az, hogy azok a csoportok, amelyek alacsonyabb kiindulási szinttel rendelkeznek (tehát alacsony az interceptjük), meredekebb emelkedést mutatnak (meredekebb slope). Ezzel szemben a magasabb

kiindulási szintű csoportoknál a prediktor hatása gyengébb (laposabb slope). Egy ilyen eredmény lehetséges magyarázata az lehet, hogy a gyengébben induló kurzusokon (ahol a hallgatók előzetes tudása alacsonyabb, ezért felkészülés nélkül alacsony pontszámot érnének el) minden egyes tanulással töltött óra rengeteget számít, mert a hallgatók a nulláról építkezve gyorsan fejlődnek. Ezzel szemben a jobb előzetes tudással érkező kurzusokon a hallgatók már eleve sokat tudnak a témáról. Náluk a felkészülés már csak a finomhangolásról szól, így egy plusz óra tanulás kevesebb pontszámnövekedést eredményez (ezt nevezik plafonhatásnak is).

A fix hatás negatív iránya esetén (29. ábra B panelje) a negatív kovariancia egyfajta padlóhatásként jelenik meg, mivel a magasabb kezdőértékkel rendelkező csoportoknál a legmeredekebb az esés. Egy ilyen eredményre jó példa lehet, ha a vizsgaidőszakban szórakozással töltött idő hatását nézzük a teljesítményre. Az amúgy is gyengébb kurzusok hallgatói a szórakozás növekedésével gyorsan elérik a legalacsonyabb teljesítményt, amely után már nem tud tovább esni a teljesítményük, így ezeknél a kurzusoknál az egyenes kevésbé lesz negatív.

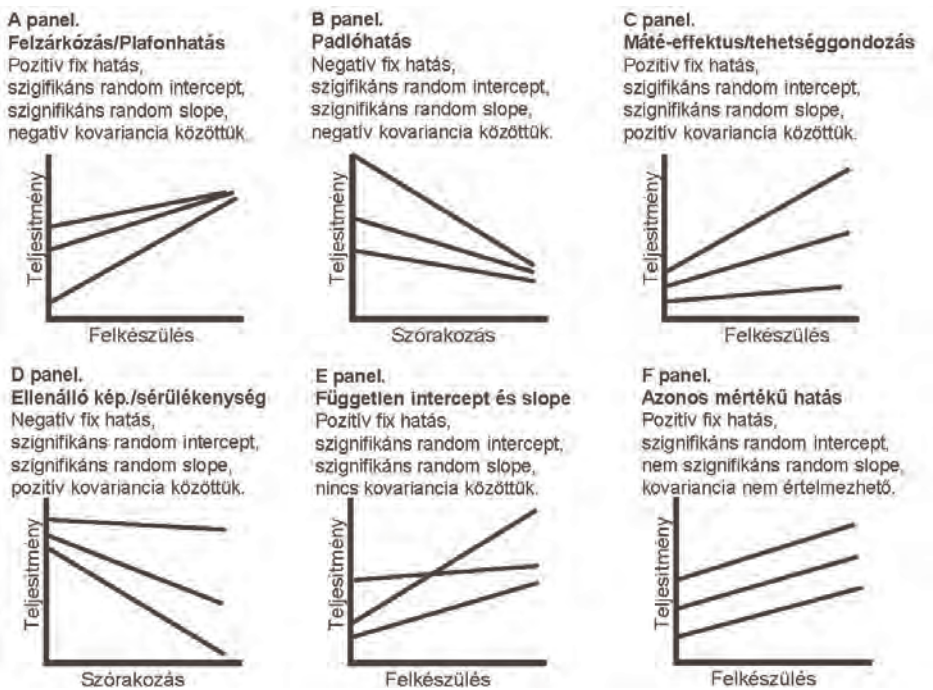
A pozitív kovariancia egy „a gazdagok még gazdagabbak lesznek" típusú jelenségre utal, amit a szociológiában Máté-effektusnak is neveznek („akinek van, annak adatik"). Ilyet látunk a tehetséggondozó programoknál is (29. ábra C panelje). Pozitív kovarianciát látunk, ha azok a csoportok, amelyek eleve magasabb kiindulási szinttel rendelkeznek (magasabb intercept), mutatják a legerősebb növekedést is (meredekebb slope). Ez a helyzet állhat elő, ha a legjobb kurzusokon nemcsak a hallgatók alapképességei jobbak, hanem motiváltabbak is, ezért jól strukturált, közös jegyzetet készítettek. Egy eleve tehetséges és motivált csoport jó jegyzetekkel sokkal többet tud profitálni minden egyes tanulással töltött órából, mint egy gyengébb képességű, motiválatlanabb csoport. A felkészülés hatékonysága itt felerősíti a már meglévő különbségeket.

Negatív fix hatás mellett a pozitív kovariancia egyfajta sérülékenységet jelent (29. ábra D panelje). Ha a kevésbé ügyes csoportok az idejüket tanulás helyett szórakozással töltik, az látványosan meg fog jelenni az eredményükben, míg az ügyesebb csoportok több szórakozás mellett is a korábban megszerzett háttértudás segítségével képesek magukat részben a felszínen tartani.

Előfordulhat az is, hogy nincs szisztematikus kapcsolat a kiindulási szint és a növekedés üteme között (29. ábra E panelje), tehát azt, hogy egy kurzus ügyese (magas intercept), semmilyen módon nem befolyásolja, hogy ott mennyire hatékony a felkészülés. Ebben az esetben a felkészülés hatékonysága teljesen független attól, hogy a hallgatók milyen előzetes tudással érkeztek a kurzusra. Lehet egy gyengén induló kurzus, ahol a felkészülés nagyon hatékony, de lehet egy másik, ami szintén gyengén indul, de a tanulás alig segít.

Végül ott, ahol a kurzusoknak nincs eltérő tanulási görbéje, tehát minden kurzusnál hasonló mértékű a fix változó hatása, és ennek megfelelően a regressziós egyenesek párhuzamosak, nincs szignifikáns random slope, így nincs értelme kovarianciáról sem beszélni (29. ábra F panelje).

Összességében tehát a random intercept és random slope közötti kovariancia feltárása egy rendkívül értékes diagnosztikai eszköz, amely mélyebb betekintést enged a csoportok közötti különbségek dinamikájába, és segít megérteni, hogy a vizsgált hatások felerősítik, gyengítik egymást, vagy függetlenek-e a már meglévő egyenlőtlenségektől.



29. ábra: Tipikus kovarianciák a random intercept és random slope között. A három vonal három különböző kurzust jelöl.

Tizedik példa: a random slope igényei

Vegyük a negyedik példát. Itt a hallgatók a félév során csak egy ZH-t írtak. A klasztereződést tekintve a hallgatók kurzusokba tartoztak, a kurzusok pedig különböző intézetekhez. A random intercepteket tartalmazó szintaxis így nézett ki:

$$\text{Teljesitmeny} \sim \text{Napszak} + (1 \mid \text{IntezetID}) + (1 \mid \text{KurzusID}).$$

Ehhez a modellhez szeretnénk a random interceptek mellé random slope-ot is adni.

Az első korlát az, hogy a Napszak random slope-ját a kurzusok random hatásához nem lehet hozzáadni, hiszen a kurzusok be vannak ágyazva a napszakba (minden kurzus vagy csak reggeli, vagy csak délutáni lehet). A random slope becslésének statisztikai előfeltétele, hogy a prediktor (amelynek a hatását vizsgáljuk, például Napszak) értékei változzanak a csoportosító változó (a random hatás, például Kurzus ID) minden szintjén belül. Ebben a példában ez a feltétel nem teljesül, ezért a modell matematikailag és konceptuálisan is értelmezhetetlen.

A második korlát az, hogy az intézet random hatása mellé sem tudunk random slope-ot rendelni. Itt nem a keresztezettség hiányával van a probléma, hiszen az Intézet ID és a Napszak egymással keresztezett változók, mert mind a Pszichológia, mind a Pedagógia Intézetnek vannak reggeli és délutáni kurzusai is. A probléma az intézetek számával van. Korábban a csak random interceptet tartalmazó modellnél is láttuk, hogy az intézet már ott is problémás random hatásként viselkedett, mert mindössze két szintje van (pszichológia és pedagógia), és ez a szintek közötti variancia ($\sigma^2_{\text{intercept}}$) becslését rendkívül bizonytalanná teszi. Egy ilyen becslés statisztikai ereje alacsony, hiszen mindössze két adatpontból próbáljuk meg egy eloszlás szórását megragadni. A random slope modellnél azonban a helyzet matematikailag is lehetetlenné válik. Egy random intercept és random slope modellt kérve arra utasítjuk a programot, hogy becsüljön meg három paramétert, az interceptek varianciáját ($\sigma^2_{\text{intercept}}$), a slope-ok varianciáját (σ^2_{slope}) és az intercept és a slope közötti kovarianciát (cov(int, slope)), de ennek a három paraméter becsléséhez a modellnek mindössze két információja van, amelyek a Pszichológia Intézet és a Pedagógia Intézet adatai. Egy ilyen egyenletrendszer aluldeterminált, statisztikailag nem lehet megoldani. A program valószínűleg hibaüzenettel le is állna.

A legtisztább és statisztikailag leginkább védhető megközelítés az, ha az intézetet átemeljük a fix hatások közé, így a modellből kikerül a problémás

random hatás. Az eredeti kérdésünket, vagyis hogy a napszak hatása eltér-e a két intézetben, továbbra is meg tudjuk vizsgálni, csak nem egy random slope, hanem egy fix interakció formájában (Napszak × Intézet). Ha ez az interakció szignifikáns, az pontosan azt jelenti, amit a random slope-pal kerestünk, hogy a napszak hatása függ az intézettől.

Ha valamilyen elméleti okból mégis ragaszkodunk ahhoz, hogy az intézetet random hatásként kezeljük, kénytelenek vagyunk a modellt egyszerűsíteni. A legkisebb egyszerűsítés az, ha a kovarianciamátrix típusát strukturálatlan (UN) helyett diagonálisra (DIAG) állítjuk. Az UN becsüli a kovarianciát is, míg a DIAG nullának tekinti azt, és csak az intercept és a slope varianciáját becsüli. Ezzel kettőre csökkentjük a becsülendő paraméterek számát, ami a modell futtatását matematikailag már lehetségessé teszi, de a becslés továbbra is rendkívül instabil marad a két intézet miatt. A következő logikus lépés, ha belátjuk, hogy az adatok nem teszik lehetővé a slope varianciájának becslését. Ilyenkor visszatérünk a csak random interceptet tartalmazó modellhez, és elfogadjuk, hogy a napszak hatásának intézetenkénti variabilitását nem tudjuk modellezni. Végső esetben, ha az intézet mint random intercept nem okozza az illeszkedés szignifikáns javulását, akár teljesen el is hagyhatjuk a modellből, ezzel feladva az intézetek közötti különbségekről szóló információt.

Tizenegyedik példa: a random slope ismételt méréseknél

A kevert modellek rugalmasságának és erejének bemutatására vegyünk egy utolsó, komplexebb példát, amely már valós kutatási helyzeteket idéz. Tegyük fel, hogy egy félév során szeretnénk megérteni a hallgatók tanulási folyamatát tíz különböző kurzuson. Azt vizsgáljuk, hogyan változik a ZH-teljesítményük az idő előrehaladtával, és hogy ezt a változást befolyásolja-e a napszak és a kurzus, amelyre járnak.

A modell változói:

a) A modellben a napszakot fix hatásként kezeljük.
b) A ZH-k sorszámát fix kovariánsként kezeljük, ami azt jelenti, hogy a ZH sorszáma egy folytonos prediktor változó, amely a félév során bekövetkező átlagos, szisztematikus változást, például az egyre jobb vagy rosszabb teljesítményt hivatott leírni, szemben a kategorikus faktorokkal, mint a napszak.

c) A modell felépítésekor kétszintű hierarchiát veszünk figyelembe, ahol a hallgatók a kurzusokba vannak beágyazva, és ezt a struktúrát a Kurzus ID random hatásával jelöljük.
d) A vizsgálat ismételt méréses jellegét a random hatásként beemelt Hallgató ID és a fix hatásként vizsgált ZH-sorszám jelöli.

A modell felépítése:

a) A Kurzus ID esetében random interceptet állítunk be, ezzel megengedve, hogy a kurzusoknak eltérő legyen a kiinduló teljesítményszintje.
b) Azt is feltételezzük, hogy a félév során tapasztalható fejlődési ütem kurzusfüggő, ezért a Kurzus ID-nél a ZH-sorszámmal egy random slope-ot is illesztünk.
c) A Hallgató ID-t is random hatásként vesszük be a modellbe, ami egy random interceptet jelent minden egyes hallgató számára. Ezzel elismerjük, hogy a teljesítményt a kurzus és a ZH-k hatásain felül stabil, egyéni különbségek is befolyásolják, tehát vannak eleve jobb és gyengébb képességű diákok.
d) A két fix hatás, a napszak és a ZH-sorszám között interakciót engedünk meg, ezzel vizsgálva, hogy a reggeli és délutáni kurzusok teljesítménye közötti eltérés növekedik-e a félév során. Ez az lme4 szintaxisában a Napszak:ZHsorszam kifejezéssel, az SPSS szintaxisában a Napszak*ZHsorszam kifejezéssel jelölhető.

Ebben a struktúrában az ismételt mérések kezelése különösen elegáns, mert az idő múlása itt félig random, félig fix módon van modellezve. A ZH-sorszám fix hatása azt a szisztematikus, mindenkinél átlagosan megfigyelhető trendet ragadja meg, amely a tanulási folyamat közös, általánosítható része. Ezzel szemben a Hallgató ID random interceptje az időben stabil, egyéni különbségeket modellezi, vagyis azt, hogy egy adott hallgató konzisztensen mennyivel teljesít jobban vagy rosszabbul az átlagnál. A longitudinális modellek ereje pontosan ebben a kettősségben rejlik, mivel egyszerre képesek ábrázolni az általános változási trendeket és az egyéni sajátosságokat.

37. doboz: Komplex, többszintű modell interakciókkal

lme4:

```
Teljesitmeny ~ Napszak + ZHsorszam + Napszak:ZHsorszam + (1 | HallgatoID) + (1 +
ZHsorszam | KurzusID)
```

SPSS:

```
MIXED Teljesitmeny BY Napszak WITH ZHsorszam
  /FIXED = Napszak ZHsorszam Napszak*ZHsorszam
  /RANDOM = INTERCEPT ZHsorszam | SUBJECT(KurzusID) COVTYPE(UN)
  /RANDOM = INTERCEPT | SUBJECT(HallgatoID) COVTYPE(VC).
```

Egy ilyen összetett modell varianciastruktúrája is árnyaltabb képet mutat (16. mátrix). Mivel a modellben kurzus- és hallgatói szintű random hatások is szerepelnek, a G-mátrix egy blokk-diagonális szerkezetet vesz fel. Paraméterszintű felrajzolásakor az egyik blokk egy 2x2-es mátrix, amely a kurzusok interceptjeinek és a ZH-sorszámra illesztett slope-jainak varianciáit, valamint a kettő közötti kovarianciát tartalmazza (itt a beállítás UN). A másik blokk egyetlen érték, a hallgatói interceptek varianciája (itt a beállítás VC). Az R-mátrix alakját az alapértelmezett Scaled Identityn hagyjuk, hiszen az R-mátrix strukturálásához nem használhatjuk fel a ZH-sorszámot, mert azt már fix hatásként használjuk.

$$G = \begin{array}{l} \\ \text{KurzusID random intercept} \\ \text{KurzusID és Zhsorszam random slope} \\ \text{HallgatóID random intercept} \end{array} \begin{array}{c} \begin{array}{ccc} \text{KurzusID random intercept} & \text{KurzusID és Zhsorszam random slope} & \text{HallgatóID random intercept} \end{array} \\ \begin{bmatrix} \sigma^2_K & \text{cov(int, slope)} & 0 \\ \text{cov(int, slope)} & \sigma^2_{\text{slope}} & 0 \\ 0 & 0 & \sigma^2_H \end{bmatrix} \end{array}$$

16. mátrix: Példa G-mátrixra random intercept és random slope esetén

1.11. Összegzés

Összegezzük a példákban bemutatott szintaxisokat lme4-ben!

38. doboz: A leggyakoribb LMM-modelltípusok lme4-szintaxisai

Fix hatások és a köztük lévő interakció:
Y ~ Fix1 + Fix2 + Fix1:Fix2
vagy rövidítve:
Y ~ Fix1*Fix2

Egy fix hatás és egy random hatás random intercepttel:
Y ~ Fix + (1 | Rand)

Egy fix és random keresztezett hatások random intercepttel:
Y ~ Fix + (1 | Rand1) + (1 | Rand2)

Egy fix és beágyazott random hatások random intercepttel:
Y ~ Fix + (1 | RandSzülő / RandGyerek)

Egy fix hatás ismételt méréssel, a sorrend figyelmen kívül hagyásával:
Y ~ Fix + (1 | RandID)

Egy fix hatás ismételt méréssel, a sorrendet figyelembe véve random intercepttel:
Y ~ Fix + (1| RandID) + (1 | RandSOR)

Ismételt mérés is fix hatásként figyelembe véve:
Y ~ Fix + FixSOR + (1| RandID)

Random intercept és random slope modell a kettő közötti kovarianciával:
Y ~ Fix + (1 + Fix | Rand)

A következőkben összefoglaljuk a fejezet során megismert mátrixalakokat is. Az itt felsorolt alakok nem fedik le az összes lehetséges struktúrát, de a gyakran használt elemek mindegyikét tartalmazzák.

39. doboz: Variancia-kovariancia mátrixalakok kislexikona (ID, VC, CS, AR(1), UN)

Scaled Identity (SI vagy ID): Minden átlós elem azonos (σ^2), az átlón kívüli elemek nullák. Legtöbbször az R-mátrix alakjának leírására használjuk, ha feltételezzük, hogy a mérési hibák minden időpontban azonos szórással rendelkeznek, és teljesen függetlenek egymástól. Ez a hagyományos lineáris regresszió alapfeltétele, ismételt méréseknél ritkán reális.

Variance Components (VC): Az átlós elemek blokkonként eltérőek lehetnek, az átlón kívüli elemek nullák. Legtöbbször a G-mátrix alakjának leírására használjuk, ha több, egymástól függetlennek feltételezett random hatásunk van (például keresztezett random hatások), és mindegyiknek külön varianciát szeretnénk becsülni. Ez a standard beállítás a G-mátrixnál, ha nincs random slope.

Diagonal (DIAG): Az átlós elemek eltérőek (σ^2_1, σ^2_2, ...), az átlón kívüli elemek nullák. Legtöbbször az R-mátrixnál használjuk, amikor a mérések hibái függetlenek egymástól, de a varianciájuk időpontonként eltérő (heteroszkedaszticitás). Például ha a félév végére a hallgatók közötti különbségek megnőnek, így a ZH-k szórása is nagyobb lesz.

Compound Symmetry (CS): Az átlós elemek azonosak (σ^2), és az összes átlón kívüli elem is azonos (cov). Legtöbbször ismételt mérések R-mátrixánál használjuk. A legegyszerűbb módja a mérések közötti függőség modellezésének. Azt feltételezi, hogy bármely két mérés korrelációja azonos, függetlenül a köztük eltelt időtől. A repeated measures ANOVA ezt a feltételt teszteli (szfericitás). Használatakor a személyID-hez nem szabad random interceptet rendelni, mert redundanciához vezetne.

Autoregressive(1) (AR(1)): Az átlós elemek azonosak (σ^2). Az átlón kívüli elemek korrelációja exponenciálisan csökken (ρ, ρ^2, ρ^3, ...) az időbeli távolság növekedésével. Legtöbbször ismételt mérések R-mátrixánál használjuk, ahol az időben közelebbi mérések jobban hasonlítanak egymásra, mint a távolabbiak. Sokkal realisztikusabb feltételezés, mint a CS.

Unstructured (UN): Teljesen rugalmas. Minden átlós elem (variancia) és minden átlón kívüli elem (kovariancia) egyedi értéket vehet fel. Használjuk G- és R-mátrix leírására is. A G-mátrixnál ez az alapbeállítás random slope modelleknél, mert lehetővé teszi a random intercept és a random slope közötti kovariancia becslését. Az R-mátrixnál akkor használjuk, ha nincs elméleti elképzelésünk a hibák mintázatáról, és hagyjuk, hogy az adatok maguk mondják el a teljes struktúrát, de vigyázni kell, mert nagyon sok paramétert becsül, ezért sok adatot igényel, és sokszor túl összetett, túl specifikus és megbízhatatlan modellt eredményez.

1.12. Modellilleszkedés

A kevert modellek középpontjában az a felismerés áll, hogy az adatainkban a variancia nem véletlenszerű zaj, hanem strukturált mintázat. A modell felépítése során lényegében egy elméleti tervrajzot, egy hipotézist fogalmazunk meg erről a struktúráról. Ez a tervrajz három kulcsfontosságú mátrixban ölt testet, a G-, R- és V-mátrixokban. A G-mátrix a klaszterező csoportok közötti varianciát írja le, vagyis azt, hogy a random hatásaink (például a kurzusok interceptjei és slope-jai) hogyan szóródnak, és hogyan kapcsolódnak egymáshoz. Az R-mátrix a csoportokon belüli varianciát, azaz a mérési hibák struktúráját modellezi. A kettőből összeálló V-mátrix pedig a teljes, megfigyelt adatokra vonatkozó variancia-kovariancia struktúra elméleti modellje. Fontos kiemelnünk, hogy a G-, R- és V-mátrixok elméleti modellek, alakjuk a kutató világról való elképzelése szerint alakul, nem pedig a tényleges adatokból kiindulva jön létre.

Az illesztés (modelfitting) nem más, mint ennek az elméleti tervrajznak a rápróbálása a valós, mért adatokra. A folyamat során a statisztikai program célja, hogy megtalálja azokat a konkrét paraméterértékeket (a fix hatás együtthatóit, a varianciákat és a kovarianciákat), amelyekkel a mi elméleti V-mátrixunk a lehető legpontosabban írja le a valós adatokban megfigyelhető szórás- és korrelációs mintázatot. Ez a folyamat nem egy egyszerű egyenletmegoldás, hanem egy iteratív, közelítő eljárás, amely addig finomítja a becsléseket, amíg a lehető legjobb egyezést nem éri el az elmélet és a valóság között.

Ennek az iteratív folyamatnak a megvalósítására a leggyakrabban a Maximum Likelihood (ML) becsléscsaládot használjuk. Az ML egy olyan statisztikai módszer, amely szisztematikus próbálgatással úgy állítja be a paraméterértékeket, hogy azok mellett a legvalószínűbb legyen a megfigyelt minta előfordulása. A Restricted Maximum Likelihood (REML) ennek egy finomított változata, amely különösen a kevert modellek esetében vált alapértelmezetté. A hagyományos ML egyszerre becsüli a fix és a random hatásokat, és hajlamos enyhén alábecsülni a varianciakomponenseket. A REML ezt a torzítást úgy korrigálja, hogy a varianciák becslésénél figyelembe veszi a fix hatások becsléséből adódó bizonytalanságot.

A modell illeszkedésének jóságát több szempontból is értékelnünk kell, mivel nincs egyetlen, mindenre jó mérőszám (mint a lineáris regressziónál az R-négyzet). A leggyakrabban azt vizsgáljuk, hogy a jelenlegi modellünk jobb-e egy egyszerűbb modellnél, tehát a relatív illeszkedés érdekel bennünket. Erre szolgál a Likelihood Ratio Test (LRT), amely megmutatja, hogy egy komplexebb modell (például random slope-pal) szignifikánsan jobb illeszkedést nyújt-e, mint egy egyszerűbb (például csak random intercepttel). Emellett az AIC és BIC

információs kritériumok is a modellek közötti választást segítik, ahol az alacsonyabb érték a jobb, takarékosabb modellre utal.

Emellett vizsgálható az abszolút illeszkedés is. Itt azt vizsgáljuk, hogy a modell önmagában mennyire írja le jól az adatokat. Ennek legfontosabb eszköze a reziduális hibák vizsgálata. Egy jó modellnek csak a véletlenszerű, strukturálatlan zajt szabad hátrahagynia. Ha a reziduális hibát ábrázolva valamilyen mintázatot (például tölcsér alakot, görbületet) látunk, az azt jelzi, hogy az elméleti tervrajzunk (a V-mátrix) valamilyen fontos szempontot figyelmen kívül hagyott a valós adatok struktúrájából. Ez az eljárás hasonló a hagyományos lineáris regresszióban használt predikáltérték-reziduális pontdiagram értékeléséhez. A jó illeszkedés tehát azt jelzi, hogy modellünk specifikációja nem mond ellent az adatok szerkezetének (mivel a reziduumokban már nem maradt szisztematikus információ), ugyanakkor fontos szem előtt tartani, hogy ez önmagában nem zárja ki más, alternatív modellek illeszkedésének lehetőségét sem.

2. Kutatási folyamat

2.1. Mintaválasztás

A kevert modellek legnagyobb ereje a csoportok közötti variancia (a random hatások) becslése. Ahhoz, hogy a program megbízhatóan meg tudja becsülni egy random hatás (például egy random intercept) szórását, elegendő számú csoportra van szüksége, amelyekből ezt a szórást kiszámíthatja. Bár nincs egyetlen, kőbe vésett szabály, a szakirodalom általában a következőket javasolja. A négyhat csoport a technikai minimum, amely alatt a legtöbb statisztikai program már nem is képes megbecsülni a varianciakomponenseket. Ha egy kutató mindössze négy-hat csoporttal rendelkezik, a random hatás varianciájának becslése rendkívül pontatlan lesz, óriási standard hibával. A statisztikai erő annyira alacsony, hogy rendkívül nehézzé válik a valós csoportszintű hatások kimutatása. Ezt a tartományt mindenképpen érdemes elkerülni, vagy megfontolni a változó fix hatásként való modellezését. A tíz-harminc csoport egyfajta realisztikus minimum egy egyszerűbb, csak random interceptet tartalmazó modellhez. Itt a becslések már stabilabbak, de még mindig lehetnek pontatlanok, főleg, ha a csoportok közötti valós variancia kicsi. A harminc-ötven vagy annál több csoport az ideális tartomány. Ekkora mintán már a random hatások varianciája is

megbízhatóan becsülhető, és van elegendő statisztikai erő a komplexebb, random slope modellek futtatásához is. Ha a kutatás fő kérdése éppen a csoportok közötti különbségek mértéke, akkor mindenképpen ezt a nagyságrendet érdemes megcélozni.

A csoportokon belüli adatpontok (például diákok egy osztályban, ismételt mérések egy személynél) száma is fontos, de más okból. A nagyobb csoporton belüli esetszám elsősorban a fix hatások becslésének pontosságát és a csoporton belüli prediktorok hatásának erejét növeli. Ha erőforrásaink korlátosak, érdemesebb inkább több, de kisebb csoport mérését választanunk a kevés, de nagy csoport vizsgálata helyett. Ismételt méréses vizsgálatoknál gyakran kevés, akár három-öt mérési pont is elegendő lehet, ha a felső szintű egységek (például személyek) száma elég magas. A random slope modellek becsléséhez azonban itt is nagyobb mérésszám szükséges.

Az LMM egyik legnagyobb előnye, hogy nincs szüksége kiegyensúlyozott adatokra. Míg a hagyományos ismételt méréses vagy beágyazott ANOVA megköveteli, hogy minden személytől ugyanannyi adatpont származzon vagy minden csoportban ugyanannyi személy legyen, addig az LMM rugalmasan kezeli a hiányzó adatokat és az eltérő csoportméreteket. Nem probléma, ha az egyik kurzusra húsz, a másikra harmincöt hallgató jár, vagy ha az egyik résztvevőtől tizenegy mérési pont van, a másiktól pedig csak nyolc a lemorzsolódás miatt. A modell az összes rendelkezésre álló információt felhasználja, ami nagyobb statisztikai erőt és realisztikusabb modellezést tesz lehetővé.

2.2. Adatgyűjtés, az adatok előkészítése

Az adatgyűjtés és -előkészítés során a legfontosabb szempontok a megfelelő hierarchikus minta gyűjtése, a változók átgondolt kódolása és az adatok szoftverek által értelmezhető formátumba, azaz hosszú formátumba való rendezése. Ahogy már volt róla szó, a kevert modellek alapkövetelménye, hogy az adatok hosszú (long) formátumban legyenek. Ez azt jelenti, hogy minden sor egyetlen megfigyelési egységet (például egy ZH-eredményt egy adott időpontban) tartalmaz, és külön oszlopok jelölik a csoportosító változókat (például Hallgató ID vagy Kurzus ID) és a mérés időpontját (ZH-sorszám). Ez a keskeny és magas adatszerkezet teszi lehetővé a szoftver számára, hogy helyesen értelmezze a hierarchiát és a mérések közötti függőségeket, szemben a széles formátummal, ahol minden ismételt mérés külön oszlopban szerepel.

A folytonos prediktorokat (például Életkor vagy a Felkészüléssel töltött idő) erősen ajánlott standardizálni vagy legalább centrálni, azaz az átlagukat

kivonni minden értékből. A nagyátlagcentrálás (grand-mean centering) különösen hasznos, mert ennek köszönhetően a modell interceptje egy értelmes, átlagos értéket fog jelenteni, az átlagos prediktorértékkel rendelkező alany becsült kimenetét. Interakciók és random slope-ok modellezésénél a centrálás a multikollinearitás csökkentésével a modell stabilitását is növeli.

2.3. Az elemzés beállításai

40. doboz: A lineáris kevert modellek (LMM) szoftveres lehetőségeinek áttekintése

A lineáris kevert modellek futtatására ma már számos statisztikai szoftver kínál robusztus és kifinomult megoldásokat. A kutatók körében vitathatatlanul az R a legnépszerűbb választás, elsősorban a rugalmasságot nyújtó lme4 csomagnak köszönhetően. A kereskedelmi szoftverek közül az SPSS MIXED parancsa a grafikus felületnek köszönhetően könnyebb belépést biztosít a kezdőknek, valamint kiemelkedik abban is, hogy mind a grafikus felületen, mind a szintaxisban rendkívül gazdag választékot kínál az ismételt mérések kovarianciastruktúráinak (az R-mátrixnak) a modellezésére. A SAS PROC MIXED eljárása különösen a klinikai kutatások és a vállalati környezetek világában elterjedt. Egyre nagyobb teret hódít a Python is, ahol a statsmodels csomag kínál LMM-funkcionalitást. Az olyan szoftverek, mint a JASP és JAMOVI az LMM-et korlátozott beállítási lehetőségekkel futtatják, ami jó belépő lehet az LMM-elemzésekbe, de bonyolultabb modelleknél gyorsan korlátokba ütközhet a kutató. Bár a szintaxisuk eltérő, ezek a programok ugyanazokon a statisztikai alapokon nyugszanak, így a köztük való választás gyakran a felhasználó preferenciáin és a kutatási környezet adottságain múlik.

Egy LMM-elemzéshez minden szoftveres környezetben van néhány alapvető beállítás és mutató, amelyet a kutatónak tudatosan kell meghatároznia. Ezek a döntések alkotják a modell gerincét, és alapvetően befolyásolják az eredmények érvényességét és értelmezhetőségét.

Meg kell határozni a változók szerepét. Melyik lesz a függő változó, melyek a fix hatások (azok a prediktorok, amelyek konkrét szintjeinek hatására vagyunk kíváncsiak), és melyek a random hatások (azok a csoportosító változók, amelyeknek a varianciáját szeretnénk modellezni). Ide tartozik a random hatások komplexitásának meghatározása is, amely során eldöntjük, hogy elegendő egy random intercept, vagy az elméletünk random slope-ot is indokol.

A modellnek meg kell adnunk azt is, hogy milyen mintázatot feltételezzen az adatokban lévő függőségekről. Ezt a G-mátrix (random hatások) és az R-mátrix

(maradékok) struktúrájának beállításával tesszük meg. A G-mátrix alakja a random intercept megengedése esetén a legtöbbször Variance Components (VC). A random slope-ot is tartalmazó modellnél döntés kérdése, hogy az alapértelmezett Unstructured (UN) típust használjuk-e, amely becsüli az intercept és a slope közötti kovarianciát is, vagy egy egyszerűbb, függetlenséget feltételező modellt választjuk (Diagonal). Az ismételt méréses adatoknál a klaszterező változó (személyazonosító) általában VC alakkal kerül a modellezésbe. Az ismételt mérések sorrendiségét jelölő változót többféleképp használhatjuk. Dönthetünk úgy, hogy fix hatásként modellezzük, ekkor természetesen sem a G-, sem az R-mátrixban nem jelenik meg. Dönthetünk úgy, hogy csak a mérések interceptjének eltérését modellezzük a G-mátrixban (VC), vagy hogy az R-mátrixban a maradékhiba strukturálásán keresztül az időbeliség dinamikáját is figyelembe vesszük. Ez utóbbi esetben legtöbbször a közelebbi mérések nagyobb hasonlóságát feltételezzük az Autoregresszive (AR(1)) beállítással. De itt is dönthetünk az egyszerűbb modell mellett pusztán az egy személyhez tartozó mérések homogén hasonlóságát feltételezve.[6] Bárhogy is döntünk, figyeljünk arra, hogy ezek ugyanarra a változóra nézve egymást kizáró beállítások, tehát például az ismételt mérések sorrendiségét vagy csak fix hatásként, vagy csak random interceptel, vagy csak a maradékhiba strukturálásán keresztül tehetjük meg.

Ha megvan a feltételezett modell, választanunk kell a két fő becslési eljárás közül. Bár a matematikai háttér komplex, a gyakorlati döntés egyértelmű szabályokon alapul. A REML (Restricted Maximum Likelihood) -módszer a legtöbb szoftver alapbeállítása. Akkor használjuk, ha a kutatás fókuszában a random hatások varianciájának pontos becslése áll, mert a REML torzítatlan becslést ad a varianciakomponensekre. Az ML-re (Maximum Likelihood) akkor van szükség, ha a célunk különböző fix hatásokat tartalmazó, egymásba ágyazott modellek összehasonlítása a Likelihood Ratio Test (LRT) segítségével.

Végül be kell állítanunk, hogy milyen mutatókat kérünk az eredményekről. A legtöbb programban automatikusan megkapjuk a fix hatások táblázatát. Ez a táblázat hasonlít leginkább egy hagyományos regressziós modell eredményéhez, és tartalmaznia kell a regressziós együtthatókat, amelyek a prediktor hatásának nagyságát és irányát mutatják, a hozzájuk tartozó standard errort, a t- vagy F-értéket és a p-értéket.

[6] Gyakori gyakorlati probléma (különösen kevés számú csoport vagy alacsony csoportok közötti variancia esetén), hogy a modell nem talál stabil matematikai megoldást. Ezt a szoftverek (például az R lme4 csomagja) konvergenciahibával vagy a jellegzetes „boundary (singular) fit" figyelmeztetéssel jelzik. Ez a szingularitás arra utal, hogy a modell túl komplex az adatok információtartalmához képest, és a szoftver a random hatás varianciáját gyakorlatilag nullának (vagy a korrelációkat -1/+1-nek) becsli. Ilyen esetekben a random struktúra egyszerűsítése javasolt.

A random hatások leírásához ki kell kérni a varianciákat, melyek azt jelölik, hogy az adott random hatás (például a kurzusok random interceptje vagy slope-ja) mentén mekkora a szóródás. Ha ez a szám nagy, az azt jelenti, hogy a csoportok között jelentős különbségek vannak. Ha közel van a nullához, az arra utal, hogy a csoportok alig különböznek, és a random hatás talán felesleges is a modellben. A random slope modellek esetén ki lehet kérni a kovariancia értékét is, mely megmutatja, hogy a random intercept és a random slope között van-e szisztematikus kapcsolat. Ezek a mutatók számszerűsítik a hierarchia és a függőség mértékét az adatokban. Ezekből tudjuk meg, hogy valóban indokolt-e a kevert modellek használata, és hogy mekkora a klaszteresődések szerepe a teljes varianciában. Az osztályon belüli korreláció (ICC) például közvetlenül ebből a táblázatból számítható.

A kevert modellek módszertanának fontos része a különböző komplexitású modellek szisztematikus összehasonlítása, hogy megtaláljuk a leginkább takarékos, de az adatokat még jól leíró struktúrát. Ennek két fő stratégiája van. Az egyik a komplextől az egyszerű felé haladás (top-down), ahol a kutató a maximálisan indokolható modellel kezd (például random intercepttel és a fontosabb prediktorokra illesztett random slope-okkal), majd egymás után próbálja meg elhagyni a nem szignifikáns tagokat, minden egyes egyszerűsítést egy Likelihood Ratio Testtel (LRT) igazolva. Ennek előnye, hogy a kiindulási modell kontrollál a lehető legtöbb varianciaforrásra, így csökkentheti az elsőfajú hibát. A másik út az egyszerűtől a komplex felé építkezés (bottom-up), ahol egy alapmodellel indulunk (például csak random intercepttel), és ehhez adunk hozzá újabb, elméletileg indokolt tagokat (például egy random slope-ot), amelyeket csak akkor tartunk meg, ha az LRT szerint szignifikánsan javítják a modell illeszkedését. Ezt a döntési folyamatot segítik a relatív illeszkedési mutatók. A -2 Log Likelihood (-2LL) a modell és az adatok közötti távolságot jelzi. Önmagában nehezen értelmezhető, de a Likelihood Ratio Test (LRT) alapját képezi, amellyel egymásba ágyazott modelleket hasonlíthatunk össze. Az AIC (Akaike Information Criterion) és BIC (Bayesian Information Criterion) a leggyakrabban használt mutatók a modellválasztáshoz. Mindkettő az illeskedést (a -2LL értéket) és a modell komplexitását (a paraméterek számát) veszi figyelembe, az alacsonyabb érték jobb illeszkedést jelez. Ezek a mutatók a takarékosság elvét érvényesítik, tehát azt a modellt részesítik előnyben, amelyik a legkevesebb paraméterrel éri el a legjobb illeszkedést. A BIC erősebben bünteti a komplexitást, mint az AIC. Az AIC és a BIC a legfontosabb eszközeink abban, hogy eldöntsük, egy komplexebb modell (például egy random slope hozzáadása) valóban megéri-e a plusz bonyolultságot.

Végül, miután kiválasztottuk a végső modellt, ellenőriznünk kell annak alapfeltételeit, amit a reziduális hibák ábrázolásával tehetünk meg. A becsült értékek és maradékok pontdiagramján a pontoknak egy véletlenszerű, mindenféle mintázatot nélkülöző felhőt kell alkotniuk a nulla vonal körül. Ha tölcsér alakot vagy görbületet látunk, az a variancia homogenitásának vagy a linearitás feltételének sérülését jelzi. Ez a vizuális ellenőrzés megegyezik a lineáris regressziónál megszokott szóráshomogenitás és linearitás tesztelésével. Végül a maradékhiba normalitásának vizsgálata hisztogrammal vagy Q-Q plottal történik.

Bár a könyv eddigi részeiben törekedtünk a beállítások szoftverfüggetlen leírására, ezen a ponton egy bekezdés erejéig kivételt teszünk, és áttekintjük az SPSS MIXED funkciójának helyes beállítását. Azért döntöttünk emellett, mert bár az SPSS menüjében lehetőség van a modell pontos és rugalmas beállítására, a menü több pontja is hírhedten félreérthető. Amikor megnyitjuk a Linear Mixed Models menüjét, az első felületen két fő területre lehet változókat húzni, a Subject és a Repeated blokkba. A Subject blokkba azokat a random változókat helyezzük, amelyek az adatok klasztereződéséért felelősek (a mi példánkban ilyen volt a Kurzus ID és a Pontozó tanár ID, valamint az ismételt mérések során a Hallgató ID). Az ismételt mérések sorrendiségét jelző változó sorsáról (nálunk ZH-sorszám) döntenünk kell. Ha fix hatásként modellezzük, akkor sem a Subject, sem a Repeated blokkba nem kell betenni. Ha csak a ZH-hoz tartozó random interceptet akarjuk modellezni, megengedve, hogy legyenek általánosan jobban és rosszabbul sikerült ZH-k, akkor a ZH-sorrend is egy klaszterező változó, és a Subject blokkba kerül. Ha a maradékhiba strukturálásán keresztül a mérések idői dinamikáját szeretnénk modellezni (például egymáshoz időben közelebbi mérések hasonlóbbak), akkor kell csak a Repeated blokkba betenni a változót. Továbblépve a faktorok és kovariánsok mezőibe csak a fix változókat kell beemelni (faktornak a csoportosító változókat, kovariánsnak a folytonosakat), a random hatásokat már megadtuk a Subject részen, azokat nem kell még egyszer közölni. A fix menüpont használata egyértelmű, itt tudunk főhatásokat és interakciókat megadni. A random hatások menüjében láthatunk egy Next gombot, amellyel az úgynevezett rétegek (layer) között lehet váltani. Minden klasztereződést okozó random változót egy-egy külön rétegre helyezzük. Ezt nem a Random Effect területén, hanem alul, a Subject groupingban tehetjük meg. Ne felejtsük el minden rétegen bekapcsolni a random hatáshoz az interceptet. Ha a modellünk megáll a random interceptek szintjén, akkor a Random Effect blokk üresen marad, mert az a random slope-ok megadására szolgál. Ha random slope-ot is tervezünk, akkor a megfelelő rétegen áttesszük a Random Effects részen a kiválasztott fix hatást, így létrehozva a fix hatás és a klaszterező tényező közötti random slope-ot. Az LMM futásához természetesen

szükség van az illeszkedés beállításaira is, de ezek már nem tartalmaznak félreérthető pontot.

2.4. Az elemzés értelmezése

A lineáris kevert modellek értelmezése többlépcsős folyamat, amely során a modell általános illeszkedésétől haladunk a specifikus hatások felé. Mielőtt bármilyen együtthatót elemeznénk, a legelső lépés a modell általános illeszkedésének felmérése és a különböző modellalternatívák összehasonlítása. Az AIC és a BIC segítségével kiválaszthatjuk a leginkább takarékos, de az adatokat még jól leíró struktúrát, míg a Likelihood Ratio Testtel formálisan is tesztelhetjük, hogy egy komplexebb modell (például random slope hozzáadása) indokolt-e egy egyszerűbbhöz képest.

Miután kiválasztottuk a statisztikailag leginkább védhető modellt, a következő lépés a modelldiagnosztika. A maradékok grafikus ábrázolásával ellenőriznünk kell, hogy a modell alapfeltételei (például a hibák normalitása és a variancia homogenitása) nem sérültek-e. A maradékok ábráján látható szisztematikus mintázatok arra utalhatnak, hogy a modellünk nem írja le tökéletesen az adatokban rejlő struktúrát.

Az értelmezést a random hatásokkal érdemes folytatni. A random hatásokról szóló táblázat varianciaértékei számszerűsítik a csoportok (például kurzusok, személyek) közötti különbségeket. A nullától jelentősen eltérő variancia igazolja, hogy a random hatás bevonása indokolt volt, és az ICC kiszámításával azt is megbecsülhetjük, hogy a teljes variancia hány százaléka magyarázható a csoportszintű különbségekkel.

A csoportszintű variabilitás megértése után térhetünk rá a kutatási hipotéziseinkre közvetlenül választ adó fix hatások táblázatára. Ennek értelmezése megegyezik a hagyományos regressziós modellekével, tehát a becsült regressziós együtthatók a prediktorok hatásának irányát és nagyságát, a hozzájuk tartozó p-értékek pedig a statisztikai szignifikanciát mutatják. Itt dől el, hogy a modellünk alapján a vizsgált változóknak van-e kimutatható, átlagos hatása a kimenetre. Szignifikáns hatások esetén az eredmény tovább elemezhető a megszokott módokon post hoc és kontrasztvizsgálatokkal.

3. Mikor használjuk és mikor ne használjuk a módszert?

A lineáris kevert modellek alkalmazása akkor indokolt, sőt statisztikailag elengedhetetlen, ha az adataink megsértik a hagyományos lineáris modellek egyik legfontosabb alapfeltételét, a megfigyelések függetlenségét. Ez a függőség legygyakrabban a beágyazott struktúrákban jelenik meg, például amikor diákok osztályokba, az osztályok pedig iskolákba vannak rendezve, vagy ismételt méréses elrendezésekben, ahol egy személytől több időpontban is gyűjtünk adatot. A módszer képes modellezni a csoportosulásból (klaszterződésből) fakadó korrelációt, így elkerüli a standard hibák alábecslését és az elsőfajú hiba megnövekedett kockázatát.

Ennek ellenére az LMM használata bizonyos esetekben felesleges vagy akár félrevezető is lehet. Nincs szükség az LMM-re, ha az adataink valóban függetlenek, például egy egyszerű keresztmetszeti kérdőíves felmérésnél, ahol minden sort egy különálló, független személy adatai alkotnak, vagy olyan ismételt méréses vizsgálatoknál, ahol minden személyhez azonos számú mérés tartozik. Ilyenkor a hagyományos lineáris regresszió vagy ANOVA a megfelelő, takarékosabb eszköz. Ne használjunk LMM-et akkor se, ha a csoportosító változónk szintjeinek száma túl alacsony (minimum négy-hat, de jellemzően tíz alatt), mert a kevés csoportból a modell nem képes megbízhatóan megbecsülni a csoportok közötti varianciát, ami instabil eredményekhez vezet. Ebben a helyzetben a robusztusabb megoldás a csoportosító változó fix hatásként való kezelése. Végül, az LMM alkalmazása akkor sem helyes, ha a függő változónk nem folytonos és normális eloszlású, vagy az összefüggések nem lineárisak, mert az LMM kizárólag ilyen adatokra alkalmazható. Bináris vagy nominális adatok esetén az általánosított lineáris kevert modellek (GLMM) jelentik a helyes választást, nemlineáris modellezéshez pedig a nemlineáris kevert modellek (NLMM).

4. A minőség biztosítása

Egy statisztikai elemzés közlésekor a célunk, hogy az olvasó számára ne csak az eredményt, hanem az egész gondolatmenetet és eljárást átláthatóvá és megismételhetővé tegyük. Az alábbi lista segít abban, hogy az LMM eredményeiről szóló beszámoló teljes körű és szakszerű legyen.

Szükséges közlés (minimum elvárás):

a) A végső modell pontos bemutatása egyértelműen megnevezve a függő változót, a fix hatásokat, valamint a random hatásokat. A random hatásoknál annak megadása, hogy melyik változókra engedtünk meg random interceptet és/vagy slope-ot.
b) Annak közlése, hogy melyik szoftverrel és csomaggal (például R lme4) és milyen becslési módszerrel (ML vagy REML) készült az elemzés.
c) A fix hatások táblázata, amely tartalmazza a fix hatások becsült együtthatóit, standard hibáit és a szignifikanciát jelző p-értékeket.
d) A random hatások táblázata, amely tartalmazza a random hatások becsült varianciáit (vagy szórásainak) és a random slope modellek esetén a köztük lévő kovarianciát/korrelációt.

Ajánlott közlés (jó gyakorlat) a szükséges közlésen felül:

a) A modellválasztás folyamatának leírása, ismertetve, hogy melyik modelleket hasonlítottuk össze (például egy random intercept és egy random slope modell), és mi alapján (például Likelihood Ratio Test, AIC/BIC) választottuk ki a végső modellt.
b) A fontosabb modellek illeszkedési mutatóinak (-2LL, AIC és BIC) közlése, ami kontextust ad a végső modell illeszkedésének jóságáról.
c) A random intercept modell(ek) esetében az ICC értékének közlése, amely megmutatja, hogy a variancia hány százaléka magyarázható a csoportszintű különbségekkel.
d) Annak közlése, hogy a folytonos változókat centrálták-e, és hogy a kategorikus változóknál melyik szint volt a referenciakategória.

Kiváló közlés az ajánlott közlésen túl:

a) A legfontosabb eredmények (különösen a szignifikáns interakciók vagy a random slope hatások) grafikus ábrázolása.
b) Annak leírása, hogy ellenőriztük a modell feltételeit (például a maradékok normalitását és homogenitását), és az eredmények rendben voltak, ami bizonyítja az elemzés statisztikai robusztusságát.
c) A p-értékek mellett a fix hatások becsléseinek 95%-os konfidenciaintervallumait is érdemes közölni, ami teljesebb képet ad a becslések pontosságáról, és segít elmozdulni a merev szignifikáns/nem szignifikáns gondolkodástól.
d) Valamilyen hatásméret (például standardizált együtthatók vagy pszeudo-R^2 értékek) közlése, ami segít az olvasónak felmérni a talált hatások gyakorlati jelentőségét a statisztikai szignifikancián túl.

5. A módszer változatai és haladó alkalmazásai

Az LMM legfontosabb és leggyakrabban használt kiterjesztése az általánosított lineáris kevert modellek (Generalized Linear Mixed Models, GLMM), amely az LMM logikáját alkalmazza a nem normális eloszlású függő változókra, lehetővé téve többek között bináris kimenetek logisztikus vagy darabszámadatok Poisson-modellezését. Ennél is tovább menve, a nemlineáris kevert modellek (Nonlinear Mixed Models, NLMM) olyan helyzetekre kínálnak megoldást, ahol a prediktorok és a kimenet közötti kapcsolat eleve nem lineáris, mint például a növekedési görbék modellezésekor.

6. Összefoglalás

A lineáris kevert modellek a lineáris regresszió kiterjesztéseként egy rugalmas és hatékony keretrendszert kínálnak a statisztikailag nem független, csoportosított adatok elemzésére. Az LMM képes kezelni a hierarchikus struktúrákat, az ismételt méréseket, a kiegyensúlyozott és kiegyensúlyozatlan vizsgálatokat és a személyenként eltérő számú méréseket azáltal, hogy a fix hatásokkal modellezett átlagos trendek mellett a random hatásokon keresztül a csoportok közötti variabilitást is modellezi.

7. Önellenőrző kérdések

1. kérdés: Mi a legalapvetőbb különbség a fix és a random hatások között?

Válasz: A fix hatások olyan prediktorok, amelyeknek minden releváns szintje jelen van a mintában, és a célunk az ezen szintek közötti konkrét különbségek megbecsülése (például kísérleti és kontrollcsoport). Ezzel szemben a random hatások olyan csoportosító változók, amelyek szintjei egy nagyobb populációból vett véletlen mintának tekinthetők, és a célunk nem az egyes szintek hatásának, hanem az általuk képviselt varianciának a megbecsülése (például osztályok, vizsgálóorvosok).

2. kérdés: Mit modellez a G-mátrix egy kevert modellben?

Válasz: A G-mátrix a modell random hatásainak variancia-kovariancia mátrixa. Lényegében ez a modelltervrajza a csoportok közötti különbségekről.

3. kérdés: Mit modellez az R-mátrix, és mikor különösen fontos a beállítása?

Válasz: Az R-mátrix a modell maradékhibájának a variancia-kovariancia mátrixa. A csoporton belüli variabilitás struktúráját írja le. Különösen ismételt méréses adatoknál fontos a beállítása, ahol feltételezhetjük, hogy a hibák nem függetlenek. Olyan struktúrákkal, mint az AR(1), modellezni tudjuk, hogy az időben közelebbi mérések hibái erősebben korrelálnak egymással.

4. kérdés: Mit reprezentál a V-mátrix, és hogyan épül fel?

Válasz: A V-mátrix a teljes, megfigyelt adatok variancia-kovariancia struktúráját reprezentálja egy csoporton belül. Ez a mátrix a random hatások (G) és a maradékok (R) hatásának összegeként áll elő. A V-mátrix írja le, hogy az elméleti modell szerint egy személy két különböző időpontban mért értéke mennyire hasonlít egymásra, figyelembe véve mind az egyéni, stabil különbségeket (G), mind az időspecifikus hibákat (R).

5. kérdés: Mire utal a VC (Variance Components) kovarianciastruktúra?

Válasz: A VC (Variance Components) egy diagonális kovarianciastruktúrát jelöl, ahol a random hatásokat korrelálatlannak tételezzük, de mindegyiknek saját, külön varianciát becsülünk. Ez az alapbeállítás a G-mátrixnál, amikor több, egymástól függetlennek gondolt random hatást (például keresztezett random hatásokat, mint (1 | HallgatoID) + (1 | KurzusID)) illesztünk a modellbe.

Ajánlott irodalom

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2022). *Multilevel and Longitudinal Modeling with IBM SPSS*. Routledge. https://doi.org/10.4324/9780367824273

Hove, D. T., Jorgensen, T. D., & van der Ark, L. A. (2024). Updated guidelines on selecting an intraclass correlation coefficient for interrater reliability, with applications to incomplete observational designs. *Psychological Methods, 29*(5), 967–979. https://doi.org/10.1037/met0000516

Hurlbert, S. H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs, 54*(2), 187–211. https://doi.org/10.2307/1942661

Jiang, J., & Nguyen, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer New York. https://doi.org/10.1007/978-1-0716-1282-8

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer. https://doi.org/10.1007/b98882

West, B. T., Welch, K. B., Gałecki, A. T., & Gillespie, B. W. (2022). *Linear Mixed Models*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003181064

XIII. AZ EREDMÉNYEK ÁLTALÁNOSÍTHATÓSÁGA ÉS PREDIKTÍV EREJE: SZIGNIFIKANCIA, BOOTSTRAP ÉS KERESZTVALIDÁCIÓ

SOLTÉSZ-VÁRHELYI KLÁRA

A fejezet a következő kérdéseket válaszolja meg:

a) Milyen elméleti alapon nyugszik a hagyományos szignifikanciatesztelés?
b) Milyen elméleti alapon nyugszik a konfidenciaintervallum számítása?
c) Milyen elméleti alapon nyugszik a bootstrap módszertana?
d) Milyen elméleti alapon nyugszik a keresztvalidáció eljárása?
e) Mi az elméleti és gyakorlati különbség a fentebb felsorolt eljárások között?

1. Bevezetés

Könyvünk utolsó fejezete elméleti keretet ad a tárgyaltaknak. Ahogy elmélettel kezdtünk, úgy egy elméleti témával is zárunk. Azt a központi kérdést járjuk körbe, hogy milyen módszerek állnak rendelkezésünkre annak eldöntésére, hogy a mintában mért értékek, a csoportok közötti különbségek vagy a változók közötti összefüggések vajon csak a mintára jellemző lokális jelenségek, vagy a teljes populációra is általánosítható igazságok. Az itt bemutatott témák nem konkrét statisztikai eljárások, hanem eltérő elméleti alapokon nyugvó koncepciók és azok gyakorlati megvalósításai. Három nagy megközelítést fogunk áttekinteni, amelyek különböző módokon válaszolják meg az eredmények általánosíthatóságának kérdését.

A legrégebbi a *szignifikanciatesztelés*, amely a korlátozott számítási kapacitás korában alakult ki, amikor a számításokat még kézzel kellett elvégezni. Ez a módszer azt a kérdést vizsgálja, hogy egy mintánkban látott mértékű hatásnak mekkora a valószínűsége egy olyan populációban, melyben nincs jelen a hatás. Logikája szerint akkor tekinthetünk egy hatást valósnak, ha a mintánkban látott eredménynek nagyon kicsi az esélye, hogy a puszta véletlen szülte. Működésének kulcsfontosságú eleme, hogy elméleti alapokon választunk hozzá eloszlásgörbéket. Még ha ennek legtöbbször nem is vagyunk tudatában, de például a t-próba alkalmazásával a t-eloszlást jelöljük ki modellünk alapjául.

A *bootstrap* már a modern számítógépek számítási kapacitását aknázza ki. Az adatok szisztematikus újra-mintavételezésével becsüli a kapott eredmények bizonytalanságát, így jóval kevésbé támaszkodik elméleti eloszlásfeltételezésekre. Lehetővé teszi például egy becslés konfidenciaintervallumának meghatározását anélkül, hogy az adatok eloszlására vonatkozó szigorú feltételezéseket tennénk. Célja azonban végső soron még mindig a klasszikus statisztikához hasonlóan a becslések bizonytalanságának számszerűsítése.

Ezzel szemben a *keresztvalidáció* gyökeresen eltér a hagyományos elméleti megközelítésektől. Az elmúlt évek hatalmas számításikapacitás-növekedésére építve a hangsúlyt az elméletvezérelt modellek validitásáról a prediktív (előre jelző) erőre helyezi át. Célja azt szimulálni, hogy a modell sosem látott adatokat milyen jósággal tudna előre jelezni. A modell jóságát azonban új adatok bevonása helyett a meglévő adatok többszöri okos felosztásával közelíti.

Bármilyen statisztikai tesztelést is végzünk, a logikája gyakorlatilag a következő gondolatra vezethető vissza: van egy kutatási kérdésünk, például

kíváncsiak vagyunk, hogy a férfiak és a nők magassága különbözik-e egymástól, vagy hogy a testi egészség és a mentális jóllét között van-e összefüggés. Olyan választ szeretnénk kapni, amely az általunk vizsgált teljes populációra általánosítható. A populáció azonban szinte soha nem érhető el a maga egészében, ezért a populációból veszünk egy kezelhető méretű mintát, és a méréseinket ezen a mintán végezzük el. A mintán találunk valamekkora hatást, például a férfiak és a nők átlagmagassága között 10 cm különbséget, az egészség és a mentális jóllét között pedig 0,3-as korrelációt. A kérdés az, hogy ez a hatás csak a mintánkra igaz, vagy általánosítható a populációra is.

Ezt a központi kérdést többféleképpen is megfogalmazhatjuk. Mennyi annak a valószínűsége, hogy tévedünk, amikor a mintában látott hatást a populációra általánosítjuk? Mennyi annak a valószínűsége, hogy a populációban valójában nincs különbség a férfiak és a nők magassága között, csak a mintavételi bizonytalanság miatt véletlenül sikerült egy olyan mintát vennünk, amelyben úgy tűnik, hogy a két csoport eltér egymástól? Ha holnap újabb mintát gyűjtenénk, vajon hasonló mintázatot látnánk abban is? Ha száz mintánk lenne, mennyire szóródna a hatás nagysága? Vajon használható-e a modellünk arra, hogy ez alapján döntéseket hozzunk a világban? Képes-e ez a modell elég jól bejósolni egy teljesen új értéket?

Már itt is látszik, hogy bár ezek a kérdések ugyanazt a problémát feszegetik, nem tökéletesen ugyanarra vonatkoznak. A kérdések egy része az eredmények általánosíthatóságára, másik része a megismételhetőségére kérdez rá. Némelyik a modell becsléseinek pontosságát firtatja, míg mások a prediktív erejét. Egyes kérdések arra vonatkoznak, hogy a modellünk pontosan írja-e le a valóságot, mások pedig arra, hogy jól használható-e a gyakorlatban.

De miért ne egyezne meg a mai minta alapján tett becslésem a holnapival? A legalapvetőbb ok a véletlen mintavételezésből adódó *mintavételezési bizonytalanság* (sampling error). A populációban vannak a populációs átlaghoz közelebb álló személyek, és vannak olyanok is, akik az átlagtól jelentősebben eltérnek. Amikor a populációból véletlenszerűen kiválasztunk néhány személyt, elkerülhetetlenül lesznek olyan minták, amelyekben arányaiban több az alacsonyabb, és olyanok, amelyekben több a magasabb értékkel rendelkező egyén. Így a különböző mintákból becsült értékek – legyen az mintaátlag, két csoport különbsége vagy változók együtt járásának mértéke – mintáról mintára kis mértékben eltérőek lesznek.

Vannak azonban más okok is arra, hogy egy minta eredménye ne legyen általánosítható a populációra vagy megismételhető egy újabb mérés során. Ezek egy része az adatfelvétel módszertani hibáiból, más része a modell hibás specifikációjából adódik, de ezeken felül egyéb okokat is számba veszünk.

Az adatfelvétel módszertani hibái:

A *nem reprezentatív* minta (sampling bias vagy általánosabban a selection bias) problémája eltér a véletlen mintavételi hibától. Itt a probléma az, hogy a mintavételi eljárás szisztematikusan torzít, és eleve olyan mintát eredményez, amely nem tükrözi a populációt. Ez nem a véletlen műve, hanem a módszertan hibája. Például egy online politikai felmérés eredményei nem feltétlenül általánosíthatók a teljes népességre, mert csak azokat éri el, akik aktívak az interneten, és érdeklődnek a téma iránt, kihagyva például az idősebb, internetet nem használó rétegeket.

A *külső érvényességet fenyegető tényezők* (threats to external validity) problémája szorosan kötődik az előző ponthoz. Arra utal, hogy egy eredmény lehet belsőleg valid és megismételhető egy adott kontextusban, de nem általánosítható más helyzetekre, időpontokra vagy populációkra. Például egy egyetemi hallgatókon végzett pszichológiai kísérlet eredményei nem biztos, hogy megismételhetők lennének egy idősebb, más kulturális háttérrel rendelkező populációban. Vagy egy 2020-as, pandémia alatti kutatás szorongásról szóló eredménye nem feltétlenül érvényes 2025-ben.

A *mérési hiba* (measurement error) szintén az adatfelvételnél előforduló lehetséges hiba. Előfordulhat, hogy a mintánk és a modellünk is tökéletes lenne, de a mérőeszközünk pontatlan vagy torzít. Ebben az esetben az, amit mérünk, nem a valós jelenség. Például ha a mentális jóllétet egy rosszul megfogalmazott, félreérthető kérdőívvel mérjük, a kapott adatok nem a tényleges jóllétet fogják tükrözni, így a belőlük levont következtetések sem lesznek megbízhatóak, hiába volt a mintavétel véletlenszerű.

A modellspecifikáció hibái:

A modellspecifikációs hibák különösen a hagyományos statisztikai elemzéseket érintik, amelyekben sok lépés elméleti döntéseken alapul.

Hibás eloszlás feltételezése. Előfordulhat, hogy rossz eloszlástípust feltételezünk a populációra (például normális eloszlásúnak tekintjük, és ezért átlagot használunk középértékként, vagy parametrikus próbát választunk, miközben a valóságban a normális eloszlás feltételezése nem helytálló).

Hibás modellstruktúra választása. Ide tartozik, ha nem megfelelő modellt választunk az összefüggések becslésére. Például kihagyunk fontos változókat, amelyek alapvetően meghatározzák az összefüggés természetét (például *confounder* változók), vagy épp túl sok változót, túl sok paramétert használunk, ami *túlillesztéshez* (overfitting) vezet. Utóbbi esetben a modell (például egy egyszerű lineáris kapcsolat helyett egy tucatnyi görbülettel) a minta véletlenszerű zaját is megtanulja.

A fentieken túl más *modellfeltételezések megsértése* is problémát okozhat, például a lineáris regresszió esetén a linearitás, a homoszkedaszticitás vagy a hibák függetlenségének hiánya. Ha ezek a feltételezések nem teljesülnek, a modell becslései torzulhatnak.

Egyéb problémák:

Az előzőeken túl a megismételhetetlen eredmények mögött a leggyakrabban a *túl kicsi mintaméret* áll. A kis mintában túl nagy a szélsőséges értékek befolyásoló ereje, ami növeli a becslések bizonytalanságát, ami miatt nehéz megkülönböztetni a valódi hatásokat a véletlen ingadozásoktól. Ez különösen a szignifikanciatesztelésnél és a hagyományos statisztikai modelleknél jelent problémát. Ha például csak öt férfit és öt nőt vizsgálunk a magasságbeli különbség tesztelésére, akkor akárcsak egy szélsőségesen magas vagy alacsony személy mintába kerülése megszüntetheti vagy másfélszeresére növelheti a két minta közötti eltérést.

Végül a tudományos élet egyik legégetőbb kihívásaként jelen lévő *megismételhetőségi válsághoz* (replication crisis) hozzájárul a kutatók elfogultsága és a szignifikáns eredmények publikálásának kényszere is. Az eredmények megismételhetőségét csökkenti, ha a kutatók szelektíven közlik az eredményeiket (p-hacking), például csak a szignifikáns összefüggéseket publikálják, vagy többszörös tesztelést végeznek a hibaszint korrekciója nélkül. Az előzőekhez kötődik az a nem túl szerencsés elemzési gyakorlat is, hogy a kutatók gyakran előzetes hipotézisek nélkül végigpásztázzák az adatbázist szignifikáns eredmények után kutatva, vagyis tulajdonképpen a zavarosban halászva (fishing for significance vagy data dredging). Ezek a problémák különösen a szignifikanciatesztelésnél gyakoriak.

A következőkben azt mutatjuk be, hogy az általánosíthatóság és megismételhetőség problémájára milyen gyakorlati megoldások születtek válaszként, külön figyelmet fordítva a három legnagyobb és legelterjedtebb módszertanra, a szignifikanciatesztelésre, a bootstrapre és a keresztvalidációra. A témához kapcsolódik, de a könyv nem tárgyalja részletesen a bayesi megközelítést, mely a hagyományos frekventista alapon nyugvó szignifikanciatesztelés tradicionális elméleti alternatívája. Említésre méltók továbbá a témában a permutációs tesztek, a replikációs vizsgálatok, a metaanalízisek és a szenzitivitási vizsgálatok.

2. A szignifikancia

A megbízhatósági mutatók elméletének megértését a leíró statisztikáknál kell kezdeni. Amikor mintát veszünk, azt egy középértékkel, például a mintaátlaggal jellemezzük (M). Emellett megadható az is, hogy a mintában lévő egyedek mennyire szóródnak az átlag körül, és ezt a szórással jelezzük (SD). Ha például kutatásunk az emberek csokoládéfogyasztását vizsgálja, és egy mintában megmérjük, hogy hetente ki hány grammot fogyaszt, akkor eredményül kaphatjuk azt, hogy M = 65, SD = 40.

A szórás csak korlátozott információval szolgál a megbízhatóságról. Arra lehet belőle következtetni, hogy az átlag mennyire jól reprezentálja a mintát, miszerint ha a szórás kicsi, az egyéni értékek viszonylag közel esnek az átlaghoz, tehát az átlag a minta legtöbb tagját jól leírja. Ha viszont a szórás nagy, akkor a minta nagyobb része található jelentősen az átlag alatt vagy felett.

Ha holnap megismételnénk a vizsgálatot, nem valószínű, hogy ismét pontosan M = 65-öt kapnánk. Miért? A véletlen mintavétel miatt. Ha véletlenül néhány, a csokoládét jobban kedvelő ember kerül a vizsgálatba, akkor kicsit nagyobb, mondjuk M = 68-as értéket kapnánk. Ha néhány, a csokoládét nem kedvelő ember (mert ilyenek is vannak), akkor némileg alacsonyabb átlag adódik. Ha pedig véletlenül Gombóc Artúr is a mintánkba kerül, ez az egyetlen szélsőséges érték is jelentősen megemelheti az átlagot.

Ha ötven mintát veszünk, és mindegyiknek kiszámoljuk az átlagát, azt fogjuk látni, hogy a mintaátlagok eloszlásának jellegzetes alakja van. Minél nagyobb és minél több mintát gyűjtünk, a mintaátlagok eloszlása a központi határeloszlás törvénye miatt annál jobban fog hasonlítani a normális eloszlásra.

Mitől függ, hogy mennyire szóródnak a mintaátlagok? Alapvetően két dologtól: egyrészt minél nagyobb az egyedek szóródása, annál könnyebb szélsőséges értékeket tartalmazó mintát venni, tehát annál nagyobb lesz a mintaátlagok várható szóródása is. Másrészt minél nagyobb a minta, egy Gombóc Artúr-féle adat annál kisebb hatással van az átlagra, kevésbé tudja azt elhúzni, tehát annál kisebb lesz a mintaátlagok szóródása. Például egy háromfős mintánál: Személy1 = 60, Személy2 = 70 és Gombóc Artúr = 120, ekkor M = 83,3. Ezzel szemben egy tízfős mintánál: Személy1–Személy9 = [60, 61, 62, 64, 65, 66, 67, 68, 70] és Gombóc Artúr = 120, ekkor M = 70,3. Látható, hogy a mintaátlag ez utóbbi esetben kevésbé mozdult el, tehát a mintaátlag nagyobb minták esetében kevésbé érzékeny a nem reprezentatív, szélsőséges értékekre. A mintaátlagok szórására

egyetlen mintából is következtethetünk, és ezt nevezzük standard errornak (standard hibának), amelynek a képlete SE = SD/√N. Ez a számítás pontosan az előbb felvázolt gondolatmenetnek megfelelően a mintán belüli szóródást és az elemszámot veszi figyelembe. Ez a mutató már pontosabb információval szolgál a megbízhatóságról: azt mutatja meg, hogy ha több mintánk lenne, akkor a belőlük számolt átlagok mennyire szóródnának. Segítségével megbecsülhetjük, hogy ha holnap megismételnénk a mintavételt, a maihoz hasonló mintaátlagra számíthatunk-e, vagy valami teljesen más eredmény fog születni.

Lépjünk tovább, és vizsgáljuk meg, hogy van-e összefüggés a személy által átélt stressz mennyisége és a csokoládéfogyasztás között, feltételezve, hogy azok, akik magasabb stresszt élnek át, nagyobb mennyiségű édességet fognak fogyasztani. Az összefüggés lineáris regresszióval vizsgálható. Tegyük fel, hogy találunk egy B = 2,5-ös meredekséget, aminek a jelentése az, hogy a stresszskálán egy pont növekedés a heti csokoládéfogyasztás 2,5 grammos növekedését eredményezi. Ha ezt a kutatást néhány tucatszor megismételjük, akkor a korábbi gondolatmenetnek megfelelően a kapott meredekségnek is lesz egyfajta szóródása, lehet, hogy egyik mintában csak 2,2 grammnyi növekedést találunk, máshol pedig 2,8 grammot is. Ez a szóródás is a standard error segítségével jellemezhető, amelynek számítása megegyezik a korábban látott logikával, értéke a személyek predikció körüli szóródásától és az elemszámtól függ. Egy további vizsgálatban ellenőrizzük, hogy vajon férfiak és nők között van-e eltérés a csokoládéfogyasztásban. Legyen a nők átlaga M = 64,8, a férfiaké 65,2. A nemek hatása ekkor a két átlag közötti 0,4 grammnyi eltérés. Ha több mintát veszünk, a véletlen mintavételezés miatt természetesen ez az eltérés is kismértékben ingadozni fog, és az ingadozás mértékét a férfi és a női mintán belüli szóródás és az elemszám fogja befolyásolni. Összességében látható, hogy standard error értéke nemcsak a mintaátlagnak lehet, hanem más mintából számolt mutatónak is, többek között egy statisztikai modellel becsült hatás nagyságának is.

Ha a kapott hatást (a regressziós egyenes meredekségét vagy a két mintaátlag különbségét) elosztom a hozzá tartozó standard error értékével, akkor egy olyan arányszámot kapok, mely megadja, hogy a mintában látott hatás mennyire emelkedik ki a hozzá feltételezett bizonytalanság sávjából. Ezt az arányszámot nevezzük statisztikai értéknek. A statisztikai érték egy mértékegység nélküli szám (azaz már nem grammban mérjük), és ilyen érték a statisztikai próbákból kapott t-érték, F-érték vagy khi-négyzet-érték. Tehát ha a stressz-csokoládéfogyasztás példájában az ötvenfős mintán a B = 2,5, és az SE = 1,1, akkor t = 2,27; ha a férfiak és nők közötti eltérés ΔM = 0,4 és a két mintából számolt közös SE = 0,7, akkor t = 0,57 (ezt legközelebb le is ellenőrizheti az olvasó, amikor lineáris regressziót vagy t-próbát futtat). Hogy a lehetséges statisztikai értékek

közül melyiket kapom meg, leginkább az adatoktól és a számítás módjától függ, de minden statisztikai értékben közös, hogy jel és zaj hányadosából adódnak. Azonos a statisztikai értékekben az is, hogy mindegyikük egy-egy nevezetes eloszláshoz kötődik, és ennek megfelelően valószínűségek rendelhetők hozzájuk.

De minek a valószínűségeit jelölik a statisztikai értékek? Ezzel el is érkeztünk a fejezet első nagy témájához, a hagyományos frekventista szignifikanciatesztelés koncepciójához. Minek szeretnénk a valószínűségét megtudni? Leginkább annak, hogy mennyire valószínű, hogy az a hatás, amelyet a mintában is látunk, a stressz t = 2,27 nagyságú hatása a csokoládéfogyasztásra, mennyire valószínű, hogy a populációban is jelen van. A hagyományos statisztikai eljárások erre a kérdésre nem tudnak közvetlenül választ adni, a t-értékhez rendelt valószínűség *nem* ezt jelenti. Akkor fordítsuk meg, mondván, hogy megelégszünk azzal is, ha tudjuk, mekkora a tévedés valószínűsége akkor, ha a minta alapján összefüggést feltételezünk a populációban stressz és csokoládéfogyasztás között, vagy különbséget férfiak és nők fogyasztása között. Ezt nevezzük elsőfajú hibának (Type I error), annak a tévedésnek a valószínűségét, hogy valami olyan hatást találtunk a mintában, amely a valóságban, a populációban nem létezik (álpozitív eredmény). Közelebb kerültünk az igazsághoz, de a statisztikai értékekhez rendelhető valószínűségek még mindig nem pontosan ezt mérik.

Tegyük fel, hogy a populációban a csokoládéfogyasztás normál eloszlást követ, illetve azt is, hogy férfiak és nők fogyasztása között nincs különbség. Ha ebből a populációból veszünk egy férfi és női mintát, annak van a legnagyobb valószínűsége, hogy a mintában sem találunk a két nem között eltérést, t-értékben kifejezve ez a t = 0. Természetesen a véletlen mintavételezés miatt annak is még viszonylag nagy valószínűsége van, hogy a két minta között találunk valamekkora eltérést (pozitív t-értéket, ha az egyik nem fogyasztása nagyobb, negatívat, ha a másiké). Az egyre nagyobb eltérésekhez egyre kisebb valószínűségek tartoznak. Elméleti szinten mindez az úgynevezett t-eloszlás alakot hozza létre. A t-eloszlás alakját a minta elemszáma befolyásolja, kisebb minták esetében a szélsőségesebb értékekhez még viszonylag nagyobb valószínűség tartozik, és az elemszám növekedésével ez mérséklődik, és a t-eloszlás közelíti a normál eloszlást. A t-eloszlás folytonos eloszlás, tehát nem adott pontokhoz kerül valószínűség hozzárendelésre, hanem mindig tartományok valószínűségével tudunk számolni. Legtöbbször azt keressük, hogy milyen valószínűség tartozik az adott t-értéknél nagyobb tartományhoz (vagy negatív t-értéknél egy adott t-értéknél kisebb tartományhoz).

Tehát a t-eloszlásból arra tudunk következtetni, hogy mekkora egy minimum adott mértékű hatás véletlen mintavételezésből adódó előfordulásának valószínűsége abban az esetben, ha a populációban nincs jelen a hatás. Ehhez

az eloszláshoz hasonlítjuk a mi mintánkban látott hatás nagyságát. Például ha ötvenfős a mintánk, akkor a minimum t = 0,57 nagyságú hatás előfordulási valószínűsége p = 0,286, azaz ha a populációban nincs férfiak és nők között eltérés a csokifogyasztásban, 28,6% valószínűsége van annak, hogy egy ekkora különbség, amelyet mi látunk a mintánkban, mégis megjelenik a mintavételezés során. Mivel ez a valószínűség elég nagy, így nem vetjük el a nullhipotézist, miszerint nincs eltérés férfiak és nők között. Ha a valószínűség kicsi lenne, akkor elvetnénk a nullhipotézist, és elfogadnánk az alternatív hipotézist, hogy férfiak és nők csokoládéfogyasztása eltérő. Ronald A. Fisher, a szignifikanciatesztelés egyik megalkotója azt javasolta, hogy egyfajta közös konvencióként 5%-os (vagy p = 0,05) valószínűséget tekinthetünk egy olyan küszöbnek, amely alatt az eredmény már statisztikailag szignifikáns. Ezt a küszöbértéket, amelyet a kutató még az elemzés előtt meghatároz, szignifikanciaszintnek nevezzük, és α-val (alfa) jelöljük. Amikor a mintából számolt p-érték kisebb az előre meghatározott α-nál (például $p < 0,05$), akkor elvetjük a nullhipotézist. A p-érték itt kötődik vissza az elsőfajú hibához, mert ezzel a döntési szabállyal hosszú távon legfeljebb 5% eséllyel követünk el elsőfajú hibát, vagyis vetünk el tévesen egy igaz nullhipotézist.

Mindennek számos elméleti és gyakorlati hozadéka van. A legfontosabb, hogy a hagyományos frekventista hipotézistesztelés nagymértékben elméletvezérelt. Az elmélet alapján határozzuk meg a statisztikai modellünket, választjuk ki a prediktor változókat és az összefüggést leíró függvényt. Elméleti megfontolás alapján választunk az adatokhoz megfelelő nevezetes eloszlást, mely alapján a kiszámolt hatáshoz valószínűséget rendelünk, és elméleti megfontolás alapján határozzuk meg azt a kritériumszintet, amely alatt egy eredményt szignifikánsnak tekintünk. Ugyan ennek legtöbbször nem vagyunk tudatában, hiszen mi, felhasználók a felszínen csak annyiról döntünk, hogy t-próbát választunk férfiak és nők csokoládéfogyasztásának összehasonlítására, de ez az egyetlen döntés számos elméleti megfontolást és feltételezést rejt magában. A frekventista hipotézistesztelés mélyen és eszszenciálisan elméletalapú. Feltételezésekkel él többek között a populációra, az elemek függetlenségére, a kiválasztott modellre, a reziduális hibák függetlenségére, normalitására és homogenitására. Észben kell azonban tartanunk, hogy az elméleti feltételezésekre való ilyen nagy fokú támaszkodásra elsősorban azért volt szükség, mert a módszertan megalkotásakor még nem állt a kutatók rendelkezésére majdnem korlátlan számítási kapacitás, mely segítségével a modernebb eljárások (többek között a bootstrap és a keresztvalidáció) működnek, ahol az eredmények megbízhatóságát a helyes elméleti döntések helyett részben a nyers számítási kapacitás biztosítja.

A másik következmény, hogy amikor a $p < 0,05$-öt olyan válaszvonalnak tekintjük, amely feketén-fehéren elválasztja a szignifikáns és nem szignifikáns eredményeket egymástól, akkor összemossuk azon eredmények megbízhatóságát, melyek éppen a határérték szignifikáns oldalára estek, azokétól, amelyeknél az elsőfajú hiba valószínűsége egyértelműen és megbízhatóan alacsony. A tisztán p-értékre való támaszkodás nem veszi figyelembe a hatás nagyságát. Emellett teret enged a p-hacking jelenségének. A kutatót csábíthatja a gondolat, hogy addig növelje az elemszámot, amíg a számára fontos eredmények szignifikánssá válnak. Hogy úgy szűrje a szélsőséges értékeket vagy alakítsa a modellt, például vegyen ki vagy adjon hozzá változókat, szűrjön almintákra, hogy a számára fontos eredmények szignifikánsakká váljanak. Összességében ez a szignifikanciavadász gyakorlat a humán kutatások elemzéseinek kellemetlenül nagy fokú megismételhetetlenségéhez vezetett a múltban.

A szignifikanciateszteléssel kapcsolatos problémákra és a megismételhetőségi válságra válaszul számos tudományos szervezet és folyóirat vezetett be publikációs reformokat a kutatási eredmények átláthatóbbá tétele érdekében. Ennek egyik legfontosabb eleme, hogy a nagy presztízsű irányelvek, mint például az Amerikai Pszichológiai Társaság (APA) kiadványi kézikönyve, ma már megkövetelik a $p < 0,05$-ös küszöbérték mechanikus használata helyett a pontos p-érték közlését (jellemzően két vagy három tizedesjegyig, például $p = 0,042$), ami sokkal több információt hordoz a statisztikai bizonyíték erejéről. Emellett szinte kötelezővé tették a hatásnagyság-mutatók (mint például a Cohen-féle d vagy a korrelációs együttható) közlését.

A p-érték-alapú döntéshozatal problémáinak enyhítésére egy másik fontos javaslat a konfidenciaintervallumok (CI) használatának előtérbe helyezése volt. A konfidenciaintervallum lényegében egy becsült megbízhatósági tartomány, melyet úgy kapunk, hogy a hatás standard errorját (SE) megszorozzuk egy, a kívánt valószínűségi szinthez (jellemzően 95%-hoz) tartozó elméleti értékkel (például t-értékkel). Ez a tartomány kiszámítható egy egyszerű átlag köré, de ami még fontosabb, a statisztikai modellekben becsült hatások, például egy regressziós együttható (B) köré is. Az értelmezés logikája rendkívül szemléletes. Ha egy hatás értéke nulla ($B = 0$), az azt jelenti, hogy a prediktor nincs hatással a kimeneti változóra. A konfidenciaintervallum egy megbízhatósági sávot von a becsült hatás köré, és az értelmezés kulcsa az, hogy ez a sáv tartalmazza-e a nulla értéket. Ha a CI-ben benne van a nulla, az azt jelenti, hogy bár a mintánkban találtunk valamekkora hatást, a populációra nézve a lehetséges valós hatásértékek között ott van a „nincs hatás” ($B = 0$) eshetősége is. Ezzel szemben minél távolabb esik a teljes intervallum a nullától, és minél szűkebb, annál biztosabbak lehetünk a hatás valódiságában és pontosságában.

Bár ez a reform elméletben sokat segített volna az eredmények árnyaltabb értelmezésében, a gyakorlati megvalósulása kevésbé volt sikeres. A legtöbb kutató ugyan közli a konfidenciaintervallumokat, de ez gyakran csak formaiság, mivel a publikáció így professzionálisabbnak tűnik. Az eredmények érdemi értelmezése során azonban továbbra is szinte kizárólag a szignifikanciára – és szerencsés esetben a hatásnagyságra – támaszkodnak, így a konfidenciaintervallum egyfajta kötelező, de kevéssé értelmezett dekorációs elemmé vált a tanulmányokban. Emellett a CI továbbra is jelentős mértékben elméletalapú, szinte minden szempontból ugyanazokra az elméleti feltételezésekre építkezik, amelyekre a p-érték is, és csupán az értelmezést teszi árnyaltabbá.

A frekventista statisztika elméleti alternatívájaként régóta létezik a bayesi megközelítés, amely gyökeresen más kérdést tesz fel. Ahelyett, hogy azt vizsgálná, mekkora a mért adatok valószínűsége egy feltételezett nullhipotézis mellett, a bayesi statisztika azt a kérdést válaszolja meg, amelyet a kutatók valójában tudni szeretnének: a begyűjtött adatok alapján mekkora a valószínűsége annak, hogy a hipotézisük igaz? A módszer lényege, hogy a kutató egy előzetes feltevéssel (prior) indul, amely számszerűsíti a hipotézis valószínűségét még az adatok ismerete előtt. Az adatok (a bizonyíték) beérkezése után ez az előzetes feltevés frissül, és egy utólagos valószínűséget (posterior) kapunk, amely már tükrözi a hipotézis plauzibilitását az új információk fényében. Bár ez a megközelítés elméletben sokkal intuitívabb választ ad az eredmények megbízhatóságára, elterjedése a gyakorlatban mégis korlátozott. Ennek oka egyrészt a számítási és elméleti összetettsége, másrészt a legfőbb kritika éppen a módszer központi elemére, a prior meghatározására irányul, amelynek kiválasztása gyakran szubjektív. Emiatt a bayesi statisztika használata viszonylag szűk szakértői körre korlátozódik, és az átlagos kutatók ritkán alkalmazzák. Fontos látni, hogy ez a módszertan sem mentes az elméleti feltételezésektől, sőt a prior eloszlás megadásának szükségessége miatt gyakran még több és komplexebb elméleti döntést követel meg a kutatótól, mint a hagyományos eljárások.

3. Bootstrap

Az első valódi elmozdulás a modellek elméleti feltételezéseken alapuló értékeléséről természetesen csak akkor jöhetett el, amikor az adatalapú, empirikus ellenőrzéshez szükséges számítási kapacitás elérhetővé vált a számítógépek fejlődésével. Bár az igazi áttörést a bootstrap hozta el, a számításigényes, újramintavételezésen alapuló statisztika alapjait már korábban lefektették. Az első ilyen próbálkozások egyike a John Tukey által az 1950-es években kidolgozott Jackknife-módszer volt. Ennek logikája, hogy szisztematikusan, egyenként kihagyja a minta elemeit, és az így kapott csökkentett mintákon újraszámolja a vizsgált mutatót, majd ezek eloszlásából becsüli a statisztika bizonytalanságát. Hasonlóan régi, R. A. Fisher nevéhez fűződő elgondolás a permutációs teszt, amelynek elve már az 1930-as években megszületett. Itt a nullhipotézist (például két csoport között nincs különbség) úgy teszteljük, hogy a csoportcímkéket véletlenszerűen összekeverjük, és a kapott empirikus eloszlásban vizsgáljuk, hogy az eredetileg mért különbség mennyire számít szélsőségesnek. Bár az elv zseniális volt, gyakorlati alkalmazása a számítógépek elterjedéséig szinte lehetetlen volt. Az első, igazán átütő sikert és széles körben alkalmazható megoldást azonban a bootstrap hozta el. A módszert Bradley Efron publikálta 1979-ben, de igazi elterjedése jóval később, nagyjából a 2010-es évektől számítható, amikor a nagy kereskedelmi szoftverek (mint az SPSS vagy a Stata) beépítették a bootstrap eljárásokat a menüvezérelt felületeikbe, így téve azt elérhetővé az átlagos, nem programozói szintű kutatók számára is.

A hagyományos statisztikai eljárások, mint a szignifikanciatesztelés, tehát azzal az alapvető problémával küzdenek, hogy a becsléseink bizonytalanságának meghatározásához kénytelenek vagyunk elméleti feltételezésekkel élni az adatok populációbeli eloszlásáról. A bootstrap egy forradalmi, számítógépes megközelítés, amely erre a problémára ad elegáns választ. Ahelyett, hogy külső elméletekre támaszkodna, azt mondja, hogy feltételezzünk annyira keveset, amennyire csak lehet, és inkább kérdezzük meg magát a mintát, hogyan viselkedne, ha a mérést újra és újra megismételnénk.

A bootstrap központi analógiája zseniálisan egyszerű. Mivel a valós populációból nem tudunk több ezer új mintát venni, hogy feltérképezzük a mintavételi bizonytalanságot, tegyünk úgy, mintha az az egyetlen rendelkezésre álló mintánk lenne a teljes populáció. Ez a minta a legjobb információnk arról, hogyan is néz ki a valóság, ezért kezeljük a populáció miniatűr, de hű tükörképének. Ebből

a minipopulációból már annyi új mintát vehetünk, amennyit csak a számítógépünk kapacitása enged. Ezt a folyamatot nevezzük újra-mintavételezésnek (resampling). A módszer kulcsa és egyben trükkje a visszatevéses mintavétel.

A módszer a gyakorlatban egy jól definiált, számítógépes algoritmus szerint működik. Az eredeti mintával kezd, tehát az eredeti N elemű mintán (például 50 ember heti csokifogyasztása), és kiszámolja a minket érdeklő mutatót (például a mintaátlagot). Ezt követően létrehozza az első bootstrap mintát. Véletlenszerűen húz N elemet az eredeti mintából, de visszatevéssel. Ez azt jelenti, hogy miután egy adatpontot kiválasztott, azt „visszateszi a kalapba", így az újra kiválasztható válik. Ennek eredményeként egy-egy eredeti adatpont az új mintában szerepelhet többször is, míg mások ideiglenesen kimaradhatnak. De miért pont visszatevéssel? Ez nemcsak egy furcsa technikai megoldás, hanem a valós mintavételezés szimulációja. A populációban sem csupán egyetlen 65 grammot fogyasztó személy van, hanem rengeteg. Ha a valódi populációból vennénk egy újabb mintát, abban is könnyen szerepelhetne több, azonos értékkel rendelkező egyén. A visszatevés ezt a logikát képezi le a mi minipopulációnkra. Az is felmerülhet kérdésként, hogy miért nem generálunk egyszerűen véletlen számokat. Azért, mert az eredeti minta hordozza a kulcsfontosságú információt az egyes értékek előfordulási arányairól. A mintánk alapján a 65 grammos fogyasztás jóval gyakoribb, mint a 150 grammos. Amikor a saját mintánkból húzunk, automatikusan tiszteletben tartjuk ezeket az arányokat, és sokkal nagyobb valószínűséggel fogunk olyan bootstrap mintát generálni, amelyben többször szerepel a 65-ös érték, mint olyat, amelyben a 150-es dominál. A minta tehát nemcsak az értékek tartományát, hanem azok sűrűségét is megadja. Ezen az új, frissen generált bootstrap mintán is kiszámoljuk ugyanazt a mutatót (például az átlagot). Ez az érték valószínűleg kissé el fog térni az eredeti minta átlagától.

És most jön az ismétlés, ismétlés, ismétlés, ami a tudás anyja. A mintavételezést és a bootstrap mintán való paraméter-számítást sokszor, jellemzően legalább ezerszer, de gyakran inkább ötezerszer megismételjük. A folyamat végén kapunk több ezer becsült értéket (például mintaátlagot). Ezt a rengeteg, pusztán az adatokból generált értéket nevezzük bootstrap eloszlásnak. Ebből az empirikus eloszlásból már mindenféle elméleti feltételezés nélkül, közvetlenül leolvashatjuk a minket érdeklő bizonytalansági mutatókat.

A bootstrap standard error nem más, mint a bootstrap eloszlás, azaz az ötezer bootstrap mintaátlag szórása. A 95%-os konfidenciaintervallumot pedig az eloszlás 2,5. és 97,5. percentilise közötti tartomány adja (ugyan létezik más számítás is, ez a legegyszerűbb, percentilisalapú bootstrap CI). A kapott konfidenciaintervallum nem feltétlenül lesz szimmetrikus. Míg a hagyományos, képleten alapuló intervallumok a pontbecslés ± hibahatár logika miatt mindig

szimmetrikusak, a bootstrap ezt a kényszert elengedi. A bootstrap eloszlás hűen tükrözi az eredeti minta sajátosságait, beleértve annak ferdeségét is. Ha az eredeti adataink nem szimmetrikusak – gondoljunk csak a Gombóc Artúr-féle kiugró értékre, amely az eloszlást jobbra elhúzza –, akkor az abból generált több ezer bootstrap mintaátlag eloszlása is ferde lesz. Amikor ebből a ferde eloszlásból levágjuk az alsó és felső 2,5%-ot, hogy megkapjuk a köztes 95%-os konfidenciaintervallumot, a kapott intervallum is aszimmetrikus lesz. A Gombóc Artúr-példánál maradva a CI felső határa valószínűleg jóval messzebb lesz a becsült átlagtól, mint az alsó határa. Ez az aszimmetria nem hiba, hanem a módszer egyik legnagyobb erőssége, ami sokkal őszintébb és realisztikusabb képet ad a becslésünk bizonytalanságáról.

Nézzünk egy példát egy húszfős (N = 20) mintával. Az eredeti mintánk (N = 20) heti csokifogyasztása grammban: [40, 45, 50, 50, 50, 55, 55, 55, 60, 60, 60, 60, 65, 65, 70, 70, 80, 90, 100, 150]. Láthatjuk, hogy a legtöbb ember 50–70 gramm között fogyaszt, vagyis közel helyezkedik el a korábban említett, populációra jellemző 65 gramm/hét fogyasztáshoz, de az eloszlás, ahogy az a valóságban is gyakori, nem tökéletesen szimmetrikus, illetve van egy kiugró, 150 grammos érték is. Ennek hatására a minta átlaga (66,5 g) magasabb, mint a mediánja (60 g), ami a jobbra elhúzódó, ferde eloszlás tipikus jele. Lássunk három lehetséges bootstrap mintát:

1. bootstrap minta (Gombóc Artúr kimaradt): [60, 50, 70, 40, 55, 65, 60, 50, 80, 55, 45, 60, 65, 70, 90, 55, 60, 50, 40, 70], az átlag M = 59,5 g.
2. bootstrap minta (a kiugró érték kétszer bekerült): [55, 150, 60, 45, 100, 65, 50, 150, 40, 55, 60, 80, 60, 70, 70, 50, 90, 65, 55, 50], az átlag M = 71 g.
3. bootstrap minta (a kiugró érték egyszer került be): [70, 60, 40, 55, 80, 100, 65, 50, 60, 60, 45, 55, 150, 65, 50, 90, 60, 55, 70, 50], az átlag M = 66,5 g.

Ha ezt a folyamatot 5000-szer elvégezzük, az átlagok eloszlása hűen fogja tükrözni az eredeti minta ferdeségét.

A csokoládéfogyasztás átlagának példája jól szemlélteti a módszer működését, de a bootstrap igazi ereje messze túlmutat az egyszerű középértékeken. A bemutatott újra-mintavételezési logika ugyanis szinte bármilyen, a mintából számolt statisztikai mutatóra alkalmazható. Kiszámíthatjuk vele nemcsak az átlag, hanem a medián, a szórás, a ferdeség vagy két változó korrelációjának konfidenciaintervallumát is.

A módszer ereje igazán a komplexebb modelleknél mutatkozik meg. Ha például egy regressziós modell együtthatójának (B) bizonytalanságára vagyunk kíváncsiak, a bootstrap algoritmus mind az 5000 újra-mintavételezett

adathalmazon újra és újra lefuttatja a teljes regressziós elemzést, és az így kapott 5000 regressziós együtthatóból építi fel a bootstrap eloszlást. A legfontosabb, hogy a bootstrap olyan mutatóknak is képes megbízhatósági intervallumot számítani, amelyeknek a hagyományos statisztikában nincs is egyszerű, képlettel leírható standard errorja. Így bármilyen hatásnagyság-mutatónak, mint például a standardizált bétaegyütthatónak vagy akár egy mediánok közötti különbségnek is meghatározhatjuk a bootstrapalapú konfidenciaintervallumát. Ez a sokoldalúság felszabadítja a kutatót az elméleti képletek korlátai alól, és lehetővé teszi, hogy szinte bármilyen, számára fontos mutató megbízhatóságát empirikusan, közvetlenül az adatokból becsülje meg.

A módszer sokoldalúsága mellett a másik, talán leggyakrabban emlegetett előnye a robusztussága. Míg a hagyományos statisztikai eljárások parametrikusak, azaz szigorú feltételezésekkel élnek a populáció eloszlására vonatkozóan – leggyakrabban a normális eloszlást feltételezik –, a bootstrap mentes ezektől a kényszerektől. Nem követeli meg az adatoktól, hogy egy adott elméleti eloszlást kövessenek, mivel az eloszlást empirikusan, magukból az adatokból építi fel. Éppen ezért a bootstrap kifejezetten hasznos, amikor az adataink ferde eloszlást mutatnak. Ahelyett, hogy egy ferde eloszlásra egy szimmetrikus, elméleti konfidenciaintervallumot erőltetne, a bootstrap az eredeti minta ferdeségét tükrözve egy aszimmetrikus intervallumot hoz létre, ami sokkal valósághűbben ábrázolja a becslés bizonytalanságát. Hasonlóképpen, a kiugró értékek jelenléte sem okoz akkora problémát. A visszatevéses mintavételezés során a kiugró érték hol kimarad, hol pedig akár többször is bekerül a bootstrap mintákba, így a módszer közvetlenül modellezi annak destabilizáló hatását a becslésre. Az eredmény egy szélesebb, de őszintébb konfidenciaintervallum, amely hűen tükrözi a kiugró érték által okozott megnövekedett bizonytalanságot.

A bootstrap talán legprovokatívabb alkalmazása egy izgalmas gondolatkísérlet: mi történne, ha magát a p-értéket bootstrappelnénk? Fontos kiemelni, hogy ez nem egy tényleges módszertan, mert ez a lépés visszahozná az elméleti feltételezéseket (hiszen a p-érték számítása ezeken alapul) egy olyan módszerbe, amely épp ezek elkerülésére született, hanem csupán egy izgalmas szemléltetés, mely hihetetlenül tanulságos szemléletformálásként szolgálhatna. Képzeljük el a folyamatot! Az eredeti mintánkon lefuttatjuk a statisztikai tesztet és megkapjuk a p-értéket. Ezt követően létrehozunk több ezer bootstrap mintát, és mindegyiken újra és újra lefuttatjuk ugyanazt a teljes statisztikai tesztet, minden alkalommal feljegyezve a kapott p-értéket. A végeredmény nem egyetlen szám, hanem egy teljes eloszlás, egy hisztogram a lehetséges p-értékekről.

Ez a hisztogram lenne az igazi szemfelnyitó élmény a kutató számára. Egy $p = 0{,}04$-es, éppen szignifikáns eredmény esetében jó eséllyel azt látnánk, hogy

a p-értékek eloszlása kényelmesen átnyúlik a 0,05-ös határon, és a lehetséges kimenetelek jelentős része (mondjuk 40%-a) valójában nem is szignifikáns. Ez rávilágítana, hogy a „bizonyított" eredményünk mennyire törékeny, és a mintavételi véletlen mennyire befolyásolta azt. Arra tanít, hogy a p-értéket ne egy fix, kőbe vésett igazságnak tekintsük, hanem egy bizonytalan, a mintavételi ingadozásnak erősen kitett mutatónak, ami segít felülkerekedni a szignifikáns/nem szignifikáns kategóriák merev, fekete-fehér logikáján.

Bár a bootstrap robusztussága és sokoldalúsága rendkívül vonzó, fontos megérteni, hogy a módszer nem egy univerzális csodaszer. A legfontosabb korlátját a „Szemét be, szemét ki" (Garbage In, Garbage Out) elve foglalja össze. A bootstrap teljes mértékben az eredeti mintára támaszkodik, és nem képes olyan információt teremteni, amely abban nincs eleve jelen. Ha az eredeti mintavételünk hibás, például torzított vagy nem reprezentatív, a bootstrap csupán ezt a torzítást fogja felerősíteni és reprodukálni az ezernyi újra-mintavételezés során. Éppen ezért téves az a gyakori feltételezés, hogy a bootstrap megoldást kínál a kis minták problémájára. Valójában egy nagyon alacsony elemszámú minta esetén a bootstrap kifejezetten félrevezető lehet, mert a kis minta a populációnak csupán egy alacsony felbontású, elnagyolt képe. Valószínűtlen, hogy egy ilyen kis adathalmaz képes megragadni a populáció valódi változatosságát, az értékek teljes skáláját és az eloszlás valódi alakját. Ha a kis minta véletlenül nem tartalmaz szélsőséges értékeket, a bootstrap eljárás soha nem fogja azokat legenerálni. A végeredmény egy mesterségesen keskeny bootstrap eloszlás lesz, ami egy megtévesztően kicsi standard errorhoz és egy irreálisan szűk konfidenciaintervallumhoz vezet. A kutató így tévesen azt a következtetést vonhatja le, hogy a becslése nagyon pontos, miközben ez a látszólagos precizitás csupán a kiindulási minta szegényességének az eredménye.

Összefoglalva: a bootstrap egy rendkívül erőteljes és intuitív eszköz, amely a modern számítási kapacitást használja fel a klasszikus statisztika egyik legnehezebb problémájának megkerülésére. Helyén használva sokkal megbízhatóbb és őszintébb képet ad az eredményeink bizonytalanságáról.

4. Keresztvalidáció

A bootstrap valóban hatalmas módszertani váltást hozott azzal, hogy elengedte a merev elméleti feltételezéseket, de a célkitűzéseit tekintve mégsem szakadt el teljesen a korábbi paradigmától. Tulajdonképpen minden eddig tárgyalt eljárás ugyanazon központi kérdés körül forgott: hogyan tudunk egy intervallumot (legyen az standard error vagy konfidenciaintervallum) rendelni a pontbecslésünk köré, amely megadja annak potenciális variabilitását. Ez a megközelítés a statisztikai inferencia világa, ahol a fő cél egy mintából becsült paraméter tulajdonságainak minél pontosabb leírása.

A következő módszertan, a keresztvalidáció ezzel egy határozott paradigmaváltással szakít. Célja már nem az, hogy a becsült érték potenciális eloszlásának egy újabb, még pontosabb leírását adja, hanem azt a rendkívül gyakorlatias kérdést kutatja, hogy a mintánkon létrehozott modellel mennyire vagyunk képesek egy teljesen új értéket helyesen bejósolni. Ezzel a hangsúly a becslés tulajdonságainak leírásáról (inferencia) átkerül a modell gyakorlati teljesítményére, azaz a predikcióra.

Inferencia (eddigi módszerek): a becslés tulajdonságainak leírása. Mennyire bizonytalan a stressz és a csokifogyasztás közötti kapcsolat erősségének becslése?

Predikció (keresztvalidáció): a modell bejósló képességének tesztelése. Ha jön egy új ember, akinek tudom a stressz-szintjét, a modellem milyen pontosan fogja megtippelni a csokifogyasztását?

A keresztvalidáció egy olyan eljárás, amely legszorosabban a gépi tanulás (machine learning) világához köthető. A gépi tanulás lényege, hogy ahelyett, hogy egy előre definiált modellt tesztelnénk, hatalmas adatmennyiséget adunk egy algoritmusnak, amely abból önállóan tanulja meg a mintázatokat, és építi fel a lehető legjobb modellt egy adott feladatra, például a csokoládéfogyasztás bejóslására. Ezek a modellek gyakran rendkívül komplexek, és sokszor úgynevezett „fekete dobozként" (black box) működnek, tehát látjuk, milyen adatok mennek be és milyen jóslat jön ki, de a belső működésük és logikájuk nehezen vagy egyáltalán nem értelmezhető.

Ha egy ilyen modellt kizárólag azon az adathalmazon értékelnénk, amelyen tanult, könnyen abba a hibába esnénk, hogy egy látszólag tökéletes, de a valóságban használhatatlan modellt alkotunk. Ez a túlillesztés (overfitting) jelensége. A túlillesztés során a modell ahelyett, hogy általánosítható mintázatokat

tanulna meg, gyakorlatilag bemagolja a tanítóadatok zaját és egyedi furcsaságait. Az ilyen modell lenyűgöző pontosságot mutat a már ismert tanítóadatokon, de csúfosan elbukik, amint új, eddig sosem látott adatokkal találkozik. Ez akár odáig is fajulhat, hogy teljesítménye rosszabb lesz, mint az egyszerű, naiv alapmodellé, például regressziónál negatív R^2-értéket produkálhat a tesztadatokon. A regressziós modelleknél az R^2 azt méri, mennyivel jobb a modellünk jóslata egy egyszerű alapmodellnél, ami nem veszi figyelembe a prediktorokat, hanem minden esetre csak a kimeneti változó átlagát tippeli. A túlillesztés során a modell nemcsak a valódi mintázatot, hanem a tanítóadatok véletlenszerű zaját is megtanulja. Amikor ezt a túlságosan specifikus modellt egy új, független tesztadathalmazon próbáljuk ki, a jóslatai annyira pontatlanok lehetnek, hogy a modell egyenese akár az adatok trendjével ellentétes irányt is felvehet, vagyis rosszabbul illeszkedne az új adatokra, mint a primitív, vízszintes átlagvonal. Mivel az R^2 az alapmodellhez képesti javulást méri, ha a modellünk valójában ront a helyzeten, az értéke negatívvá válik. A negatív R^2 tehát egy vészjelzés, azt mutatja, hogy a modell túl van illesztve, és használhatatlan új adatok bejóslására.

A túlillesztés elkerülésének alapvető megoldása, hogy a modell teljesítményét egy olyan független adatcsomagon kell tesztelni, amelyet a modell a tanulás során még soha nem látott. Ennek a logikának a gyerekcipőben járó változatával már a hagyományos statisztikában is találkozhattunk, például amikor egy feltáró faktoranalízis eredményeit egy megerősítő faktoranalízissel ellenőrizzük. Sőt, a legtisztább párhuzamot az a pszichometriában ajánlott gyakorlat mutatja, amikor egy CFA segítségével véglegesített kérdőívmodellt egy teljesen új, független mintán ismételt CFA-val validálnak (bár ez a gyakorlat a valóságban sajnos ritkán valósul meg). Ez a „fejlesztés az egyik mintán, ellenőrzés a másikon" elv tökéletesen megegyezik a gépi tanulás tanító és tesztelő adathalmazokra bontásának alapkoncepciójával.

Ahogy láttuk, a modell valós teljesítményének mérésére a legegyszerűbb megoldás, ha a rendelkezésre álló mintánkat két részre osztjuk: egy nagyobb tanító (training) és egy kisebb tesztelő (vagy validációs) adathalmazra. A gépi tanulás korai szakaszaiban, nagyjából a 2000-es évek közepéig ténylegesen ez volt a fő elemzési gyakorlat, tehát a kutatók az adathalmazukat véletlenszerűen kettéosztották, az egyik felén tanítottak, a másikon pedig egyetlen egyszer teszteltek. Gyakori volt az a gyakorlat is, amely a mintát három részre osztja: egy tanító, egy tesztelő és egy végső, teljesen érintetlen validációs adathalmazra. Akár két, akár három részre bontottak, a cél az volt, hogy a tanuláshoz a lehető legtöbbet használják fel, de az ellenőrzésre is maradjon egy elegendően nagy és reprezentatív minta. A gyakorlatban ez a két részre osztásnál a leggyakrabban 70-30% vagy 80-20% körüli arányt jelent.

Ennek a megközelítésnek azonban van egy nyilvánvaló és fájó hátránya. A modell végleges változatának felépítéséhez az adatainknak csupán egy részét, például 70%-át használtuk fel. A maradék 30%-ot, bár értékes információkat tartalmaz, lényegében „kidobtuk" a tanulási folyamatból, és kizárólag tesztelésre használtuk. Ez a veszteség közvetlenül a modell pontosságának kárára mehet, hiszen egy olyan modellt kapunk, amely kevesebb adatból tanult, mint amennyi a rendelkezésünkre állt.

A tanító-tesztelő felosztás adatvesztésének problémájára a megoldást a keresztvalidáció (cross-validation) ötlete hozta el. Bár a koncepció alapjait már korábban lefektették, és a módszertan modern formáját is formalizálták az 1970-es években olyan statisztikusok, mint Seymour Geisser és M. Stone, a koncepció és a mindennapi gyakorlat között évtizedes szakadék tátong. A módszer ugyanis rendkívül számításigényes, hiszen a modellt többször is újra kell építeni. Emiatt a gyakorlatban még a 2000-es évek elején is csak elvétve használták. A széles körű elterjedésére egészen addig kellett várni, amíg a személyi számítógépek teljesítménye és a szoftveres támogatás el nem érte azt a szintet, ahol a többszöri modellezés már nem jelentett leküzdhetetlen akadályt.

A keretrendszer megalkotói azt ismerték fel, hogy amennyiben ez a számítási kapacitás már rendelkezésre áll, semmi akadálya annak, hogy a mintát ne csak egyszer, hanem szisztematikusan, többféleképpen is felosszuk. Így minden egyes adatpont sorra kerülhet a „vizsgázó", azaz a tesztelőadat szerepében. Vegyük például a korábbi 80-20%-os felosztást. Ahelyett, hogy egyszer félretennénk 20%-ot, ezt a folyamatot megismételjük ötször, minden alkalommal egy másik 20%-ot jelölve ki tesztadatnak. Ezt nevezzük 5-szörös (5-fold) keresztvalidációnak. A modellünket tehát ötször építjük fel, minden alkalommal az adatok különböző 80%-án tanítva, majd a maradék, addig sosem látott 20%-on tesztelve. A folyamat végére minden adatpont pontosan egyszer szerepelt a tesztelő, és négyszer a tanító adathalmazban, így a modell betanításához és a minőségének ellenőrzéséhez is a teljes mintát felhasználtuk. Értelemszerűen az öt, némileg eltérő adathalmazon tanított modell különböző teljesítménymutatókat fog produkálni a saját tesztadatain. A keresztvalidáció végső eredménye nem egyetlen szám, hanem ez az öt teljesítménymutató. Ezek átlaga adja a modell várható, általánosító képességének robusztus becslését, a szórásuk pedig arról ad felvilágosítást, hogy mennyire stabil a modell teljesítménye, azaz mennyire érzékeny a tanítóadatok összetételének kisebb változásaira.

Jogosan merül fel a kérdés, hogy ha ötször építettünk modellt, akkor a kapott teljesítménymutató melyikhez tartozik az öt közül. A meglepő válasz az, hogy egyikhez sem. Fontos megérteni, hogy a keresztvalidáció során létrehozott modellek csupán ideiglenes, eldobható eszközök. Egyetlen feladatuk, hogy

a teljesítmény becsléséhez szükséges próbákat elvégezzék. A folyamat végén, miután megkaptuk a megbízható teljesítménybecslést, ezeket az ideiglenes modelleket elfelejtjük, és egyetlen, végső modellt tanítunk be a teljes, 100%-os adathalmazon. A keresztvalidáció során kapott átlagos teljesítményérték tehát ennek a végső, minden adatot látott modellnek a várható, jövőbeli teljesítményére adja a legjobb becslésünket.

A keresztvalidáció feladata tehát nem a modell paramétereinek finomhangolása, hanem a prediktív erejének megbecslése. A folyamat végén a modellhez kapunk egy átlagos teljesítményértéket (például egy átlagos R^2-et) és annak szórását. Ez a fókusz alapvetően eltér a korábbi módszerekétől, és a különbség a mért bizonytalanság természetében rejlik. Míg az elméleti alapon számolt és a bootstrapalapú konfidenciaintervallum is a paraméter becslésének bizonytalanságát mérte, addig a keresztvalidáció a modell teljesítménye értékelésének bizonytalanságát adja meg. Az előbbi azt a kérdést válaszolja meg, hogy „Menynyire vagyok biztos a regressziós együtthatóm (B) értékében?", és egy plauzibilis tartományt ad magára a paraméterre. Ezzel szemben a keresztvalidáció arra a kérdésre felel, hogy „Mennyire vagyok biztos abban, hogy a modellem R^2 értéke valóban 0,75 lesz új adatokon?". A teljesítménymutatók szórása a különböző adatfelosztásokon megmutatja, mennyire stabil maga az értékelésünk. Ez a megkülönböztetés tökéletesen összefoglalja a paradigmaváltást, mivel már nem egy belső modellparaméter megbízhatóságát vizsgáljuk, hanem a modell egészének külső, gyakorlati teljesítményét és stabilitását.

A keresztvalidáció óriási előnye, hogy a korábbi, egyszerű felosztással szemben a minta 100%-át felhasználja mind a tanulási folyamatban, mind a modell várható teljesítményének becslésében. Ennek a pontosságnak az ára a megnövekedett számítási igény, hiszen a modellt többször is fel kell építeni, egy k-szoros validációnál k-szor az ideiglenes tesztekhez, majd egyszer utoljára (+1) a teljes adathalmazon.

A folyamat végén kapott értékelés rendkívül informatív. Ha nagy a különbség a keresztvalidáció során kapott átlagos teljesítmény és a végső, 100%-os adathalmazon tanított modell saját magán visszamért teljesítménye között, az a túlillesztés egyértelmű jele. A keresztvalidált érték mutatja a modell reális, várható teljesítményét, míg a modell saját tanítóadatain mért, ennél sokkal magasabb értéke egy túlzó, túlságosan optimista becslés. A nagy szakadék a kettő között tehát arra utal, hogy nem megfelelő modellt vagy beállításokat választottunk.

Felmerül a kérdés, hogy hány részre (k) érdemes osztani az adatainkat. A válasz egy gyakorlati kompromisszum: annyira, amennyit a számítási lehetőségeink megengednek. A leggyakoribb választás az 5-szörös vagy 10-szeres

keresztvalidáció. A lehetőségek felső határát a mintánk elemszáma (N) adja meg. Ezt a speciális esetet, amikor k = N, „Hagyj-ki-egyet" keresztvalidációnak (Leave-One-Out Cross-Validation, LOOCV) nevezzük. Ez a leginkább alapos választás, amennyiben a számítási kapacitásunk ezt lehetővé teszi. Ebben az esetben N darab modellt építünk, mindegyiket N-1 adaton tanítva, és mindegyikkel csupán azt az egyetlen, kihagyott adatpontot jósoljuk meg. Az N darab jóslás hibáinak átlaga adja a modell teljesítményének precíz becslését. Matematikai modellek is bizonyítják, hogy az így kapott validációs értékek rendkívül jól becslik a modell valós teljesítményét új, eddig ismeretlen adatokon.

A keresztvalidáció megbízhatóságának alapvető garanciája az, hogy az egyes modellezési ciklusok (a tanítás-tesztelés párok) semmilyen módon ne kommunikáljanak egymással. Más szóval: el kell kerülnünk az adatszivárgást (data leakage). De mit jelent ez a gyakorlatban? Azt, hogy a modell tanítása során a tesztadatoknak szigorúan érintetlennek és ismeretlennek kell maradniuk a hozzájuk tartozó tanulóadatokon végzett elemzések számára. Bármilyen olyan lépést, amely a teljes adathalmazból tanul meg valamilyen paramétert, tilos a keresztvalidációs ciklus előtt elvégezni.

Lássunk erre egy egyszerű példát! Tegyük fel, hogy szeretnénk egy változót az elemzés előtt standardizálni. A rossz, de csábítóan egyszerű megközelítés az lenne, ha a teljes, 100%-os adathalmazon határoznánk meg a standardizáláshoz szükséges átlagot és szórást, végeznénk el a standardizálást, és csak ezután indítanánk a keresztvalidációt. Ezzel a lépéssel a tesztadatokban lévő információ (például egy extrém magas érték jelenléte) már beszivárgott a tanítóadatok feldolgozásába, hiszen befolyásolta a közösen meghatározott átlagot és szórást. A tesztadatok így már nem tekinthetők teljesen függetlennek és ismeretlennek. A helyes eljárás az, hogy maga a standardizálás is a keresztvalidációs cikluson belül történik. Minden egyes tanuló-teszt páros esetében kizárólag az aktuális tanító adathalmazon határozzuk meg az átlagot és szórást és végezzük el a standardizálást, a tesztadatokon pedig szintén elvégezzük a standardizálást, de a tanulóadatok átlagát és szórását mint paramétert felhasználva. Ez a szigorú elkülönítés biztosítja, hogy a modell értékelése valóban egy olyan adathalmazon történik, amelynek semmilyen információja nem befolyásolta a modell felépítését, így a kapott teljesítménymutató a modell valós általánosítóképességét tükrözi.

Ha a keresztvalidáció ennyire robusztus és hatékony, jogosan merül fel a kérdés, hogy miért nem használjuk egyszerűen a hagyományos statisztikai eljárások helyett. A válasz az, hogy a keresztvalidáció koncepciója túlzottan összefonódott a gépi tanulás módszertanával, és a gépi tanulásnak más a funkciója, mint a hagyományos statisztikai elemzéseknek, a két módszertan alapvetően

más kérdésre ad választ. A gépi tanulás és a hagyományos statisztika nem egymás versenytársai, hanem különböző célt szolgáló, eltérő filozófiájú eszközök.

A hagyományos statisztika célja a megértés és a magyarázat (inferencia). A kutató nem azzal a feladattal indul, hogy egy tucat változóból létrehozza a lehető legjobb modellt. Ehelyett van egy konkrét kíváncsisága, egy elméletvezérelt kérdése a világról: „A stressz összefügg-e a csokoládéfogyasztással?" A statisztikai modell csupán egy eszköz ennek a specifikus hipotézisnek a tesztelésére. A kutatót nem feltétlenül érdekli, hogy a heti desszertre szánt költségkeret esetleg sokkal jobban bejósolja a csokifogyasztást, hanem ő a stressz elméleti szerepére kíváncsi. A célja, hogy a modell segítségével egy kicsit többet tudjon meg a világ egy őt érdeklő szeletének működéséről.

Ezzel szemben a gépi tanulás a bejóslásra (predikció) fókuszál. Itt a kiindulópont nem egy elméleti kérdés, hanem egy gyakorlati probléma: hogyan tudjuk a lehető legpontosabban megbecsülni egy jövőbeli esemény kimenetelét? A cél a leghatékonyabb modell megtalálása, amely a rendelkezésre álló adatokból a legjobb teljesítményt nyújtja új, ismeretlen eseteken. Nem számít, hogy a modell egy bonyolult „fekete doboz", a lényeg a gyakorlati pontossága.

A két gondolkodásmód közötti különbséggel egy kicsit más formában már a hagyományos statisztikán belül is találkozhattunk. A hierarchikus regresszió során a kutató elméleti alapon, előre meghatározott sorrendben lépteti be a változókat a modellbe, hogy egy specifikus hipotézist teszteljen. Ezzel szemben a stepwise regresszió egy ad hoc, adathajhász algoritmus, amely pusztán a statisztikai mutatók alapján, elméleti megfontolások nélkül próbálja megtalálni a leghatékonyabb prediktorkombinációt. Az előbbi az inferencia, az utóbbi pedig a predikció leegyszerűsített logikáját képviseli. A különbséget talán egy analógia világítja meg a legjobban. A hagyományos statisztikus olyan, mint egy közgazdász, aki a gazdasági válságok okait és a mögöttes elméleti mechanizmusokat akarja megérteni. Az adatelemző, aki gépi tanulást használ, pedig olyan, mint egy tőzsdei kereskedő, aki a leghatékonyabb modellt keresi a részvényárfolyamok holnapi mozgásának bejóslására anélkül, hogy feltétlenül értenie kellene a teljes világgazdaság mély elméleti összefüggéseit.

Az, hogy a keresztvalidációt szinte kizárólag a gépi tanulással azonosítjuk, elsősorban történelmi és kulturális okokra vezethető vissza, nem pedig módszertani kényszerre. A gépi tanulás világában a puszta túlléléshez volt rá szükség a túlillesztés elkerülése miatt, míg a hagyományos statisztika „jól elvolt" a saját, elméleti alapú mutatóival.

Felmerülhet a kérdés, hogy nem lehetne, hogy egy hagyományos, elméletvezérelt regressziós modellt, ahol a prediktorokat mi magunk elméleti alapon választottuk ki, egyszerűen keresztvalidációval értékeljünk.

A válasz az, hogy de, abszolút lehetséges! Ahelyett, hogy a modellt lefuttatnánk a teljes adathalmazon, és megnéznénk a végső R^2-értéket, elindíthatunk egy 10-szeres keresztvalidációt. A program tízszer újraépíti pontosan ugyanazt a regressziós modellt, és minden alkalommal kiszámolja az R^2-et a félretett tesztadatokon. A végeredmény nem egyetlen R^2, hanem tíz különböző érték, amelyek átlaga egy sokkal őszintébb, realisztikusabb becslést ad a modell valódi prediktív erejére, szórásuk pedig a modell stabilitását mutatja meg. Azonban ez a szemlélet még messze nem tekinthető a mindennapi kutatási gyakorlat bevett részének, sokkal inkább egy olyan iránynak, amerre a megbízhatóbb tudományos megismerés érdekében haladni érdemes. Érdekességként megemlítjük, hogy ez az integráció azért már elkezdődött bizonyos területeken, hiszen vannak olyan modern statisztikai eljárások, amelyek beépítették a keresztvalidációszerű gondolkodást. Kiváló példa erre a PLS-SEM-modellekben használt Q^2-predict érték kiszámítása, amely egy úgynevezett blindfolding eljáráson alapul, amely a keresztvalidációhoz közeli módszertan. A kutatók azonban legtöbbször nincsenek tisztában azzal, hogy ilyenkor keresztvalidációt alkalmaznak, mert ezeket a mechanizmusokat nem ezen a néven, hanem az adott módszertanhoz kötődő, specifikus elnevezésen ismerik.

Fontos azonban látni, hogy a keresztvalidáció felhasználásának mélysége a két megközelítésben gyökeresen eltér. Amikor egy hagyományos, elméletvezérelt modellt értékelünk, a keresztvalidáció egy korlátozottabb, passzívabb szerepet tölt be. A kutató a legfontosabb döntéseket, mint az adattisztítást, a változók kiválasztását, a modell típusát, már a folyamat előtt meghozza. A keresztvalidáció itt csupán egy végső minőség-ellenőrzés egy már kész, előre specifikált modellen. Ezzel szemben a gépi tanulásban a keresztvalidáció a teljes modellépítési és -kiválasztási folyamat központi, aktív eleme. A cikluson belül versenyeztetünk többféle modellt, itt hangoljuk a paramétereiket, sőt a helyes eljárás szerint még az olyan adat-előkészítési lépések is, mint az adatok skálázása, minden tanulóadaton külön-külön történik az adatszivárgás elkerülése végett.

Ez a hibrid megközelítés egyfajta hidat képez a két világ között. Lehetővé teszi, hogy a kutató továbbra is a saját, elméletvezérelt modelljét tesztelje (inferencia), de annak értékeléséhez már a gépi tanulásból ismert robusztus, empirikus validációs eljárást használja (predikció). Ez egy végső valóság-ellenőrzés, mielőtt elkezdenénk a p-értékeket és az együtthatókat értelmezni, meggyőződhetünk róla, hogy a modellünk egyáltalán rendelkezik-e bármiféle általánosítható prediktív erővel. Ez a szemlélet rámutat, hogy a két világ nem elválasztva, hanem egymást kiegészítve tudja a leghatékonyabban szolgálni a tudományos megismerést.

5. Összefoglaló

Az elemzések megbízhatóságának értékelésében a tárgyalt módszereknek eltérő, egymást kiegészítő szerepük van, amit az elemzés három szintjén érdemes áttekinteni. Hasonlítsuk össze a hagyományos konfidenciaintervallumokat, a szignifikanciatesztelést, a bootstrap eljárást, valamint a keresztvalidációt a minta leíró mutatóinak szintjén, a modell paramétereinek szintjén és a teljes modell értékelésének szintjén.

A minta leíró mutatóinak szintjén, például a mintaátlagnál, a hagyományos konfidenciaintervallum egy elméleti feltételezéseken alapuló becslést ad arra, mekkora lenne a mintaátlag variabilitása többszörös mintavétel esetén. A bootstrap ugyanezt a variabilitást számolja ki, de elméleti feltételezések nélkül, pusztán adatalapú, empirikus úton. Ezen a szinten a szignifikanciatesztelésnek és a keresztvalidációnak nincs értelmezhető szerepe.

A modell paramétereinek szintjén, például egy regressziós együtthatónál, a helyzet már árnyaltabb. Az előzőhöz hasonlóan itt is kiszámítható a paraméter variabilitása, bizonytalansága mind elméleti, mind bootstrap alapon. Ezt a bizonytalanságot figyelembe véve pedig már a paraméter által leírt hatás általánosíthatóságáról is kapunk információt. A szignifikanciatesztelés elméleti feltételek segítségével teszteli, hogy a hatás kellően elkülönül-e a nullától, míg a bootstrap ugyanezt adatalapúan, a konfidenciaintervallum nulltartalmának vizsgálatával teszi. A keresztvalidáció itt sem értelmezett, hiszen annak célja nem egy-egy paraméter, hanem a teljes modell teljesítményének értékelése.

A teljes modell szintjén, például az R-négyzet értékelésénél mutatkoznak meg a legnagyobb különbségek. A hagyományos gyakorlatban itt már ritkán számolunk konfidenciaintervallumot. A modell értékelése jellemzően az R-négyzet nagysága és a modell egészére vonatkozó szignifikanciateszt (F-statisztika) együttes mérlegelésével történik. Tehát a szignifikanciatesztelés az R^2 értékelése mellett történik meg annak eldöntésére, hogy szabad-e és érdemes-e a modellt használni. Bár a bootstrap elvileg az R^2-re is számolható lenne, ez a gyakorlatban nem terjedt el. A keresztvalidáció azonban pontosan ezen a szinten tudna igazán kibontakozni, ha a gyakorlatban elterjedne. Egyfajta pontosabb, jobb értékelése lehetne a modellnek, amit a modellhez tartozó szignifikancia mellett vagy akár helyette is használhatnánk annak eldöntésére, hogy szabad-e az adott modellel dolgozni. Ez a gondolatmenet elvezet egy modern, robusztus kutatási munkafolyamat javaslatához. Képzeljük el,

hogy a keresztvalidációt egyfajta kapuőrként alkalmazzuk. Ha a modell ezen a teszten átmegy, azaz a keresztvalidált teljesítménye elfogadható és stabil, az zöld utat ad a második lépéshez, a modell belső paramétereinek (például a regressziós együtthatóknak) részletes, hagyományos statisztikai értelmezéséhez, amellyel már a konkrét elméleti kérdéseinkre kaphatunk választ. Így a keresztvalidáció nemcsak a prediktív modellezés eszköze lenne, hanem az elméleteket tesztelő, a világot jobban megismerni kívánó kutatások megbízhatóságát is alapvetően növelhetné.

6. Ellenőrző kérdések

1. kérdés: Mit mér a statisztikai szignifikanciatesztelés során kapott p-érték? Válassza ki a helyes definíciót! A helyes választ a kérdéssor végénél helyeztük el.

A) Annak a valószínűségét, hogy a kutatási (alternatív) hipotézisünk igaz.
B) Annak a valószínűségét, hogy a mintánkban látott eredményt a puszta véletlen okozta.
C) Annak a valószínűségét, hogy legalább akkora hatást mérünk a mintában, mint amit mértünk, feltéve, hogy a populációban valójában nincs hatás (a nullhipotézis igaz).
D) Annak a valószínűségét, hogy tévedünk, amikor a szignifikáns eredmény alapján elvetjük a nullhipotézist.

2. kérdés: Mi a bootstrap módszer központi logikája? Miért a visszatevéses újra-mintavételezés a folyamat kulcsa?

Válasz: A bootstrap központi logikája, hogy az egyetlen rendelkezésre álló mintánkat a teljes populáció miniatűr tükörképének tekinti. Mivel a valódi populációból nem tudunk több ezer új mintát venni a bizonytalanság felmérésére, ezért ebből a minipopulációból generálunk több ezer új, úgynevezett bootstrap mintát. A visszatevés biztosítja, hogy az új minták tiszteletben tartsák az eredeti minta eloszlását. Egy olyan értéknek, amelyik gyakori volt az eredeti mintában,

minden egyes húzásnál nagyobb esélye van bekerülni, így a bootstrap minták hűen tükrözik az eredeti adatok sűrűségét és mintázatát.

3. kérdés: Mi a keresztvalidáció elsődleges célja? Miért jelent előrelépést a minta egyszerű, egyszeri tanító és tesztelő adathalmazra való felosztásához képest?

Válasz: A keresztvalidáció elsődleges célja, hogy egy robusztus és reális becslést adjon a modell várható, jövőbeli teljesítményére új, eddig sosem látott adatokon. A fő feladata a modell általánosítóképességének felmérése és a túlillesztés elkerülése. Az egyszerű, egyszeri tanító-tesztelő felosztáshoz képest két fő okból jelent előrelépést: kiküszöböli az adatvesztést és megszünteti a szerencsefaktort. Egyetlen felosztásnál a kapott teljesítménymutató nagyban függhet a véletlentől (hogy éppen milyen adatok kerültek a teszthalmazba). A keresztvalidáció a teljesítményt több különböző felosztáson méri, majd átlagolja, így egy sokkal stabilabb és megbízhatóbb becslést ad a modell valódi, várható teljesítményére.

4. kérdés: Miben áll a legfőbb elméleti különbség a hagyományos szignifikanciatesztelés és a bootstrapalapú konfidenciaintervallumok között a becslések bizonytalanságának meghatározásában?

Válasz: A legfőbb elméleti különbség a bizonytalanság becslésének forrásában rejlik. A hagyományos szignifikanciatesztelés (és a hagyományos CI) elméletvezérelt. Előre feltételezi, hogy az adatok egy bizonyos elméleti eloszlásból (például normális eloszlásból) származnak, és ezen feltételezések alapján, matematikai képletek segítségével számítja ki a bizonytalanságot. A bootstrapalapú konfidenciaintervallum ezzel szemben adatalapú és empirikus. Nem él szigorú elméleti feltételezésekkel az adatok eloszlásáról. Helyette a rendelkezésre álló mintából, számítógépes újra-mintavételezéssel hozza létre a lehetséges kimenetelek eloszlását, és ebből, közvetlenül az adatokból méri le a bizonytalanságot.

5. kérdés: Miben áll a legfőbb elméleti különbség a bootstrap és a keresztvalidáció módszertana között?

Válasz: Míg a bootstrap a becslések bizonytalanságát vizsgálja, addig a keresztvalidáció a modell teljesítményét értékeli. A bootstrap az inferencia (statisztikai következtetés) eszköze. Arra a kérdésre válaszol, hogy „Mennyire bizonytalan egy adott paraméter becslése?”. A célja egyetlen mutató (például egy regressziós

együttható) megbízhatósági tartományának (konfidenciaintervallumának) meghatározása. A fókusz a modell egy belső tulajdonságának precizitásán van. A keresztvalidáció a predikció (bejóslás) eszköze. Arra a kérdésre ad választ, hogy „Mennyire jól működik a teljes modell a gyakorlatban, új adatokon?". A célja a modell egészének külső, gyakorlati teljesítményének (például R^2-ének) felmérése. A fókusz a modell általánosítóképességén, a gyakorlati hasznosságán van.

Válasz az 1. kérdésre: C

Ajánlott irodalom

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*(7748), 305–307. https://doi.org/10.1038/d41586-019-00857-9

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Efron, B., & Hastie, T. (2021). *Computer Age Statistical Inference, Student Edition*. Cambridge University Press. https://doi.org/10.1017/9781108914062

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-031-38747-0

Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology, 8*(1), Article 33267. https://doi.org/10.1525/collabra.33267

Rooij, M. de, & Weeda, W. (2020). Cross-Validation: A Method Every Psychologist Should Know. *Advances in Methods and Practices in Psychological Science, 3*(2), 248–263. https://doi.org/10.1177/2515245919898466

Rust, J., Kosinski, M., & Stillwell, D. (2021). *Modern psychometrics: The science of psychological assessment*. Routledge.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond „ p < 0.05". *The American Statistician, 73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wilcox, R. R. (2021). *Introduction to robust estimation and hypothesis testing*. Elsevier.

Irodalom

Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research. *Journal of Medicine, Surgery, and Public Health, 2*, 100051. https://doi.org/10.1016/j.glmedi.2024.100051

Amrhein, V., Greenland, S., & McShane, B. (2019). *Scientists rise up against statistical significance*. *Nature, 567*(7748), 305–307. https://doi.org/10.1038/d41586-019-00857-9

Arseniev-Koehler, A. (2024). Theoretical foundations and limits of word embeddings: What types of meaning can they capture? *Sociological Methods & Research, 53*(4), 1753–1793. https://doi.org/10.1177/00491241221140142

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173

Bartoš, F., Maier, M., Wagenmakers, E.-J., Nippold, F., Doucouliagos, H., Ioannidis, J. P. A., Otte, W. M., Sladekova, M., Deresssa, T. K., Bruns, S. B., Fanelli, D., & Stanley, T. D. (2024). Footprint of publication selection bias on meta-analyses in medicine, environmental sciences, psychology, and economics. *Research Synthesis Methods*, *15*(3), 500–511. https://doi.org/10.1002/jrsm.1703

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, *2*, 8–14. https://doi.org/10.1016/j.npls.2016.01.001

Biggs, R., Preiser, R., Vos, A. de, Schlüter, M., Maciejewski, K., & Clements, H. (Eds.). (2021). *The Routledge Handbook of Research Methods for Social-Ecological Systems*. Routledge. https://doi.org/10.4324/9781003021339

Blair, G., Coppock, A., & Humphreys, M. (2023). *Research design in the social sciences: Declaration, diagnosis, and redesign*. Princeton University Press.

Braga, A. A., Turchan, B. S., Papachristos, A. V., & Hureau, D. M. (2019). Hot spots policing and crime reduction: An update of an ongoing systematic review and meta-analysis. *Journal of Experimental Criminology, 15*(3), 289–311. https://doi.org/10.1007/s11292-019-09372-3

Braun, V., & Clarke, V. (2021). Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, *21*(1), 37–47. https://doi.org/10.1002/capr.12360

Braun, V., & Clarke, V. (2021). To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health*, *13*(2), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE.

Braun, V., & Clarke, V. (2023). Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. *Health Psychology Review*, *17*(4), 695–718. https://doi.org/10.1080/17437199.2022.2161594

Brohinsky, J., Marquart, C., Wang, J., Ruis, A. R., & Shaffer, D. W. (2021). Trajectories in Epistemic Network Analysis. In A. R. Ruis & S. B. Lee (Eds.), *Communications in Computer and Information Science. Advances in Quantitative Ethnography* (Vol. 1312, pp. 106–121). Springer International Publishing. https://doi.org/10.1007/978-3-030-67788-6_8

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.

Bruckmaier, G., Krauss, S., Binder, K., Hilbert, S., & Brunner, M. (2021). Tversky and Kahneman's Cognitive Illusions: Who Can Solve Them, and Why? *Frontiers in Psychology*, *12*, 584689. https://doi.org/10.3389/fpsyg.2021.584689

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity*, *56*(3), 1391–1412. https://doi.org/10.1007/s11135-021-01182-y

Campbell, K., Orr, E., Durepos, P., Nguyen, L., Li, L., Whitmore, C., Gehrke, P., Graham, L., & Jack, S. (2021). Reflexive Thematic Analysis for Applied Qualitative Health Research. *The Qualitative Report*. Advance online publication. https://doi.org/10.46743/2160-3715/2021.5010

Carlson, R., Bauer, J., & Manning, C. D. (2025). *A New Pair of GloVes*. https://doi.org/10.48550/arXiv.2507.18103

Carlson, S. (1985). A double-blind test of astrology. *Nature*, *318*(6045), 419–425. https://doi.org/10.1038/318419a0

Chang, H.-Y., & Tsai, C.-C. (2023). Epistemic Network Analysis of Students' Drawings to Investigate Their Conceptions of Science Learning with Technology. *Journal of Science Education and Technology*. Advance online publication. https://doi.org/10.1007/s10956-022-10026-9

Charmaz, K. (2025). *Constructing grounded theory. Introducing qualitative methods*. SAGE.

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models. *Sociological Methods & Research*, *36*(4), 462–494. https://doi.org/10.1177/0049124108314720

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Clauser, B. E., & Bunch, M. B. (Eds.). (2021). *The history of educational measurement: Key advancements in theory, policy, and practice*. Routledge.

Cole, G. G. (2021). The Grievance Studies Affair; One Funeral at a Time: A Reply to Pluckrose, Lindsay, and Boghossian. *Sociological Methods & Research*, *50*(4), 1937–1945. https://doi.org/10.1177/00491241211009949

Conaghan, C., Daly, E., & Ryan, L. (2024). Reflexive thematic analysis of interviews with postprimary educators' experience and attitudes toward concussion. *Psychology in the Schools*, *61*(9), 3714–3727. https://doi.org/10.1002/pits.23257

Coutts, J. J., & Hayes, A. F. (2023). Questions of value, questions of magnitude: An exploration and application of methods for comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *55*(7), 3772–3785. https://doi.org/10.3758/s13428-022-01988-0

Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.

Csíkos, C. (2020). *A neveléstudomány kutatásmódszertanának alapjai*. ELTE.

Dean, G. O., & Kelly, I. W. (2003). Is astrology relevant to consciousness and psi? *Journal of Consciousness Studies*, *10*(6-7), 175–198.

Di Paola, N., Chari, S., Iannacci, F., & Kraus, S. (2025). Configurational theory in business and management research: Status quo and guidelines for the application of qualitative comparative analysis (QCA). *Technological Forecasting and Social Change*, *211*, 123907. https://doi.org/10.1016/j.techfore.2024.123907

Dijkstra, T. K., & Henseler, J. (2015). Consistent Partial Least Squares Path Modeling1. *MIS Quarterly*, *39*(2), 297–316. https://doi.org/10.25300/MISQ/2015/39.2.02

Doval, E., Viladrich, C., & Angulo-Brunet, A. (2023). Coefficient Alpha: The Resistance of a Classic. *Psicothema*, *35*(1), 5–20. https://doi.org/10.7334/psicothema2022.321

Dunivin, Z. O. (2025). Scaling hermeneutics: a guide to qualitative coding with LLMs for reflexive content analysis. *EPJ Data Science*, *14*(1). https://doi.org/10.1140/epjds/s13688-025-00548-8

Duşa, A. (2019). *QCA with R*. Springer. https://doi.org/10.1007/978-3-319-75668-4

Dybro Liengaard, B. (2024). Measurement invariance testing in partial least squares structural equation modeling. *Journal of Business Research*, *177*, 114581. https://doi.org/10.1016/j.jbusres.2024.114581

Efron, B., & Hastie, T. (2021). *Computer Age Statistical Inference, Student Edition*. Cambridge University Press. https://doi.org/10.1017/9781108914062

Eisenstein, J. (2019). *Introduction to natural language processing. Adaptive computation and machine learning*. The MIT Press.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Espregueira-Mendes, J., Karlsson, J., Musahl, V., & Ayeni, O. R. (Eds.). (2025). *Orthopaedic Sports Medicine*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-65430-6

Feyerabend, P. (2010). *Against method*. Verso.

Finlay, L. (2021). Thematic Analysis. *European Journal for Qualitative Research in Psychotherapy*(11), 103–116. https://doi.org/10.24377/EJQRP.article3062

Fiss, P. C., Cambré, B., & Marx, A. (Eds.). (2013). *Configurational Theory and Methods in Organizational Research*. Emerald Group Publishing Limited. https://doi.org/10.1108/rso

Folliard, K. J., Crozier, K., & Kamble, M. M. W. (2024). A very, very lonely, unmagical time. The lived experience of perinatal anxiety: A longitudinal interpretative phenomenological analysis. *Midwifery*, *136*, 104070. https://doi.org/10.1016/j.midw.2024.104070

Forer, B. R. (1949). The fallacy of personal validation; a classroom demonstration of gullibility. *Journal of Abnormal Psychology*, *44*(1), 118–123. https://doi.org/10.1037/h0059240

Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alsahfi, T., & Alshemaimri, B. (2025). BERT applications in natural language processing: a review. *Artificial Intelligence Review*, *58*(6). https://doi.org/10.1007/s10462-025-11162-5

Gelman, A., Hill, J., & Vehtari, A. (2020). *Regression and Other Stories*. Cambridge University Press. https://doi.org/10.1017/9781139161879

Geremew, Y. M., Huang, W.-J., & Hung, K. (2024). Fuzzy-Set Qualitative Comparative Analysis as a Mixed-Method and Analysis Technique: A Comprehensive Systematic Review. *Journal of Travel Research*, *63*(1), 3–26. https://doi.org/10.1177/00472875231168619

Gliner, J. A., Morgan, G. A., & Leech, N. L. (2017). *Research methods in applied settings: An integrated approach to design and analysis*. Routledge Taylor & Francis Group.

Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating Model Fit of Measurement Models in Confirmatory Factor Analysis. *Educational and Psychological Measurement*, *84*(1), 123–144. https://doi.org/10.1177/00131644231163813

Greckhamer, T., Furnari, S., Fiss, P. C., & Aguilera, R. V. (2018). Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. *Strategic Organization*, *16*(4), 482–495. https://doi.org/10.1177/1476127018786487

Greckhamer, T., Misangyi, V. F., & Fiss, P. C. (2013). The Two QCAs: From a Small-N to a Large-N Set Theoretic Approach. In P. C. Fiss, B. Cambré, & A. Marx (Eds.), *Configurational Theory and Methods in Organizational Research* (Vol. 38, pp. 49–75). Emerald Group Publishing Limited. https://doi.org/10.1108/S0733-558X(2013)0000038007

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. https://doi.org/10.48550/arXiv.2203.05794

Groskurth, K., Bluemke, M., & Lechner, C. M. (2024). Why we need to abandon fixed cutoffs for goodness-of-fit indices: An extensive simulation and possible solutions. *Behavior Research Methods*, *56*(4), 3891–3914. https://doi.org/10.3758/s13428-023-02193-3

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)*. SAGE.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hajtman, E., & Hajtman, B. (2018). *Gyakorló példák a pszichológus hallgatók statisztikai tantárgyainak anyagához*. Akadémiai Kiadó. https://doi.org/10.1556/9789634542513

Hak, T., Jaspers, F., & Dul, J. (2013). The Analysis of Temporally Ordered Configurations: Challenges and Solutions. In P. C. Fiss, B. Cambré, & A. Marx (Eds.), *Configurational Theory and Methods in Organizational Research* (Vol. 38, pp. 107–127). Emerald Group Publishing Limited. https://doi.org/10.1108/S0733-558X(2013)0000038009

Hankar, M., Kasri, M., & Beni-Hssane, A. (2025). A comprehensive overview of topic modeling: Techniques, applications and challenges. *Neurocomputing*, *628*, 129638. https://doi.org/10.1016/j.neucom.2025.129638

Hardwicke, T. E., Thibault, R. T., Clarke, B., Moodie, N., Crüwell, S., Schiavone, S. R., Handcock, S. A., Nghiem, K. an, Mody, F., Eerola, T., & Vazire, S. (2024). Prevalence of Transparent Research Practices in Psychology: A Cross-Sectional Study of Empirical Articles Published in 2022. *Advances in Methods and Practices in Psychological Science*, *7*(4), Article 25152459241283477. https://doi.org/10.1177/25152459241283477

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2022). *Multilevel and Longitudinal Modeling with IBM SPSS*. Routledge. https://doi.org/10.4324/9780367824273

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900–1902. https://doi.org/10.1126/science.1128898

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hill, C. E., & Knox, S. (2021). *Essentials of consensual qualitative research*. American Psychological Association. https://doi.org/10.1037/0000215-000

Hove, D. T., Jorgensen, T. D., & van der Ark, L. A. (2024). Updated guidelines on selecting an intraclass correlation coefficient for interrater reliability, with applications to incomplete observational designs. *Psychological Methods*, *29*(5), 967–979. https://doi.org/10.1037/met0000516

Howard, M. C., Boudreaux, M., Cogswell, J., Manix, K. G., & Oglesby, M. T. (2025). A literature review of model fit and model comparisons with confirmatory factor analysis: Formalizing the informal in organizational science. *Applied Psychology*, *74*(1), Article e12592. https://doi.org/10.1111/apps.12592

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hunyadi, L., & Vita, L. (2019). *Statisztika I.* Akadémiai Kiadó. https://doi.org/10.1556/9789634542216

Hurlbert, S. H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs*, *54*(2), 187–211. https://doi.org/10.2307/1942661

Husserl, E. (2012). *Ideas: General introduction to pure phenomenology*. Routledge.

Ivanović, J. (2016). *Bevezetés a pedagógiai kutatás módszertanába*. Újvidéki Egyetem.

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning*. Springer International Publishing. https://doi.org/10.1007/978-3-031-38747-0

Jenkins, R. (2025). *Pierre Bourdieu*. Routledge.

Jiang, J., & Nguyen, T. (2021). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer New York. https://doi.org/10.1007/978-1-0716-1282-8

Johnson, S. J., Murty, M. R., & Navakanth, I. (2024). A detailed review on word embedding techniques with emphasis on word2vec. *Multimedia Tools and Applications*, *83*(13), 37979–38007. https://doi.org/10.1007/s11042-023-17007-z

Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*.

Kakamad, F. H., Abdalla, B. A., Abdullah, H. O., Omar, S. S., Mohammed, S. H., Ahmed, S. M., Mohammed, K. K., Hassan, H. A [Hemn A.], Baba, H. O., Ahmed, J. O., Mustafa, M. Q., Omar, D. A., Salih, R. Q., Rahim, H. M., Hussein, D. A., Hassan, M. N., Mikael, T. M., Hassan, H. A [Hunar A.], & Najar, K. A. (2024). Lists of predatory journals and publishers: a review for future refinement. *European Science Editing*, *50*, Article e118119. https://doi.org/10.3897/ese.2024.e118119

Kalkbrenner, M. T. (2024). Choosing Between Cronbach's Coefficient Alpha, McDonald's Coefficient Omega, and Coefficient H: Confidence Intervals and the Advantages and Drawbacks of Interpretive Guidelines. *Measurement and Evaluation in Counseling and Development*, *57*(2), 93–105. https://doi.org/10.1080/07481756.2023.2283637

Kaposi, J., & Szőke-Milinte, E. (Eds.). (2022). *Kutatási módszerek pedagógusjelölteknek*. Pázmány Péter Katolikus Egyetem.

Kassai, S. (2020). *Using Interpretative Phenomenological Analysis (IPA) to Assess Recovery Processes : A Qualitative Analysis of Experience and Identity* (Vol. 19). L'Harmattan. https://doi.org/10.56037/978-963-414-617-9

King, N. (2019). *Interviews in qualitative research*. SAGE.

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. The Guilford Press.

Kontra, J. (2011). *A pedagógiai kutatások módszertana: Egyetemi jegyzet*. Kaposvári Egyetem.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Korbmacher, M., Azevedo, F., Pennington, C. R., Hartmann, H., Pownall, M., Schmidt, K., Elsherif, M., Breznau, N., Robertson, O., Kalandadze, T., Yu, S., Baker, B. J., O'Mahony, A., Olsnes, J. Ø.-S., Shaw, J. J., Gjoneska, B., Yamada, Y., Röer, J. P., Murphy, J., … Evans, T. (2023). The replication crisis has led to positive structural, procedural, and community changes. *Communications Psychology*, *1*(1), 3. https://doi.org/10.1038/s44271-023-00003-2

Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *The European Journal of General Practice*, *24*(1), 120–124. https://doi.org/10.1080/13814788.2017.1375092

Krippendorff, K. (2019). *Content Analysis: An Introduction to Its Methodology*. SAGE. https://doi.org/10.4135/9781071878781

Kuckartz, U., & Rädiker, S. (2023). *Qualitative content analysis: Methods, practice and software*. SAGE.

Kuhn, T. S. (2012). *The structure of scientific revolutions*. The University of Chicago Press.

Lagerspetz, M. (2021). "The Grievance Studies Affair" Project: Reconstructing and Assessing the Experimental Design. *Science, Technology, & Human Values*, *46*(2), 402–424. https://doi.org/10.1177/0162243920923087

Lakens, D. (2022). Sample Size Justification. *Collabra: Psychology*, *8*(1), Article 33267. https://doi.org/10.1525/collabra.33267

Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.). (2017). *Handbook of Learning Analytics*. Society for Learning Analytics Research. https://doi.org/10.18608/hla17

Lennox, C., & Payne-Mann, C. (2025). An Explanation of Path Analysis and Recommendations for Best Practice. *Contemporary Accounting Research*, Article 1911-3846.70019. Advance online publication. https://doi.org/10.1111/1911-3846.70019

Lim, W. M. (2025). What Is Qualitative Research? An Overview and Guidelines. *Australasian Marketing Journal*, *33*(2), 199–229. https://doi.org/10.1177/14413582241264619

Marx, A., & Dusa, A. (2011). Crisp-Set Qualitative Comparative Analysis (csQCA), Contradictions and Consistency Benchmarks for Model Specification. *Methodological Innovations Online*, *6*(2), 103–148. https://doi.org/10.4256/mio.2010.0037

Maxwell, S. E., Delaney, H. D., & Kelley, K. (2024). *Designing experiments and analyzing data: A model comparison perspective*. Routledge Taylor & Francis Group.

Mayring, P. (2022). *Qualitative content analysis: A step-by-step guide*. SAGE.

McCammon, L. C., Gillen, P., McLaughlin, D., & Kernohan, W. G. (2025). An Interpretative Phenomenological Analysis That Seeks to Describe and Understand the Personal Experience of Burnout in General Practitioners. *Qualitative Health Research*, *35*(1), 118–131. https://doi.org/10.1177/10497323241260738

McElreath, R. (2020). *Statistical Rethinking*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429029608

McLevey, J., Scott, J., & Carrington, P. J. (Eds.). (2024). *The Sage handbook of social network analysis*. SAGE.

McNeish, D., & Dumas, D. (2025). Reliability representativeness: How well does coefficient alpha summarize reliability across the score distribution? *Behavior Research Methods*, *57*(3), 93. https://doi.org/10.3758/s13428-025-02611-8

McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, *28*(1), 61–88. https://doi.org/10.1037/met0000425

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook*. SAGE.

Neuendorf, K. A. (2017). *The Content Analysis Guidebook*. SAGE. https://doi.org/10.4135/9781071802878

Ng, C. K. J., Kwan, L. Y. J., & Chan, W. (2024). A Note on Evaluating the Moderated Mediation Effect. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*(2), 340–356. https://doi.org/10.1080/10705511.2023.2201396

Nicmanis, M. (2024). Reflexive Content Analysis: An Approach to Qualitative Data Analysis, Reduction, and Description. *International Journal of Qualitative Methods*, *23*, Article 16094069241236603. https://doi.org/10.1177/16094069241236603

NISO CRediT Working Group. *ANSI/NISO Z39.104-2022, CRediT, Contributor Roles Taxonomy*. NISO.

Nizza, I. E., Farr, J., & Smith, J. A. (2021). Achieving excellence in interpretative phenomenological analysis (IPA): Four markers of high quality. *Qualitative Research in Psychology*, *18*(3), 369–386. https://doi.org/10.1080/14780887.2020.1854404

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, *73*, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, *19*, Article 1609406919899220. https://doi.org/10.1177/1609406919899220

Oana, I.-E., Schneider, C. Q., & Thomann, E. (2021). *Qualitative Comparative Analysis Using R*. Cambridge University Press. https://doi.org/10.1017/9781009006781

Pearl, J. (2013). *Causality*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Pessoa, A. S. G., Harper, E., Santos, I. S., & Da Gracino, M. C. S. (2019). Using Reflexive Interviewing to Foster Deep Understanding of Research Participants' Perspectives. *International Journal of Qualitative Methods*, *18*, Article 1609406918825026. https://doi.org/10.1177/1609406918825026

Phillips, D. C. (2014). *Encyclopedia of educational theory and philosophy*. SAGE.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. Springer. https://doi.org/10.1007/b98882

Poploski, K. M., Smith, C., & Popchak, A. J. (2025). Biases and Confounders. In J. Espregueira-Mendes, J. Karlsson, V. Musahl, & O. R. Ayeni (Eds.), *Orthopaedic Sports Medicine* (pp. 1–15). Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-65430-6_116-1

Popper, K. R. (2002). *The logic of scientific discovery*. Routledge.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Preiser, R., García, M. M., Hill, L., & Klein, L. (2021). Qualitative content analysis. In R. Biggs, R. Preiser, A. de Vos, M. Schlüter, K. Maciejewski, & H. Clements (Eds.), *The Routledge Handbook of Research Methods for Social-Ecological Systems* (pp. 270–281). Routledge. https://doi.org/10.4324/9781003021339-23

(2024). *Proceedings of the 14th Learning Analytics and Knowledge Conference*. ACM.

Rácz, J., Karsai, S., & Tóth, V. (2023). *Kvalitatív pszichológia*. ELTE Eötvös Kiadó.

Ragin, C. C. (2000). *Fuzzy-set social science*. University of Chicago Press.

Ragin, C. C. (2014). *The comparative method : moving beyond qualitative and quantitative strategies*. University of California Press.

Reid, J. W., Parrish, J., Syed, S. B., & Couch, B. (2024). Finding the Connections: A Scoping Review of Epistemic Network Analysis in Science Education. *Journal of Science Education and Technology*. Advance online publication. https://doi.org/10.1007/s10956-024-10193-x

Richtig, G., Berger, M., Koeller, M., Richtig, M., Richtig, E., Scheffel, J., Maurer, M., & Siebenhaar, F. (2023). Predatory journals: Perception, impact and use of Beall's list by the scientific community-A bibliometric big data study. *PloS One*, *18*(7), e0287547. https://doi.org/10.1371/journal.pone.0287547

Ringle, C. M., Sarstedt, M., Sinkovics, N., & Sinkovics, R. R. (2023). A perspective on using partial least squares structural equation modelling in data articles. *Data in Brief*, *48*, 109074. https://doi.org/10.1016/j.dib.2023.109074

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm : An R Package for Structural Topic Models. *Journal of Statistical Software*, *91*(2). https://doi.org/10.18637/jss.v091.i02

Roemer, E., Schuberth, F., & Henseler, J. (2021). HTMT2–an improved criterion for assessing discriminant validity in structural equation modeling. *Industrial Management & Data Systems*, *121*(12), 2637–2650. https://doi.org/10.1108/IMDS-02-2021-0082

Rooij, M. de, & Weeda, W. (2020). Cross-Validation: A Method Every Psychologist Should Know. *Advances in Methods and Practices in Psychological Science*, *3*(2), 248–263. https://doi.org/10.1177/2515245919898466

Roos, J. M., & Bauldry, S. (2022). *Confirmatory Factor Analysis*. SAGE. https://doi.org/10.4135/9781071938959

Rosseel, Y. (2012). lavaan : An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rönkkö, M., & Cho, E. (2022). An Updated Guideline for Assessing Discriminant Validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Ruis, A. R., & Lee, S. B. (Eds.). (2021). *Communications in Computer and Information Science*. *Advances in Quantitative Ethnography*. Springer International Publishing. https://doi.org/10.1007/978-3-030-67788-6

Rust, J., Kosinski, M., & Stillwell, D. (2021). *Modern psychometrics: The science of psychological assessment*. Routledge.

Sántha, K. (2009). *Bevezetés a kvalitatív pedagógiai kutatás módszertanába*. Eötvös József Kiadó.

Sántha, K. (2015a). *Multikódolt adatok kvalitatív elemzése*. Eötvös József Kiadó.

Sántha, K. (2015b). *Trianguláció a pedagógiai kutatásban*. Eötvös József Kiadó.

Sántha, K. (2022). *Kvalitatív tartalomelemzés*. Eötvös József Kiadó.

Saqr, M., & López-Pernas, S. (Eds.). (2024). *Learning Analytics Methods and Tutorials*. Springer International Publishing. https://doi.org/10.1007/978-3-031-54464-4

Sarstedt, M., Adler, S. J., Ringle, C. M., Cho, G., Diamantopoulos, A., Hwang, H., & Liengaard, B. D. (2024). Same model, same data, but different outcomes: Evaluating the impact of method choices in structural equation modeling. *Journal of Product Innovation Management*, *41*(6), 1100–1117. https://doi.org/10.1111/jpim.12738

Sathyanarayana, S., & Mohanasundaram, T. (2024). Fit Indices in Structural Equation Modeling and Confirmatory Factor Analysis: Reporting Guidelines. *Asian Journal of Economics, Business and Accounting*, *24*(7), 561–577. https://doi.org/10.9734/ajeba/2024/v24i71430

Savalei, V., & Reise, S. P. (2019). Don't Forget the Model in Your Model-based Reliability Coefficients: A Reply to McNeish. *Collabra: Psychology*, *5*(1), Article 36. https://doi.org/10.1525/collabra.247

Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge University Press.

Schoemann, A. M., Moore, E. W. G., & Yagiz, G. (2025). How and why to follow best practices for testing mediation models with missing data. *International Journal of Psychology : Journal International De Psychologie*, *60*(1), e13257. https://doi.org/10.1002/ijop.13257

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*, *3*(3), 9–45. https://doi.org/10.18608/jla.2016.33.3

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., & Mislevy, R. (2009). Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning. *International Journal of Learning and Media*, *1*(2), 33–53. https://doi.org/10.1162/ijlm.2009.0013

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic Network Analysis: A Worked Example of Theory-Based Learning Analytics. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), *Handbook of Learning Analytics* (pp. 175–187). Society for Learning Analytics Research. https://doi.org/10.18608/hla17.015

Sharma, P. N., Liengaard, B. D., Hair, J. F., Sarstedt, M., & Ringle, C. M. (2023). Predictive model assessment and selection in composite-based modeling using PLS-SEM: extensions and guidelines for using CVPAT. *European Journal of Marketing*, *57*(6), 1662–1677. https://doi.org/10.1108/EJM-08-2020-0636

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *European Journal of Marketing*, *53*(11), 2322–2347. https://doi.org/10.1108/EJM-02-2019-0189

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Siebert-Evenstone, A. L., Arastoopour Irgens, G., Collier, W., Swiecki, Z., Ruis, A. R., & Williamson Shaffer, D. (2017). In Search of Conversational Grain Size: Modeling Semantic Structure using Moving Stanza Windows. *Journal of Learning Analytics*, *4*(3). https://doi.org/10.18608/jla.2017.43.7

Silver, C., & Lewins, A. (2014). *Using Software in Qualitative Research: A Step-by-Step Guide*. SAGE. https://doi.org/10.4135/9781473906907

Skaaning, S.-E. (2011). Assessing the Robustness of Crisp-set and Fuzzy-set QCA Results. *Sociological Methods & Research*, *40*(2), 391–408. https://doi.org/10.1177/0049124111404818

Smith, J. A. (2011). Evaluating the contribution of interpretative phenomenological analysis. *Health Psychology Review*, *5*(1), 9–27. https://doi.org/10.1080/17437199.2010.510659

Smith, J. A. (2024). Travelling in time: Using interpretative phenomenological analysis (IPA) to examine temporal aspects of the personal experience of issues in health and well-being *European Journal of Psychotherapy & Counselling*, *26*(3-4), 235–252. https://doi.org/10.1080/13642537.2024.2388663

Smith, J. A., Flowers, P., & Larkin, M. (2022). *Interpretative phenomenological analysis: Theory, method and research*. SAGE.

Smith, J. A., & Nizza, I. E. (2022). *Essentials of interpretative phenomenological analysis*. American Psychological Association.

Stapley, E., O'Keeffe, S., & Midgley, N. (2022). Developing Typologies in Qualitative Research: The Use of Ideal-type Analysis. *International Journal of Qualitative Methods*, *21*, Article 16094069221100633. https://doi.org/10.1177/16094069221100633

Stroup, W. W., Ptukhina, M., & Garai, J. (2024). *Generalized Linear Mixed Models*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429092060

Sun, Y., Hsieh, J., & Gao, X. (2025). Integrating Qualitative Comparative Analysis With Reflexive Thematic Analysis in Theme Development. *International Journal of Qualitative Methods*, *24*, Article 16094069251318749. https://doi.org/10.1177/16094069251318749

Swiecki, Z., Ruis, A. R., Farrell, C., & Shaffer, D. W. (2020). Assessing individual contributions to Collaborative Problem Solving: A network analysis approach. *Computers in Human Behavior*, *104*, 105876. https://doi.org/10.1016/j.chb.2019.01.009

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics*. Pearson.

Tan, Y., Swiecki, Z., Ruis, A. R., & Shaffer, D. (2024). Epistemic Network Analysis and Ordered Network Analysis in Learning Analytics. In M. Saqr & S. López-Pernas (Eds.), *Learning Analytics Methods and Tutorials* (pp. 569–636). Springer International Publishing. https://doi.org/10.1007/978-3-031-54464-4_18

Tang, J., Liu, G., Bai, J., & Jiang, J [Junrong] (2024). The impacts of peer assessment on critical thinking competence: An epistemic network analysis. *Journal of Hospitality, Leisure, Sport & Tourism Education*, *35*, 100515. https://doi.org/10.1016/j.jhlste.2024.100515

Taylor, N., Maduesesi, O., Vasiliou, V. S., & Thompson, A. R. (2025). The experience of living with vitiligo in Nigeria: A participatory Interpretative Phenomenological Analysis. *Journal of Health Psychology*, *30*(5), 1120–1135. https://doi.org/10.1177/13591053241261684

Terry, G., & Hayfield, N. (2021). *Essentials of thematic analysis*. American Psychological Association. https://doi.org/10.1037/0000238-000

Tinto, V. (2017). Through the Eyes of Students. *Journal of College Student Retention: Research, Theory & Practice*, *19*(3), 254–269. https://doi.org/10.1177/1521025115621917

Tomán, E. (2024). Genetikai örökség – családi test : Egy öröklődő betegség embodiment-tapasztalata. *Alkalmazott Pszichológia*, *26*. https://doi.org/10.17627/ALKPSZICH.2024.4.37

Trainor, L. R., & Bundon, A. (2021). Developing the craft: reflexive accounts of doing reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *13*(5), 705–726. https://doi.org/10.1080/2159676X.2020.1840423

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Vargha, A. (2015). *Matematikai statisztika - Pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. Pólya.

Vargha, A. (2019). *Többváltozós statisztika dióhéjban*. Pólya.

Vargha, A. (2020). *Normális vagy? És ha nem? - Statisztikai módszerek nem normális eloszlású változókkal pszichológiai kutatásokban*. Pólya.

Viberg, O., Baars, M., Mello, R. F., Weerheim, N., Spikol, D., Bogdan, C., Gasevic, D., & Paas, F. (2024). Exploring the nature of peer feedback: An epistemic network analysis approach. *Journal of Computer Assisted Learning*, *40*(6), 2809–2821. https://doi.org/10.1111/jcal.13035

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond " p < 0.05". *The American Statistician*, *73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wen, C., Clough, P., Paton, R., & Middleton, R. (2025). Leveraging large language models for thematic analysis: a case study in the charity sector. *AI & SOCIETY*. Advance online publication. https://doi.org/10.1007/s00146-025-02487-4

West, B. T., Welch, K. B., Gałecki, A. T., & Gillespie, B. W. (2022). *Linear Mixed Models*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003181064

Wilcox, R. R. (2021). *Introduction to robust estimation and hypothesis testing*. Elsevier.

Williams, L. J., Williams, A. R., & O'Boyle, E. H. (2024). Assessment of Path Model Fit: Evidence of Effectiveness and Recommendations for use of the RMSEA-P. *Organizational Research Methods*, *27*(1), 64–87. https://doi.org/10.1177/10944281221124946

Wolf, M. G., & McNeish, D. (2023). Dynamic : An R Package for Deriving Dynamic Fit Index Cutoffs for Factor Analysis. *Multivariate Behavioral Research*, *58*(1), 189–194. https://doi.org/10.1080/00273171.2022.2163476

Zahavi, D. (2021). Applied phenomenology: why it is safe to ignore the epoché. *Continental Philosophy Review*, *54*(2), 259–273. https://doi.org/10.1007/s11007-019-09463-y

Zhang, B., Zhou, Y., & Li, D. (2025). Can Human Reading Validate a Topic Model? *Sociological Methodology*, *55*(1), 59–90. https://doi.org/10.1177/00811750241265336

Zhao, L., Echeverria, V., Swiecki, Z., Yan, L., Alfredo, R., Li, X., Gasevic, D., & Martinez-Maldonado, R. (2024). Epistemic Network Analysis for End-users: Closing the Loop in the Context of Multimodal Analytics for Collaborative Team Learning. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 90–100). ACM. https://doi.org/10.1145/3636555.3636855

Angol–magyar terminológiai ekvivalenciajegyzék

Angol szakkifejezés	Magyar megfelelő
Adjacency matrix	Szomszédsági mátrix (ENA-nál)
Akaike Information Criterion (AIC)	Akaike-féle információs kritérium
Availability heuristic	Elérhetőségi heurisztika
Average Variance Extracted (AVE)	Átlagos magyarázott variancia
Bag-of-Words (BoW)	Szózsákmodell
Balanced design	Kiegyensúlyozott elrendezés
Bayesian Information Criterion (BIC)	Bayesi információs kritérium
Between-subject design	Független mintás elrendezés
Bias	Torzítás
Big Q / Small q	„Nagy Q” (kvalitatív paradigma) / „Kis q” (kvalitatív adat)
Block-diagonal matrix	Blokk-diagonális mátrix
Bootstrap / Bootstrapping	Bootstrap eljárás (visszatevéses újra-mintavételezés)
Bracketing	Zárójelezés (fenomenológiában)
Calibration	Kalibráció (QCA-nál)
Centroid	Centroid (súlypont)
Codebook	Kódkönyv
Comparative Fit Index (CFI)	Komparatív illeszkedési index
Compound Symmetry (CS)	Összetett szimmetria (kovarianciastruktúra)
Confidence Interval (CI)	Konfidenciaintervallum

Confirmatory Factor Analysis (CFA)	Megerősítő faktorelemzés
Confounding variable / Confounder	Zavaró tényező / Konfounder
Consistency	Konzisztencia (QCA-nál)
Convergent Validity	Konvergens validitás
Covariance	Kovariancia
Coverage	Lefedettség (QCA-nál)
Crisp set	Éles halmaz
Cross-loading	Kereszttöltés
Cross-validation	Keresztvalidáció
Data leakage	Adatszivárgás (gépi tanulásnál)
Degrees of Freedom (df)	Szabadságfok
Dependent variable	Függő változó
Dimensionality reduction	Dimenziócsökkentés
Discriminant Validity	Diszkrimináns validitás
Document-Term Matrix (DTM)	Dokumentum-term mátrix
Double hermeneutic	Kettős hermeneutika
Endogenous variable	Endogén (belső) változó
Epistemic Network Analysis (ENA)	Episztemikus hálózatelemzés
Equifinality	Ekvifinalitás
Exogenous variable	Exogén (külső) változó
Falsifiability	Cáfolhatóság (falszifikálhatóság)
File drawer problem	Fiókprobléma (publikációs torzítás)
Fixed effect	Fix hatás
Fuzzy set	Életlen (fuzzy) halmaz
Goodness of Fit	Illeszkedés jósága
Grand-mean centering	Nagyátlagcentrálás
Grounded Theory (GT)	Megalapozott elmélet

Heterotrait-Monotrait Ratio (HTMT)	HTMT-mutató (diszkrimináns validitás)
Independent variable	Független változó
Indicator / Item	Indikátor / Tétel
Interaction effect	Interakciós hatás
Inter-coder reliability	Kódolói megbízhatóság
Intraclass Correlation (ICC)	Osztályon belüli korreláció
Latent variable	Látens változó
Leave-One-Out Cross-Validation (LOOCV)	„Hagyj-ki-egyet” keresztvalidáció
Lemmatization	Lemmatizálás (szótövesítés szótárral)
Likelihood Ratio Test (LRT)	Likelihood-arány teszt
Linear Mixed Model (LMM)	Lineáris kevert modell
Loading (Factor loading)	Faktorterhelés / Töltés
Logical remainder	Logikai maradék (QCA-nál)
Long format	Hosszú adatformátum
Longitudinal study	Longitudinális vizsgálat
Manifest / Observed variable	Manifeszt / Megfigyelt változó
Maximum Likelihood (ML)	Maximum Likelihood becslés
Measurement error	Mérési hiba
Measurement model	Mérési modell
Mediator	Mediátor változó
Mixed Methods	Kevert módszertan
Model fit	Modellilleszkedés
Modification Index (MI)	Modifikációs index
Moderator	Moderátor változó
Moving window	Mozgó / Csúszó ablak (ENA-nál)
Multicollinearity	Multikollinearitás
Multigroup analysis	Többcsoportos elemzés

Named Entity Recognition (NER)	Névelem-felismerés
Natural Language Processing (NLP)	Természetesnyelv-feldolgozás
Necessary condition	Szükséges feltétel
Nested design	Beágyazott elrendezés
Null hypothesis	Nullhipotézis
Overfitting	Túlillesztés
Parsimony	Parszimónia (takarékosság elve)
Path Analysis (PA)	Útvonalelemzés
Path coefficient	Útvonal-együttható
p-hacking	p-hacking (szelektív közlés)
Population	Alapsokaság
Precision & Recall	Precizitás és Felidézés (NLP-metrikák)
Qualitative Comparative Analysis (QCA)	Kvalitatív összehasonlító elemzés
Random effect	Random (véletlen) hatás
Random Intercept	Random intercept (véletlen tengelymetszet)
Random Slope	Random slope (véletlen meredekség)
Reflexive Thematic Analysis (RTA)	Reflexív tematikus elemzés
Reliability	Megbízhatóság (reliabilitás)
Repeated measures	Ismételt mérések
Representativeness heuristic	Reprezentativitási heurisztika
Residual	Reziduális (maradék) hiba
Restricted Maximum Likelihood (REML)	Korlátozott Maximum Likelihood becslés
Root Mean Square Error of Approx. (RMSEA)	RMSEA illeszkedési mutató

Sample	Minta
Saturation	Telítődés / Szaturáció
Scaled Identity	Skálázott identitás (mátrix)
Selection bias	Kiválasztási torzítás
Sentiment Analysis	Szentimentelemzés
Significance level (p-value)	Szignifikanciaszint (p-érték)
Stanza	Sztanza (elemzési egység ENA-ban)
Standard Error (SE)	Standard hiba
Stemming	Szótövesítés (algoritmikus)
Stop words	Stopszavak (töltelékszavak)
Structural Equation Modeling (SEM)	Strukturális egyenletmodellezés
Structural model	Strukturális modell
Sufficient condition	Elégséges feltétel
Super-ordinate theme	Fölérendelt téma
Test set	Tesztadatbázis
TF-IDF	Term Frequency-Inverse Document Frequency
Topic Modeling	Témaelemzés
Training set	Tanító adatbázis
Triangulation	Trianguláció
Truth table	Igazságtábla (QCA-nál)
Tucker-Lewis Index (TLI)	Tucker–Lewis-index
Type I Error	Elsőfajú hiba
Type II Error	Másodfajú hiba
Validity	Érvényesség (validitás)
Variance	Variancia
Variance Components (VC)	Varianciakomponensek
Wide format	Széles adatformátum
Within-subject design	Összefüggő mintás elrendezés
Word Embedding	Szóbeágyazás

Glosszárium

Adatszivárgás (Data leakage): gépi tanulásnál (például NLP) fellépő hiba, amikor a tanító adatbázisba olyan információ kerül, amely a tesztadatokból származik (például a teljes mintán végzett standardizálás miatt), így a modell teljesítménye hamisan jónak tűnik.

Alapsokaság (Population): az a teljes csoport vagy halmaz, amelyre a kutatás eredményeit általánosítani kívánjuk, és amelyből a mintát vesszük.

Alulazonosított modell (Underidentified model): olyan statisztikai modell (például CFA, SEM), amelyben több a becsülendő paraméter, mint a rendelkezésre álló független információ, ezért matematikailag nem megoldható.

Átvihetőség (Transferability): a kvalitatív kutatások egyik minőségi kritériuma; azt jelenti, hogy a kutató olyan részletes leírást ad (sűrű leírás), amely alapján az olvasó eldöntheti, hogy az eredmények alkalmazhatók-e más kontextusban is.

AVE (Average Variance Extracted): a konvergens validitás mérőszáma CFA- és PLS-SEM-modelleknél. Azt mutatja meg, hogy a látens változó az indikátorainak varianciájából átlagosan mekkora részt magyaráz (az elfogadott érték > 0,50).

Beágyazott elrendezés (Nested design): olyan hierarchikus adatszerkezet vagy kísérleti elrendezés, ahol az egyik változó szintjei (például diákok) kizárólag egy másik, felsőbb szintű változó (például osztály) egyetlen szintjéhez tartoznak.

Beágyazott modellek (Nested models): statisztikai modellek összehasonlításánál használt fogalom; az „A" modell beágyazott a „B" modellbe, ha az „A" modell megkapható a „B"-ből bizonyos paraméterek rögzítésével (például nullára állításával).

Big Q paradigma: Braun és Clarke fogalma a minőségi kutatásra, amely nemcsak az adatok típusában (szöveg), hanem a kutatás filozófiájában (szubjektivitás, reflexivitás, jelentéskonstrukció) is eltér a pozitivista megközelítéstől.

Blokk-diagonális mátrix: olyan mátrixszerkezet lineáris kevert modelleknél, ahol a variancia-kovariancia értékek blokkokba rendeződnek (például csoportokon belül), a blokkokon kívüli értékek pedig nullák (függetlenség).

Bootstrap: számításigényes statisztikai eljárás, amely a rendelkezésre álló mintából visszatevéssel történő sokszori újra-mintavételezéssel becsüli meg a paraméterek eloszlását és standard hibáját anélkül, hogy feltételezné a normális eloszlást.

Centroid: az episztemikus hálózatelemzésben (ENA) a hálózat súlypontja a vizualizált térben; a csoportok vagy egyének átlagos pozícióját jelöli.

CFA (megerősítő faktorelemzés): olyan statisztikai módszer, amellyel egy előzetesen feltételezett mérési modell (melyik tétel melyik faktorhoz tartozik) illeszkedését teszteljük az adatokon.

Cohen-kappa (κ): a kódolói megbízhatóság mérőszáma, amely korrigálja a véletlenszerű egyezés esélyét két értékelő között kategorikus adatok esetén.

Confounder (zavaró változó): olyan harmadik változó, amely kapcsolatban áll mind a független, mind a függő változóval, és ha nem kontrolláljuk, hamis összefüggést (vagy az összefüggés hiányát) mutathat ki közöttük.

Cronbach-alfa: a belső konzisztencia (reliabilitás) hagyományos mérőszáma, amely azt mutatja, hogy egy skála tételei mennyire mérnek egy irányba.

Csúszó ablak (Moving window): az ENA és szövegelemzés eszköze; egy meghatározott méretű keret, amely végighalad az adatokon, hogy az egymáshoz közeli elemek közötti kapcsolatokat feltárja (például együttes előfordulást egy sztanzában).

Deduktív kódolás: olyan elemzési stratégia, ahol a kódokat és kategóriákat előzetes elméletek vagy keretrendszerek alapján, „felülről lefelé" határozzuk meg.

Diszkriminánsvaliditás: annak mértéke, hogy egy látens változó mennyire különbözik más, a modellben szereplő látens változóktól (nem olvadnak-e össze).

Ekvifinalitás: a QCA alapelve, mely szerint ugyanaz a kimenet (eredmény) többféle, eltérő oksági útvonalon (konfiguráción) keresztül is létrejöhet.

Elérhetőségi heurisztika: kognitív torzítás, amikor egy esemény valószínűségét az alapján ítéljük meg, hogy milyen könnyen jutnak eszünkbe rá példák.

Elsőfajú hiba (Type I error): tévesen elvetjük az igaz nullhipotézist (azt hisszük, van hatás, pedig nincs).

Endogén változó: a modell belső változója, amelyet más változók magyaráznak (a nyilak rá mutatnak). Megfelel a függő változónak.

Episztemikus keret: egy szakmai közösségre jellemző gondolkodásmód, amely a tudás, készségek, értékek, identitás és ismeretelmélet hálózatából áll.

Exogén változó: a modell külső változója, amelyet a modell nem magyaráz (nyíl indul belőle, de nem mutat rá). Megfelel a független változónak.

Faktortöltés (Loading): azt mutatja meg, hogy egy adott indikátor (például kérdőíves tétel) milyen erősen és milyen irányban kapcsolódik a mögötte álló látens változóhoz.

Falszifikálhatóság: Karl Popper kritériuma, mely szerint egy állítás akkor tudományos, ha elvileg létezik olyan megfigyelés, amely cáfolhatná.

Fix hatás (Fixed effect): olyan magyarázó változó a modellben, amelynek minden releváns szintjét megmértük (például Nem: férfi/nő; Kezelés: kapott/nem kapott), és az eredményeket konkrétan ezekre a szintekre vonatkoztatjuk.

Formatív mérési modell: olyan modell (főleg PLS-SEM-ben), ahol az indikátorok „okozzák" vagy alkotják a látens változót, nem pedig a látens változó tükröződik bennük.

Heterogén diagonális mátrix: olyan kovarianciastruktúra, ahol a csoportok (például ismételt mérések időpontjai) varianciája eltérő lehet, de egymástól függetlenek.

Hierarchikus adatok: olyan adatszerkezet, ahol az elemzési egységek magasabb szintű csoportokba rendeződnek (például tanulók osztályokba, mérések személyeken belül).

Hitelesség (Credibility): a kvalitatív kutatás belső érvényességének megfelelője; azt jelzi, hogy az eredmények mennyire hűen tükrözik a résztvevők valóságát.

Hosszú formátum (Long format): adatszerkezeti forma, ahol minden megfigyelés (például minden időpont) külön sorban szerepel, így egy személyhez több sor is tartozhat. LMM-elemzéshez szükséges.

HTMT (Heterotrait-Monotrait Ratio): a diszkrimináns validitás modern mérőszáma; ha értéke magas (például > 0,85), a két konstruktum nem különül el eléggé.

ICC (Intraclass Correlation – osztályon belüli korreláció): azt mutatja meg, hogy a teljes variancia mekkora hányada tulajdonítható a csoporttagságnak (klaszteresődésnek). Ha értéke magas, indokolt a többszintű modellezés.

Illeszkedési mutatók (Fit indices): számértékek (például CFI, RMSEA), amelyek jelzik, mennyire felel meg az elméleti modell az adatok empirikus szerkezetének.

Induktív kódolás: „alulról felfelé" építkező elemzés, ahol a kódokat közvetlenül az adatokból, előzetes elméleti keret nélkül hozzuk létre.

Inferencia: következtető statisztika; a mintából származó adatok alapján a populációra vonatkozó tulajdonságok becslése (például p-értékkel, konfidencia-intervallummal).

Interpretatív fenomenológiai elemzés (IPA): kvalitatív módszer, amely az egyének megélt tapasztalatainak és jelentésadásának mélyreható, idiografikus vizsgálatára összpontosít.

Kalibráció: a QCA-módszerben az a folyamat, amikor a nyers adatokat (például jövedelem) elméleti tudás alapján halmaztagsági értékekké (0-1) alakítjuk.

Keresztvalidáció (Cross-validation): a modell prediktív erejének ellenőrzése úgy, hogy az adatokat több részre osztjuk, és a modellen tanult összefüggéseket a modell számára ismeretlen (teszt) adatokon értékeljük.

Kettős hermeneutika: az IPA alapelve: a résztvevő értelmezi a saját világát, a kutató pedig értelmezi a résztvevő értelmezését.

Konfidenciaintervallum (CI): egy becsült tartomány, amely egy adott valószínűséggel (jellemzően 95%) tartalmazza a populáció valódi paraméterét.

Konzisztencia (Consistency): QCA-ban az a mérőszám, amely megmutatja, hogy egy feltétel (vagy konfiguráció) jelenléte mennyire megbízhatóan vonja maga után a kimenetelt.

Konvergens validitás: annak mértéke, hogy egy konstruktum indikátorai mennyire mérik ugyanazt a dolgot (például AVE segítségével mérve).

Kovariancia: két változó együtt járásának mértéke; azt mutatja meg, hogy az értékek hajlamosak-e egyszerre növekedni vagy csökkenni.

Krippendorff-alfa: robusztus kódolói megbízhatósági mutató, amely bármilyen mérési szintű adatra és tetszőleges számú kódolóra alkalmazható.

Látens változó: olyan elméleti konstrukció (például attitűd, intelligencia), amelyet közvetlenül nem tudunk mérni, de indikátorok segítségével modellezhetünk.

Lefedettség (Coverage): QCA-ban az a mérőszám, amely megmutatja, hogy egy adott megoldási útvonal a kimenetel eseteinek mekkora hányadát magyarázza.

Lineáris kevert modell (LMM): olyan statisztikai modell, amely fix és random hatásokat is tartalmaz, így alkalmas a függő (klaszterezett vagy ismételt) adatok kezelésére.

Logikai minimalizálás: a QCA eljárása, amely során a bonyolult igazságtáblasorokat a Boole-algebra szabályai szerint egyszerűsítjük, hogy megtaláljuk a lényegi oksági mintázatokat.

Manifest változó: közvetlenül megfigyelt, mért változó (például egy válasz a kérdőíven).

Mediátor: olyan változó, amely közvetíti a hatást a független és a függő változó között (választ ad a „hogyan" kérdésre).

Megerősítő faktorelemzés (CFA): olyan eljárás, amelyben teszteljük, hogy az adatok illeszkednek-e egy előre meghatározott faktorstruktúrához.

Megfigyelések függetlensége: a klasszikus lineáris modellek alapfeltétele, mely szerint egyik adatpont értéke nem befolyásolhatja a másikat. Ennek sérülésekor (például csoportosulás) van szükség LMM-re vagy beágyazott ANOVA-ra.

Mérési hiba: a mért érték és a valódi érték közötti különbség. A SEM egyik előnye, hogy képes ezt a hibát elkülöníteni a modellben.

Mérési modell: a SEM azon része, amely a látens változók és a manifeszt indikátoraik kapcsolatát írja le.

Moderátor: olyan változó, amely befolyásolja két másik változó kapcsolatának erősségét vagy irányát (választ ad a „mikor" vagy „kinek" kérdésre).

Multikollinearitás: probléma a regressziós modellekben, amikor a magyarázó változók túl erősen korrelálnak egymással, ami instabillá teszi a becslést.

Névelem-felismerés (NER): NLP-eljárás, amely a szövegben található tulajdonneveket (személyek, helyek, szervezetek) azonosítja és kategóriákba sorolja.

Nullhipotézis (H0): az a feltételezés, hogy a vizsgált hatás nem létezik (nincs különbség, nincs összefüggés). A statisztikai próba célja ennek elvetése.

P-érték: annak valószínűsége, hogy a kapott eredményt (vagy annál extrémebbet) a véletlen szülte, feltételezve, hogy a nullhipotézis igaz.

Parszimónia (takarékosság elve): az az elv, mely szerint a legegyszerűbb, legkevesebb feltételezéssel élő magyarázatot vagy modellt kell választani, amely még jól leírja az adatokat.

Precizitás (Precision): NLP-ben és osztályozásnál használt mutató; a pozitívnak ítélt esetek közül mennyi volt valóban pozitív.

Prediktív relevancia (Q^2): a PLS-SEM-modell előre jelző képességének mérőszáma; ha pozitív, a modell releváns predikciót ad.

Pszeudoreplikáció: statisztikai hiba, amikor nem független adatokat (például ismételt méréseket) függetlenként kezelünk, mesterségesen megnövelve a mintanagyságot és az elsőfajú hiba esélyét.

Random hatás (Random effect): olyan változó a modellben, amelynek szintjei egy nagyobb populációból vett mintát képeznek (például iskolák, vizsgálati alanyok), és célunk a varianciájuk becslése, nem a szintek összehasonlítása.

Random Intercept: a kevert modellekben az a beállítás, amely megengedi, hogy minden csoportnak (például személynek) saját, egyedi kiindulási értéke (alapvonala) legyen.

Random Slope: a kevert modellekben az a beállítás, amely megengedi, hogy a hatás meredeksége (erőssége) csoportonként eltérő legyen.

Reflexivitás: a kutató tudatos önvizsgálata arra vonatkozóan, hogyan befolyásolják saját előfeltevései, háttere és jelenléte a kutatási folyamatot és az értelmezést.

Reflexív tematikus elemzés (RTA): kvalitatív elemzési módszer, amely a témák aktív, kutató általi létrehozását és a szubjektivitás tudatos használatát hangsúlyozza.

Reflektív mérési modell: olyan modell, ahol a látens változó a „valódi ok", és ez tükröződik az indikátorokban (az indikátorok korrelálnak).

Reziduális (Maradék): az az érték, amelyet a modell nem tud megmagyarázni; a mért adat és a becsült adat különbsége.

Scaled Identity: a legegyszerűbb kovarianciastruktúra, amely feltételezi, hogy a hibák függetlenek és azonos varianciájúak (homoszkedaszticitás).

Stopszavak: s nyelv leggyakoribb, önálló jelentéssel alig bíró szavai (például a, az, és, hogy), amelyeket szövegelemzésnél gyakran kiszűrnek.

Strukturális egyenletmodellezés (SEM): olyan többváltozós statisztikai módszer, amely egyesíti a faktoranalízist és az útvonalelemzést, lehetővé téve látens változók közötti oksági láncolatok vizsgálatát.

Szaturáció (adattelítődés): a kvalitatív kutatásban az a pont, amikor az újabb adatok gyűjtése már nem hoz új információt vagy kódot.

Szemantikus tartalomelemzés (SCA): a szövegek nyílt, manifeszt tartalmának szisztematikus, kategóriaalapú leírására és számszerűsítésére törekvő módszer.

Szentimentelemzés: NLP-eljárás, amely a szövegek érzelmi töltetét (pozitív, negatív, semleges) azonosítja.

Szignifikancia: statisztikai döntés arról, hogy egy eredmény valószínűleg nem a véletlen műve.

Szóbeágyazás (Word Embedding): olyan NLP-technika, amely a szavakat numerikus vektorokká alakítja úgy, hogy a hasonló jelentésű szavak matematikailag közel kerülnek egymáshoz.

Szózsákmodell (Bag-of-Words): szövegelemzési megközelítés, amely a szöveget a benne lévő szavak rendezetlen halmazaként kezeli, csak a gyakoriságokra fókuszálva.

Témaelemzés (Topic Modeling): algoritmikus eljárás (például LDA), amely nagy szövegkorpuszokban rejlő rejtett tematikus struktúrákat tár fel a szavak együttes előfordulása alapján.

Tokenizáció: a szövegfeldolgozás első lépése, a szöveg darabolása elemi egységekre (szavakra, írásjelekre).

Trianguláció: az eredmények érvényességének növelése többféle adatforrás, módszer, kutató vagy elmélet együttes alkalmazásával.

Útvonalelemzés (Path Analysis): a változók közötti közvetlen és közvetett (mediált) hatások rendszerének statisztikai vizsgálata.

Validitás (érvényesség): annak mértéke, hogy a kutatás vagy mérőeszköz valóban azt méri-e, amit mérni szándékozott.

Variancia: a változó értékeinek szóródása az átlag körül; az információmennyiség mértéke a statisztikában.

Variance Components (VC): kovarianciastruktúra, amely a random hatások varianciáját becsüli, de feltételezi azok függetlenségét (nulla kovariancia).

Zavaró tényező kezelése: az a folyamat (például randomizálás, statisztikai kontroll), amellyel a harmadik változók torzító hatását kiküszöböljük az elemzésből.