

ELMÉLET–MÓDSZERTAN

DR. DUSEK TAMÁS

Területi statisztika, valószínűségszámítás és statisztikai következtetéselmélet*

A statisztikai módszerek tárgyalhatók egyrészt az alkalmazási területüktől függetlenül, formalizáltan, a matematikai elemekre koncentrálnak, másrészt az alapvető – és egyébként megkérdőjelezhetetlen – matematikai ismereteken túlmutatva, az *elemzési módszerek alkalmazási területének* (mint például nemzetgazdasági, üzemgazdasági, pénzügyi, biztosítási, földrajzi, orvosi, mérnöki, meteorológiai, biológiai, mezőgazdasági, nyelvészeti, pszichológiai kérdéseknek) sajátosságait figyelembe véve. A tisztán formális tárgyalásnak látszólag megvan az az előnye, hogy a módszerek elsajátítása után azok végtelen sok konkrét helyzetben alkalmazhatók. Valójában azonban az alkalmazási terület szempontjait figyelembe vevő tárgyalás sem zárja ki a módszerek más területeken történő használatát, sőt hozzászoktat ahhoz, hogy a gyakorlati elemzések során a megfelelő elemzési eszközök kiválasztásához és az eredmények helyes értelmezéséhez egyszerre vegyük figyelembe a módszertani és tartalmi szempontokat. Elemzések végzésekor a statisztikus nem maradhat pusztán statisztikus, közgazdasági tartalmú elemzésnél közgazdásznak, biológiaiánál biológusnak stb. is kell lennie.

Ideális esetben a valamilyen szűkebb tudományterület művelői számára készült statisztikai kézikönyvekben, tankönyvekben egymással szervesen összekapcsolódnak a statisztikai és szaktudományi fogalmak és tartalmak, és külön tárgyalják a szaktudományi alkalmazás speciális kérdéseit. A közgazdászok számára készült könyvekben a statisztika leíró részét illetően talán a súlyozással és a csoportosított adatokból álló sokaságokból képzett mutatókkal kapcsolatos kérdések mellőzése említendő hiányosságként, hiszen az eltérő méretű megfigyelési és elemzési egységek rendkívül gyakoriak a gazdasági-társadalmi statisztikák során. A következtető statisztika tárgyalása a leíró résznél formálisabb, függetlenül a címben megjelenő különböző jelzőktől vagy az előszóban deklarált céloktól. Nem esik a kérdés súlyának megfelelő mértékű szó sem a vizsgálat tárgyát képező adatbázisok keletkezésének körülményeiről, sem a gazdaság- és társadalomstatisztikai adatokra annyira jellemző rendkívül sokoldalú, nem mintavételi hibákból eredő bizonytalanságokról (Morgenstern 1963, Kish 1989).¹

Különösen éles az ellentmondás a valószínűségszámítás és a következtető statisztika matematikai szigora, valamint a módszerek alkalmazhatóságának nem matematikai

* A tanulmány a Bolyai János Kutatási Ösztöndíj támogatásával készült.

¹ „A statisztika nagy része – tanfolyamok, könyvek, folyóiratok – elsősorban vagy kizárólagosan csak a statisztikai analízissel foglalkozik, és csak elvétve fordul elő, hogy a vizsgálatok tervezéséről is szót ejtenek; az irodalom becslésekkel és számítási eljárásokkal foglalkozik, és nem adatgyűjtéssel vagy kiválasztási eljárásokkal. A legtöbb mese a statisztikáról e szavakkal kezdődik: „adva van egy valószínűségi változó n különböző értéke”, és valódi adatok esetén itt szinte minden szó félrevezető” (Kish 1989, 3. o.).

szempontú kérdéseit mellékesen kezelő gyakorlat között. Míg a következtető statisztika valószínűségelméleti megalapozásának illusztrálására többnyire még a dobókockák és különféle szerencsejátékok szolgálnak, a következő lépésben már országok idősoraiból vagy több területegységre vonatkozó adatokból származó, rendkívül komplex jelenségekkel összefüggő, a bizonytalanságot nem, a matematikai értelmű véletlent azonban nélkülöző, egyszer előforduló, aggregált, súlyozott számsorokra alkalmaznak olyan módszereket, amelyeket korlátlanul megismételhető véletlen tömegjelenségekkel kapcsolatban fejlesztettek ki.

Tanulmányomban a valószínűségszámítás formális része, a szemléltetésre felhasznált példái és területi adatokra történő gyakorlati alkalmazása közötti ellentmondások vázlatos áttekintése a célom. Mindenekelőtt a területi statisztika fogalmának értelmezési különbségeit, valamint a területi statisztika sokaságait és megfigyelési egységeit tárgyalom. Ezután térek rá a valószínűségszámítás és a klasszikus következtető statisztika területi gazdasági-társadalmi adatokra történő alkalmazásának ritkán tárgyalt kérdésére. Nem foglalkozom az ettől sok szempontból eltérő problémákat felvető Bayes-i következtetésselmélettel.

A területi statisztika különböző értelmezései

Ahogy maga a statisztika szó is többjelentésű, a területi statisztika összetételt is többféle értelemben használják. Egyrészt maguknak a területi statisztikai adatoknak az elemzését, azon belül főként az országos szint alatti bontásban rendelkezésre álló adatok elemzését értik rajta. Másrészt területi statisztikaként lehet hivatkozni magukra a területi statisztikai adatokra, adatbázisokra. Harmadrészt nevezhetik területi statisztikának az általános statisztikai módszereknek egyszerűen a területi adatokra történő alkalmazását. Negyedrész a területi statisztika feladatát meg lehet határozni az általános statisztikai módszerek területi adatokra történő alkalmazása során fellépő, a területiség létéből következő kérdések tárgyalásaként, ami az előző, egyszerű és számos módszertani csapdát rejtő megközelítéstől lényegesen eltérő álláspont. Végül, ötdrész, területi statisztikaként lehet tekinteni az eredendően területi, térparamétert tartalmazó elemzési eljárások vizsgálatára, tárgyalására, leírására.²

Az első megközelítés szellemében született munkák két részre oszthatóak: egyrégiós és többrégiós elemzésekre. Az egyrégiós elemzések egy része nem foglalkozik a területi összehasonlítás kérdéseivel, pusztán országos szint alatti egyetlen területegység pontszerű elemzését adja. Ez a megközelítés általánosan is meglehetősen elterjedt: a területi statisztika ezek alapján régiók, megyék, kistérségek, települések statisztikai adatok segítségével történő leírásával foglalkozik. A többrégiós elemzések többnyire az egyes országok területi különbségeit tárgyalják, eltérő térfelosztás és területi skála mellett.

A második értelmezés alá tartoznak a területi adatbázisok keletkezésének, minőségének, tartalmának a gyakorlati elemzések során fontossá váló általános és konkrét kérdései (*Faluvégi 2002, Klonkai 2002, Nemes Nagy 2005b 9–45. o., Waffenschmidtné 2002*).

² A fogalom sokrétűsége miatt minden bizonnyal további jelentéstartalmak elkülönítése és ettől eltérő tipizálás is lehetséges.

A Területi Statisztika folyóiratban, ahol valamennyi megközelítés megjelenik, a túlnyomó többséget az első két típusba tartozó tanulmányok alkotják.

Álláspontom szerint a területi statisztika mint módszertudomány feladata a negyedik és ötödik pontban megfogalmazottak vizsgálata (például a súlyozás, aggregálás, módosítható területi egység problémája, speciális területi adattípusok és elemzési eszközök).³ Ezek legalább általános szintű ismerete elengedhetetlen ahhoz, hogy az első megközelítés kapcsán említett, alkalmazott területi statisztikának is nevezhető elemzések módszertanilag megalapozottak legyenek. Mivel területi adatokkal nemcsak a társadalmi és gazdasági élet elemzésekor, hanem különböző léptékben a természetföldrajzi, biológiai és más kérdések kapcsán is találkozunk, az általános területi módszertani kérdések több szaktudományban is felmerülnek, mindenhol speciális sajátosságokkal átitatva.

Mindamellet semmilyen problémát nem jelent, ha valaki a területi statisztikát említve az első, második és/vagy harmadik pont szerinti értelmezésre is gondol. Az eltérő értelmezések kapcsán két fontos nehézséget látok. Egyrészt, ha az első három értelmezés a negyedik és ötödik pontban megfogalmazott *feladatokról megfeledezve* történik, vagyis ha például valaki azzal a kimondatlan feltevéssel él, hogy a területegységek statisztikai elemzése során nem fordul elő semmilyen, a területiségből fakadó sajátos kérdés. Ez a többnyire implicit feltevés rendkívül káros, mivel a gazdasági-társadalmi adatok térbelisége olyan nehézségek elé állítja a statisztikát, amelyek a hagyományos analitikus statisztikai gondolkodás keretei között nem vetődnek fel. Ezek a nehézségek azonban a térbeliség kérdésének mellőzése esetén feltáratlanok maradnak, ez pedig a fel sem ismert módszertani hibák elkövetésének veszélyével jár.

A másik, az előzővel ellentétes előjelű nehézség abból adódik, ha a területi statisztikát leszűkítik a negyedik és ötödik pontra, elfeledkezve mind az alkalmazott területi kutatások fontosságáról, mind az általános statisztikával való kapcsolatáról, mind arról, hogy a társadalmi-gazdasági adatok lényegük szerint mindig bírnak területi tartalommal is. Ez a szemlélet a területi statisztika felesleges és túlzott önállósulásához, az általános statisztikától izolált részdiszciplínává válásához vezet (ez az angol szakirodalomban bekövetkezett, magyar szempontból azonban nem reális veszély; lásd Nemes Nagy József áttekintését a hazai helyzetről (Nemes Nagy 2001, Nemes Nagy 2005a)). Ez együtt jár az általános statisztikával való aszimmetrikus kapcsolatok kialakulásával: az általános statisztika szemszögéből a területi statisztika jelentéktelen, periferiális, sajátos módszertani kérdésekkel foglalkozó részterületnek tűnhet, miközben valójában magának az általános statisztikának az alapjai is érintettek vizsgálataiban.

A területi statisztika sokaságai és megfigyelési egységei

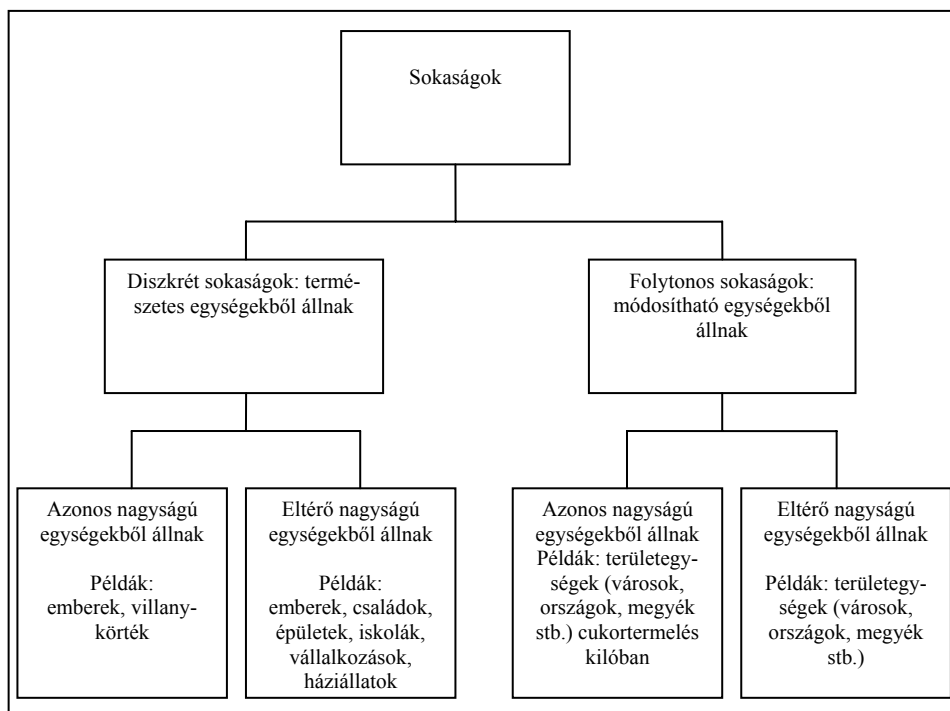
A sokaságok állhatnak egymástól természetes módon elkülönülő, valamint folytonos, egymástól önkényesen elhatárolt, módosítható egységekből. Az ugyanazon sokaságba tartozó természetes és módosítható egységek is lehetnek azonos és eltérő nagyságúak.

³ A területi elemzések alapjai című könyvem negyedik és ötödik fejezetét teljesen ennek a témának szenteltem. Az ott írtak ismerete könnyebben követhetővé teszi jelen tanulmány gondolatmenetét. Az ott, valamint könyvem további részeiben már tárgyalt témakörökre ebben a tanulmányban csak a minimálisan szükséges mértékben térek ki.

Az 1. ábrán mindezt néhány példa feltüntetésével láthatjuk. Ennek a kategorizálásnak a hangsúlyozását csak az indokolja, hogy a statisztikai kézikönyvek egy része egyáltalán nem vesz tudomást ezen eltérő típusokról és ennek elemzésbeli következményeiről, és minden adathalmazt természetes, azonos súlyú egységekből állónak tekint.⁴

1. ábra

A sokaságok típusai az alapelemek folytonossága és nagysága szerint



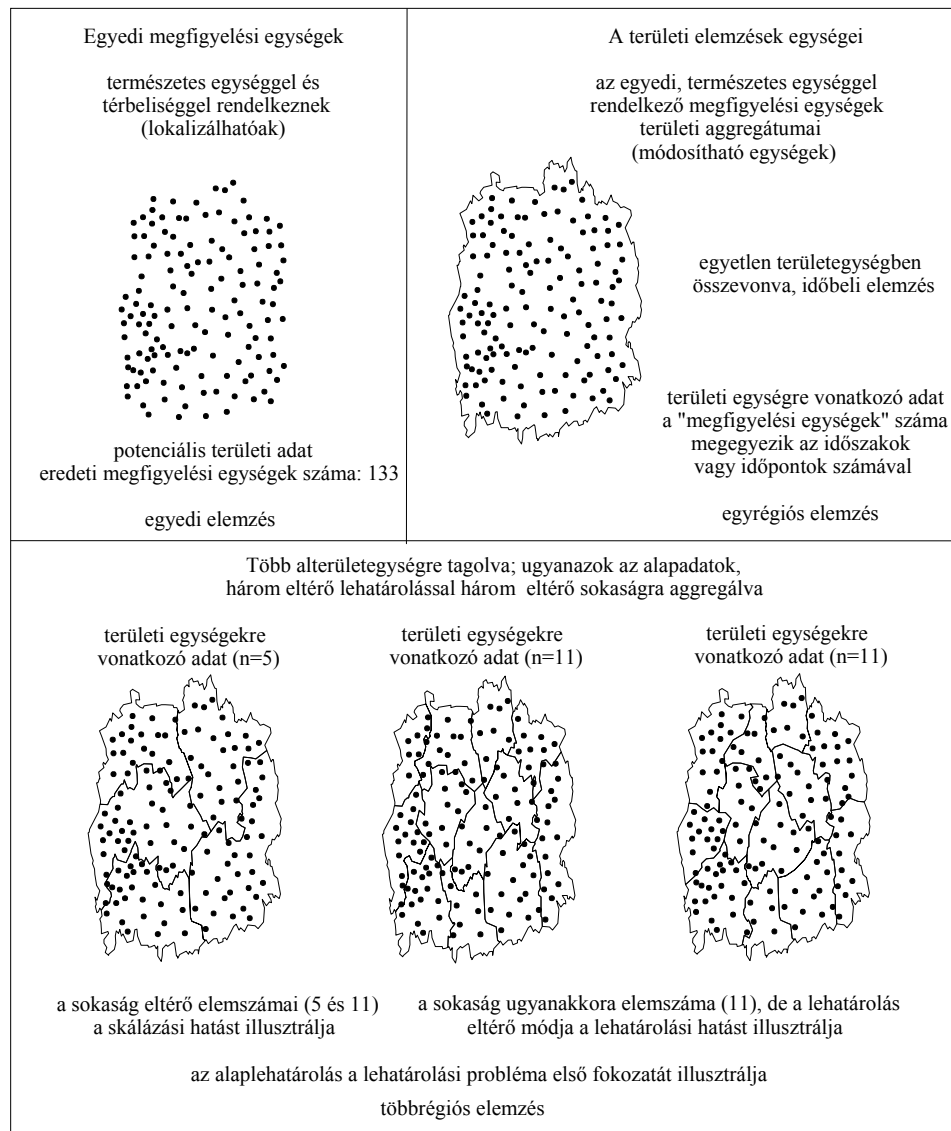
Az embereknek és a területegységeknek az azonos és eltérő nagyságú egységekből álló sokaságok közötti feltüntetése azt mutatja, hogy a sokaság objektumainak azonos és eltérő nagyságú volta nem abszolút érvényű, a sokaság elemeihez tartozó változtathatatlanság jellemző, hanem a vizsgálat céljától függ. A területi statisztika sokaságai a második kategóriába tartoznak: folytonos sokaságok, módosítható egységekből állnak, nem természetes módon adottak, akár egy társadalmi, akár egy természetföldrajzi vagy meteorológiai jellemzőről legyen szó. A területi statisztika társadalmi-gazdasági életre vonatkozó adatai a megfigyelt jellemzők területi ismérvei szerint lokalizált és csoportosított adatok. Ebben az értelemben tehát származtatottak, az *eredeti megfigyelési egység és az elemzési egység különbözik*: nem a területegység a megfigyelés elsődleges tárgya, hanem valami-

⁴ Ezt a kategorizálást is tovább lehetne finomítani, azonban a tanulmány témája szempontjából ezt a részletet nem gondolom szükségesnek.

lyen lokalizálható gazdasági-társadalmi objektum (emberek, vállalkozások, intézmények, szervezetek stb.) a hozzájuk kötődő jellemzőkkel együtt.⁵

2. ábra

Az egyedi megfigyelések és a területi adatok közötti kapcsolat



⁵ A kérdés ilyen tárgyalásmódja némileg leegyszerűsített, eltekint a különböző típusú adatok eltérő sajátosságaitól (például a területegység egészére értelmezhető minőségi adatoktól vagy az áramlásokra vonatkozó adatoktól); részletesebben lásd Dusek (2004), 93–106. o. Az adatok forrásának különbözőségeire sem térek ki.

Későbbi következtetésem nem vonatkoznak majd a területi sajátosságokkal is rendelkező, de egyedi adatokra és az azokkal végzett műveletekre, például a demográfiai mutatók számítására. Ezek csak területileg értelmezhetők, lévén egyedek összességére, vagyis csoportokra és nem egyedekre vonatkoznak, és így a módosítható területi egység problémája jelentkezik náluk (Dusek 2004). Mivel a sokaság alapelemei természetes egységek, ezért a statisztikai következtetésemélet szempontjából más, könnyebben kezelhető kategóriába tartoznak, mint amikor a sokaság alapelemei már maguk is területileg aggregált elemekből állnak. A későbbiekben ezekkel foglalkozom, vagyis nemcsak a szűk értelemben vett területi elemzésekkel (amelyek nem országos, hanem attól eltérő szinten zajlanak), hanem az összes egyrégiós elemzéssel is. Így megállapításaim érvényesek lesznek például az országokra vonatkozó idősorokkal vagy ágazati bontásban aggregált országos adatokkal végzett műveletekre is.

Mindez leolvasható a 2. ábráról és a hozzá tartozó szövegről is. A területi adatok az egyedi adatoktól számos más tulajdonság mellett abban is különböznek, hogy mértékükre a területi lehatárolás módja is hatást gyakorol. Azt a kérdést, hogy az így létrejött területi adatokat lehet-e „valószínűségi változóknak”, véletlen mintavételek, ellenőrzött kísérletek vagy megfigyelések eredményének tekinteni, a következő alfejezetekben tekintem át.

A valószínűségszámítás és a statisztikai következtetésemélet alkalmazásának területi

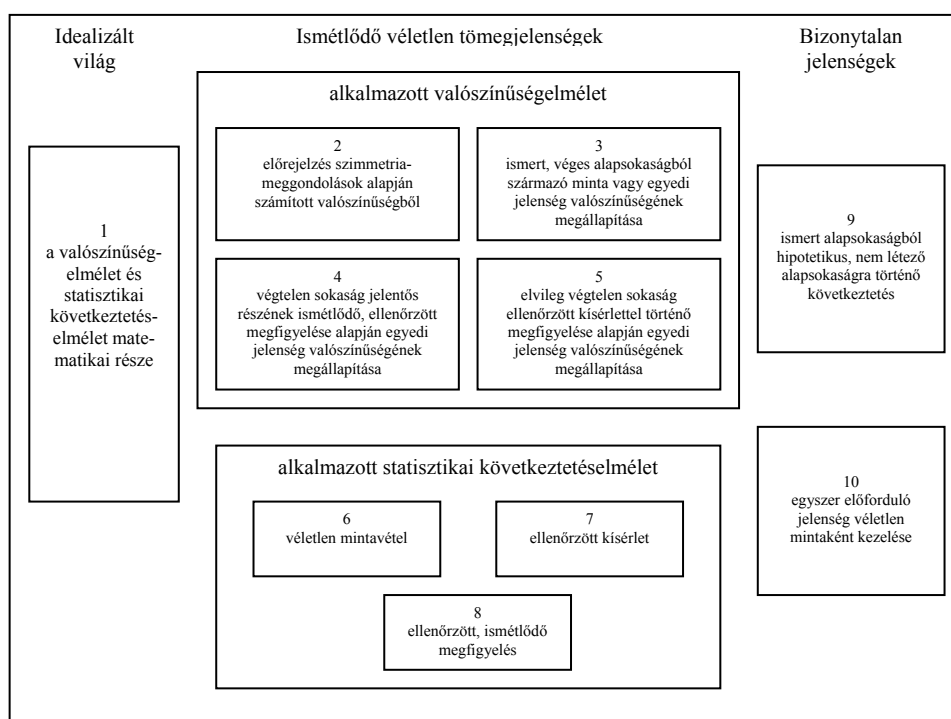
A valószínűségszámítást és a statisztikai következtetéseméletet tárgyaló kézikönyvek futó áttekintése is elég ahhoz, hogy az elmélet alkalmazásának egymástól karakteresen eltérő területeit különböztethessük meg. Ezek közül tízet jelenítettem meg a 3. ábrán. Sajnálatos módon a kézikönyvek ezen megkülönböztetések fontosságára nem kellő mértékben hívják fel a figyelmet, és olykor egymástól gyökeresen eltérő példák kerülnek egymás mellé; az alkalmazott kutatások pedig magától értetődően a módszerek használatával, és nem a módszerek eltérő alkalmazásaiból fakadó különbségekkel foglalkoznak. Ebben az alfejezetben az ábra bal oldalán és közepén látható nyolc területet tekintem át röviden, a maradék kettővel kapcsolatos problémákat – amelyek a területileg aggregált statisztikai adatokkal végzett műveletekkor is jelentkeznek – pedig hosszabban, a tanulmány hátralévő részében tárgyalom majd.

A véletlen tömegjelenségek matematikai elemzése olyan idealizált koncepciókra épít, mint a véletlen esemény, valószínűség, valószínűségi eloszlás, tetszés szerint megismételhető, standardizált körülmények között zajló kísérletek. Gondolatban végezhetünk végtelen hosszúságú kísérleteket, a valóságban ez természetesen nem lehetséges, mint ahogyan a tapasztalati jelenségek kapcsán nem létezik elméleti valószínűség, normális eloszlás, és semmilyen más folytonos eloszlás. Az idealizált koncepciókon alapuló matematikai elmélet tapasztalati relevanciáját az adja, hogy *képesek vagyunk absztrakt meghatározásait a tapasztalati jelenségekkel összekapcsolni*, így tételeit alkalmazni tudjuk a tapasztalatban megfigyelhető jelenségek leírására, elemzésére, magyarázatára és előrejelzésére. A *tapasztalat bizonyítja* például, hogy ugyanazon objektumra vonatkozó egyes mérési eredmények *diszkrét sokaságai jól közelíthetők a folytonos normális elosz-*

lással. Mindazonáltal a gondolkísérleteket és az elméleti sokaságokat lehet tekinteni a valószínűségszámítás alkalmazása első szintjének.

3. ábra

A valószínűségszámítás és a statisztikai következtetésemélet alkalmazásának eltérő területei



Az alkalmazott valószínűségelmélet az ismert nagyobb sokaságokból von le valószínűségi jellegű következtetéseket a sokaság kisebb részére vonatkozóan, többnyire jövőbeli események előrejelzése céljából. Az előrejelzések vagy szimmetriameggondolásokon, vagy múltbeli, ismert, tapasztalati gyakoriságokon alapszanak, amelyek jövőbeli érvényesülésére semmilyen elméleti garancia nincs. Utóbbi esetben az előrejelzés lehetőségét a *tapasztalati gyakoriságok stabilitása* teremti meg.

Szimmetriameggondolásokon alapuló valószínűségszámításra alapvetően kombinatorikus úton megoldható problémák egy részénél, másodsorban pedig nagyon egyszerű fizikai jelenségeknél van lehetőség. Utóbbira példa a kockadobás vagy egy érme feldobása: a szabályos kockának hat egyforma lapja, 12 egyforma éle és nyolc egyforma csúcsa van, tömegeloszlása egyenletes, ebből megállapítható, hogy valamennyi lapjára egyhatod valószínűséggel esik. A kombinatorikán alapuló valószínűségszámításkor a lehetséges végkimenetek és a hozzájuk tartozó elemi események megszámlálásán kívül szükség van az elemi események bekövetkezési valószínűségének ismeretére is, illetve legegyszerűbb esetben az elemi eseményeket azonos valószínűségűnek lehet tekinteni. A

lottóhúzásnál és egyes szerencsejátékoknál erre megvan a lehetőség, a sportfogadásnál és általában a hétköznapi életben viszont nincs.

A valószínűségszámítást tárgyaló könyvek a véletlen kísérlet meghatározása kapcsán többnyire tudatosan egybemossák a véletlen mintavételeket, az ellenőrzött kísérleteket és az ellenőrzött, ismételhető megfigyeléseket.⁶ Ennek az az oka, hogy a valószínűség általános matematikai háttere a három esetben azonos, miközben az információk megszerzésének gyakorlati különbségei jelentősek. Véletlen mintavételnél jól definiálható alapsokasággal állunk szemben; a véletlenséget – vagyis az alapsokaság elmei mintába kerülésének azonos, illetve ismert esélyét – azonban olykor csak feltételezni tudjuk, semmint ténylegesen és objektíven biztosítani. Az ellenőrzött kísérletek elvileg bármennyiszer megismételhetők, a kísérletek végzése emberi irányítás alatt áll. Ezzel szemben az ismételhető megfigyelések végzésekor korlátozó tényező, hogy a jellemző tulajdonságaikban azonosnak tekinthető jelenségeket nem mi idézzük elő, csak megfigyeljük eltérő időpontokban és/vagy helyeken. Véletlen mintavételre példa lehet a politikai közvéleménykutatás, ellenőrzött kísérletre egy termék méretének vagy tömegének a megmérése, ismételhető megfigyelésekre pedig az időjárás vagy egy étterem forgalmának a megfigyelése.

A valószínűségszámítás és statisztikai következtetéselmélet utóbbi két típusú jelenségre történő alkalmazásának alapját a kísérletek, illetve megfigyelések tömegszerűsége, ismételhetősége és véletlen jellege adja. Ez biztosítja, hogy a tapasztalati valószínűségi eloszlások megismerhetők legyenek, és ezek segítségével az ismétlődő egyedi jelenségekhez is számszerű valószínűség legyen rendelhető, valamint hogy véletlen mintavétellel következtethessünk ismeretlen alapsokasági jellemzőkre. Elvi nehézség a következtetéselmélet alkalmazásával kapcsolatban az első ábra közepén látható hét területen nem lép fel, a kísérletek és megfigyelések gyakorlati lebonyolítása azonban nagy gondosságot feltételez, és a vizsgálat tárgyának alapos ismeretét követeli meg.

Determinisztikus jelenségek, véletlen tömegjelenségek és nem ismétlődő, bizonytalan jelenségek

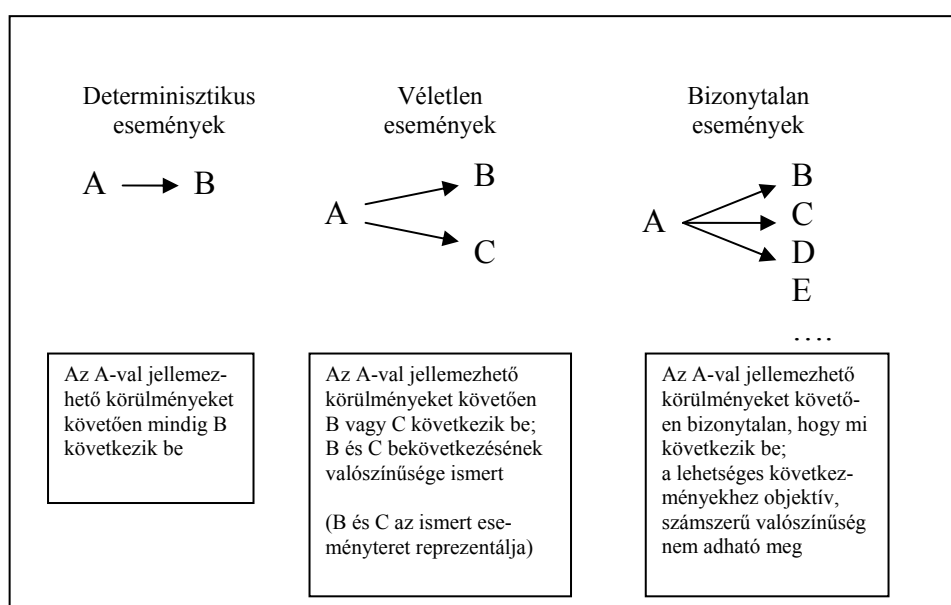
A determinisztikus jelenségek esetében az eseményeket befolyásoló tényezők megismerhetők, és ismeretükben az események kimenetele előre jelezhető. A véletlen tömegjelenségeknél az események mikroszinten nem determinisztikusak, de ismerjük az események lehetséges kimeneteleit, és ezeknek számszerű valószínűségét is meghatározhatjuk. A bizonytalan jelenségek a véletlen jelenségekkel csak annyi közös vonással rendel-

⁶ Három példa: „Minden valószínűségelméleti probléma valamely véletlen kísérlettel kapcsolatos. A „kísérlet” szót itt általánosabb értelemben használjuk. Kísérletnek nem csupán egy jelenség mesterséges előállítását nevezzük, hanem általában egy jelenség megfigyelését, függetlenül attól, hogy azt mi vagy rajtunk kívül álló okok hozták létre” (Prékopa 1980, 11. o.). „A továbbiakban véletlen jelenségen és kísérleten ugyanazt fogjuk érteni, magát a véletlen kísérlet szót pedig abban a tágabb értelemben használjuk, hogy nemcsak mesterségesen előállított jelenséget, hanem bármely természeti vagy egyéb jelenség megfigyelését értjük rajta” (Meszéna–Ziermann 1981, 23. o.). „A kísérlet kifejezést a valószínűségszámításban általánosabban értelmezzük, mint azt a köznap szöveghasználatban tesszük. E fogalomkörbe tartozik a jelenségek megfigyelése, a tudatosan nem tervezett események mérése is” (Dévényi–Gulyás 1988, 23. o.). Egyes könyvek ugyanakkor ki sem térnek a véletlen kísérlet meghatározására. Tanulmányomban a kísérlet kifejezést nem az előző három idézetnek megfelelő értelemben használom, mivel a mintavétel és megfigyelést nem értem bele.

keznek, hogy mindkét típusú egyedi esemény kimenetelével kapcsolatban bizonytalan-ság uralkodik. A három típusú jelenséget mutatja be a 4. ábra.

4. ábra

Determinisztikus, véletlen és bizonytalan események



A véletlen tömegjelenségek mellett a nem ismétlődő, bizonytalan jelenségeknek a valószínűségi számítás eszközeivel történő kezelése a jelenségek determinisztikus és véletlen eseményekre történő felosztásán alapul. Ez a felosztás történhet explicit, kifejezett módon, valamint ki nem mondottan, csupán az alkalmazásokból, példákban kitéve.⁷ Ez a dichotómia ugyanakkor a valószínűségi számítás elméletileg és filozófiaiilag sem indokolható alkalmazásaihoz vezetett. A nem determinisztikus eseményeket, amelyek kimenetele nem határozható meg egyértelműen a róluk rendelkezésre álló információk alapján, az előbb említett két csoportba lehet osztani. Egy részüket be lehet sorolni homogén események egy nagyobb osztályába, másik részüket homogén osztályokba sorolását komplexitásuk nem teszi lehetővé.⁸

A véletlen események kimenetelével kapcsolatban bizonytalan-ság uralkodik, az események azon osztályáról, amelybe az esemény tartozik, viszont mindent tudunk. Így az

⁷ Példa a kifejezett meghatározásra: „Az életben vannak tehát olyan jelenségek, amelyeknél a tekintetbe vett körülmények nem határozzák meg egyértelműen a jelenség létrejöttét, megvalósulásának mikéntjét, végbemenetelét. Ezeket a jelenségeket, eseményeket véletlen jelenségeknek, véletlen eseményeknek, vagy csak röviden eseményeknek nevezzük, megkülönböztetve őket az ún. szükségszerű, más szóval determinisztikus jelenségektől. (...) Az a tudomány, amely a véletlen jelenségek összefüggéseit, törvényszerűségeit kutatja, a valószínűségelmélet, valószínűségi számítás” (Meszéna–Ziermann 1981, 17–18. o.). Ez a meghatározás (mellyel máshol is találkozhatunk) messzemenően nem szerencsés, hiszen nem determinisztikus eseményekkel a valószínűségi számításon kívül foglalkozik például a nyelvészet, a pszichológia, a történelem, a földrajz, a biológia, az orvostudomány stb.

⁸ Nem tartozik tanulmányom tárgyához, hogy a kaotikus eseményeket hova soroljuk: önálló kategóriába, a determinisztikus, avagy a véletlen jelenségek körébe.

egyedi esemény kimeneteléhez objektív, számszerű valószínűség rendelhető. Ezzel szemben a bizonytalan események változatlan körülmények között mindössze egyszer fordulnak elő. Kimenetelük rendkívül sok, külön-külön is nehezen meghatározható tényező komplex kölcsönhatásán múlik. Ezeket vizsgálat tárgyává lehet tenni, és meg lehet becsülni a jövőbeli kimenetekre vonatkozó valószínűségeket; ezek azonban szubjektív karakterűek maradnak akkor is, ha számszerű formában fejezik ki őket. Ellenőrzésükre nincs lehetőség, mert változatlan körülmények között az esemény nem ismételtető meg.

A valószínűségszámítás eszközeivel elemezhető véletlen eseményekkel találkozunk például a következő területeken: termodinamika, szerencsejátékok, biztosítások, a genetika és öröklődés törvényei, vállalati készletgazdálkodás, banki pénzforgalom. Bizonytalan eseményekre hétköznapi példát szolgáltatnak a sportesemények eredményei⁹ és a parlamenti választások kimenetele; a tudomány teréről pedig a település- és úthálózat kialakulása, a telephelyválasztás, egy ország makrogazdasági mutatóinak az alakulása, a nyelvi változások terjedése. Ezek mind elemezhetők tudományos eszközökkel, a valószínűségszámítás alkalmazásának a feltételei azonban nem teljesülnek velük kapcsolatban.

A véletlen és bizonytalan események közötti gyakorlati határvonal ugyanakkor nem éles. Egyrészt gyakorlatilag megoldhatatlan az a kérdés, hogy hányszor kell ismétlődnie egy eseménynek ahhoz, hogy tömegszerűnek legyen tekinthető. Másrészt komplex, egyedi jelenségeknél is lehetőség van olyan kísérletekre, illetve megfigyelésekre, amelyeknél a változatlan körülmények követelménye nem tartható, de lehetőség nyílik az eseményt befolyásoló egy vagy kevés számú tényező változtatása mellett a maradék, nagyon nagy számú összetevő véletlenszerű elrendezésére, és így figyeljük meg, hogy a változtatott tényező(k) milyen hatást gyakorol(nak) vizsgálatunk tárgyára. Ilyen bonyolult kísérletekre nyílik lehetőség például a biológiában és az orvostudományban.

A véletlen és bizonytalan események megkülönböztetése a valószínűségszámítás Kolmogorov-féle formalisztikus megalapozása keretein kívül esik: „Ennek megfelelően az 1. §-ban a valószínűségi mező fogalmát mint meghatározott feltételeket kielégítő halmazrendszert értelmezzük. Az, hogy mik ezeknek a halmazoknak az elemei, a valószínűségszámítás tisztán matematikai felépítése szempontjából teljesen közömbös” (Kolmogorov 1982, 11. o.). A valószínűségelmélet gyakorisági megalapozása nem választja el egymástól a tiszta és alkalmazott valószínűségelméletet, nagyobb gondot fordít az ismételtetés követelményének explicit meghatározására. Enélkül a valószínűségszámítás csupán számok bizonyos csoportjainak tiszta matematikai tulajdonságaival foglalkozik. „A valószínűségelméletnek semmi kezdenivalója nincs az ilyen kérdésekkel: „Van-e valószínűsége annak, hogy Nagy-Britannia a közeli jövőben háborúba keveredik Egyiptommal?” Hasonlóan, a bibliai történetekre vonatkozó valószínű történeti igazság sem tartozik érdeklődési körébe. (...) Ezen kérdések, amelyek adekvát módon tárgyalhatók különböző nézőpontokból (politikai, szociológiai, történeti stb.), partikuláris helyzetekre vonatkoznak, és mint az „akkor és ott” kérdései, nem válaszolhatóak meg

⁹ Lehetetlen például az 1954-es labdarúgó-világbajnokság döntőjét azonos körülmények között kétszer lejátszani.

elméletünkkel. (...) A valószínűségelmélet tárgyát a változatlan körülmények között nagyon gyakran ismétlődő kísérletek vagy megfigyelések hosszú sorozatai képezik” (Mises 1964, 1–2. o.).

Területi adatok: populáció vagy minta?

A statisztikai sokaságokat hagyományosan populációkra és mintákra osztják fel. A populáció azt az alapsokaságot jelenti, amire vizsgálatunk tárgya, valamint annak következtetési irányulnak, a minta pedig ennek az alapsokaságnak egy részsokaságát.¹⁰ Egy adott sokaság mintának vagy populációnak tekintése részben a vizsgálat céljától függ. Ha érdeklődésünk középpontjában egy gimnázium száz érettségiző diákja áll, akik közül harmincnek van középfokú angol nyelvvizsgálója, akkor nincs létjogosultsága ezen adat (valamint valamennyi további számított mutatószám) konfidenciaintervallumáról vagy szignifikanciájáról beszélni. Ezek a százfős sokaságot leíró statisztikai értelemben teljes mértékben jellemezni fogják, minta hiányában nincs mintavételből fakadó hiba. Objektív értelemben ugyanazon adatokat azonban lehet egyidejűleg populációnak és mintának is tekinteni, például a száz érettségizőt „a gimnazisták”, „a 18 éves korosztály”, „a tanulók” időben és térben meghatározott populációjából származó mintának.¹¹

Mintákról populációkra vonatkozó, matematikai statisztikán alapuló következtetéseket számos egymástól koncepcionálisan eltérő esetben vonnak le. Ezekből négy fontos típust láthatunk az 1. táblázatban. Az első, fogalmilag egyszerű esetben akkor vonhatunk le következtetést a mintánkból, ha az alapsokaság elemeinek mintába kerülési valószínűsége ismert volt. A második esetben, amely többnyire természettudományos, műszaki és biológiai alkalmazásokban fordul elő, a ténylegesen megvalósult kísérleteket vagy megfigyeléseket a potenciálisan megvalósulható, ismeretlen kísérletek összességéből származó véletlen mintának tekintjük. A harmadik, már nagyobb elvi nehézségekkel járó esetben a mintából nem arra a populációra vonatkozóan szeretnénk következtetéseket levonni, amelyből a minta közvetlenül származik, hanem a származási populációval rokonnak tekintett, esetleg nagyobb, létező, de adatait tekintve nem, vagy hiányosan ismert populációra.¹² Negyedik esetben bármilyen populációt, számszerű információt képzeletben egy hipotetikus szuperpopulációból vett véletlen mintának tekintjük. Ez a koncepció a minta fogalmának újraértelmezéséhez vezetett, hiszen ennek alapján a populációk a szuperpopulációkhoz képest „véletlen minták” lehetnek, amikre alkalmazhatóak a statisztikai következtetésemélet eszközei, konfidenciaintervallumok adhatóak meg, és hipotéziseket lehet vizsgálni velük kapcsolatban.

¹⁰ A minta szót használják magukra a mintát alkotó mintaelemekre is; erről és kapcsolódó kérdésekről részletesebben lásd Podani (1997), 12–23. o.

¹¹ További kérdésekről lásd Kish (1989), 41–44. o. A példában szereplő száz érettségiző az említett nagyobb sokaságokból származó *véletlen* mintának ugyanakkor nem tekinthető.

¹² „Mintából származó eredményeket nem ritkán alkalmaznak olyan sokaságokra, amelyek az eredeti terv szerinti célsokaságtól különbözhetnek jellegükben, összetételükben, és ezen túlmenően térben és időben. Az ilyenfajta alkalmazások szükségessé és ennél fogva helyeselhetők. Példa: Michigan 1986-os mintájából vonhatunk le következtetéseket Ohio-ra vagy az Egyesült Államokra vonatkozóan 1986-ban, de akár 1990-ben is” (Kish 1989, 42. o.).

1. táblázat

A statisztikai következtetésemélet egyes alkalmazásai

Eset	Amiből következtetnek	Amire következtetnek
1	Véletlen minta	Meghatározható, létező alapsokaság
2	Kísérletek és megfigyelések	Potenciális alapsokaság
3	Véletlen minta	Kibővített alapsokaság
4	Populáció	Hipotetikus szuperpopuláció

A véletlen minta koncepciójának kiterjesztése először tehát az ellenőrzött kísérletekre és megfigyelésekre történt meg, majd a teljes körű felmérések populációját tekintették mintának egy hipotetikus szuperpopulációhoz képest, végül pedig a populáció–szuperpopuláció koncepción alapulva mindent véletlen mintaként kezdtek el kezelni, például a tetszés szerint létrejött, egyedi jelenségekre vonatkozó, sokszorosán összetett, súlyozott mutatók hányadosainak logaritmusát is. A populációk véletlen mintaként kezelése mögött meghúzódó indokok felderítése meglehetősen nehéz feladat, hiszen a valószínűség-számítás és a véletlen minták formalizált matematikai tárgyalását követően *átmenet nélkül* történik ezen eljárások alkalmazása *bármilyen sokaságra*, a mindennapi elemzési rutin számára pedig fel sem vetődnek ezek a kérdések.¹³

2. táblázat

A populációk véletlen mintaként kezelése mögött meghúzódó indokok

Sor-szám	Tömör megnevezés	Az érv lényege
1.	Területiség	A vizsgálat a térben máshol is elvégezhető lett volna, a konkrét adatok egyetlen területi reprezentációját jelentik csupán a lehetséges sokaságoknak.
2.	Időbeliség	A vizsgálat az időben máskor is elvégezhető lett volna, a konkrét adatok egyetlen időpontbeli vagy időszakbeli reprezentációját jelentik csupán a lehetséges sokaságoknak.
3.	Általánosíthatóság	Az adatok általánosíthatóságának eldöntésére van szükség.
4.	Valószínűségi kezelhetőség	A nem determinisztikus folyamatok sztochasztikus (valószínűségi) folyamatként kezelhetők.
5.	Magasabb minőség	A következtető statisztika fölényben van a leíró statisztikával szemben.
6.	Kiszámítható jelentőség I.	A szignifikanciavizsgálat segítség a kapcsolatok jelentőségének eldöntésében.
7.	Mérési hiba	Az adatok véletlen és független mérési hibával terheltek.
8.	Mérési mód	Az adatok többféleképpen mérhetők, a mérés aktuális megvalósulása csak egy a sok lehetőség közül.
9.	A területegység csak rész	A területegységre vonatkozó adatok mintavétel eredményei, nem a teljes sokaság megfigyeléséből származnak.
10.	Átrajzolhatóság	A területi egység módosítható, s ez a problémák forrása.
11.	Kiszámítható jelentőség II.	Az adatok nem tekinthetők mintának, de szórásuk és eltérő elemszámuk miatt szignifikanciavizsgálatok elvégezhetők velük.
12.	Randomizáció	A randomizációs eljárásokhoz nincs szükség mintára.
13.	A szokás hatalma	Minden sokaság minta, a szignifikanciaszint közlése automatikusan hozzátartozik az eredményekhez.

A 2. táblázatban 13 olyan indokot gyűjtöttem össze, amelyek felhasználhatók a populációk véletlen mintaként történő kezelésére. Az egyes érvek között természetesen számos átfedés fordul elő; különösen az első három érv tekinthető ugyanazon gondolat

¹³ A hazai és nemzetközi irodalomból könnyen lehetne sorolni a példákat; tanulmányomnak azonban nem konkrét tanulmányok, hanem egy általánosan bevett gyakorlat bírálata a célja, ezért ilyen tanulmányokra nem hivatkozom.

eltérő kifejezéseinek. Ezek érvényességének kérdésével nem általánosságban foglalkozom, hanem konkrétan a területi adatok kapcsán tárgyalom őket. Az alábbiakban néhány példát sorolok fel ezekre az adatokra vonatkozóan: bruttó hazai termék, egy lakosra jutó élelmiszer-fogyasztás értéke, százezer lakosra jutó házépítések száma, háztartások átlagos fogyasztási kiadásai, háztartások átlagos jövedelme, különböző árindexek stb.¹⁴ Ilyen és ehhez hasonló típusú adatokkal gyakran lehet találkozni, sőt, ezek előfordulása gyakoribb a gazdasági elemzésekben az eredeti megfigyelési egységekre vonatkozó adatoknál. Az elemzési egységek lehetnek ekkor települések, országos szint alatti terület egységek, országok együttese, valamint ezen adatok egy területegységre vonatkozó havi, negyedéves, éves idősorai is.

A felsorolt érvek nem a szigorú¹⁵ matematikai elméletben gyökereznek; a területi adatok vonatkozásában a tizenegyediket és tizenkettediket leszámítva¹⁶ egyiket sem osztom, a többségét félreértésen alapulónak tartom. Véleményem szerint a gazdasági és társadalmi életre vonatkozó területi statisztikai adatok mindegyike adott földrajzi helyen, adott időpontban vagy időszakban megfigyelt sokaságra vonatkozó, számszerű formában rendelkezésre álló *történeti információ*. A statisztikai adatok a valószínűségi számítás eszköztárának bevetése nélkül teljes mértékben megfelelnek annak a célnak, hogy adott társadalmi-gazdasági jellemzőket számszerűen leírjanak. Jelentésük nagy részét abból nyerik, hogy időben és térben változó nagyságúak; állandóságuk vagy nagyfokú időbeli és térbeli stabilitásuk mellett általában csökken érdekességük, mivel ekkor többnyire az adatok megfigyelésétől függetlenül is közismert általános szabályszerűségről van szó. A statisztikának ettől teljesen eltérő alkalmazásával találkozunk a természeti jelenségek tanulmányozása kapcsán: ekkor a statisztikai következtetéseméletet a mesterségesen elszigetelt, vagy standardizált, összehasonlítható körülmények között végzett, szabványosított mérési eredmények közötti szabályosságok megállapítására lehet felhasználni. A kvantitatív formájú általánosíthatóság alapját a természet alkotóelemeinek minőségi és viselkedési egyöntetűsége jelenti.

A területi adatokra a térben és időben potenciálisan végtelenül sokféleképpen realizálódni képes adatokból származó véletlen mintaként tekinteni meglehetősen furcsa, szinte misztikus jellegű hipotézis.¹⁷ Ezt a hipotézist véleményem szerint elméleti vagy tapasztalati úton sem indokolni és alátámasztani, sem cáfolni nem lehetséges. Célszerűsége ugyanakkor több mint kétséges, hiszen egy önkényes, a jelenségek térbeli és időbeli eloszlására vonatkozó igazolhatatlan feltevessel terheli és kuszálja össze a leíró, történeti vizsgálatokat is. Nehézséget okoz az is, hogy a vizsgált adatok nem véletlen mintavétel eredményei: ez ismét nem egy szilárd tudományos bázison alapuló feltevés, hanem egy igazolhatatlan elemzési gyakorlat.¹⁸

¹⁴ További példák találhatók Dusek (2004) 105. oldalán, a 19. táblázatban.

¹⁵ Szigorún azt értem, hogy pontosan definiált, világos fogalmakat használó és az állításokat (az alapigazságokat leszámítva) bizonyító.

¹⁶ Ez a két pont a felsorolás teljesebbé tétele érdekében került bele az érvek közé, hiszen ezek alapján nem a populációk mintaként kezelése miatt végeznek szignifikanciavizsgálatokat; lásd később.

¹⁷ Ahhoz, hogy az ilyen adatokat véletlen mintaként kezelhessük, Feller következő példájához hasonlólt kellene elgondolnunk: „Tekintsünk végtelen sok Földet, válasszunk ki véletlenszerűen egyet közülük...” (Feller, 1978, 20. o.).

¹⁸ A statisztikai kézikönyvek általában nem foglalkoznak sokat az értelmetlen szignifikanciapróbákkal. Kivétel Freedman és társai könyve (Freedman et al. 2005, 604–624. o.), amely a matematikai háttér ismertetése helyett többet foglalkozik az

Az első két érveléssel kapcsolatban Gould a következő konklúzióra jut: a földrajzi kutatásban „nagyon gyakran az egész populáció vizsgálható, a szignifikanciavizsgálatok eredményét azonban mégis lelkiismeretesen közlik. Azonban az egész populációt vizsgálva, mire következtetünk eredményeinkből? Ekkor olyan kifogásokkal élünk, próbálkozva a populációra vonatkozó tesztek használatát igazolni, hogy „egy adott időpontbeli mintát” vettünk, vagy „a minta egy nagyobb populációt képvisel, amely a vizsgálatunk tárgyát képező régiókon kívüli térségekből áll”. Ezek az érvek nagyon gyengén hangzanak a véletlen mintavétel szigorú feltevéseinek fényében. A megfigyelt mérőszámok, akár a korrelációról vagy a regresszióról legyen szó, *leírják* a változók közötti kapcsolat szorosságát. A hagyományos szignifikanciatesztek elvégzésének gondolata, hogy valamilyen következtetésre jussunk egy nagyobb alapsokaságra vonatkozóan, amikor nagyobb alapsokaság nem létezik, enyhén szólva különösnek tűnik” (Gould 1970, 442. o.).

Rátérve a negyedik érve, mint korábban írtam, a determinizmus hiánya nem egyenlő a folyamatok sztochasztikus, valószínűségi jellegével.

A következtető statisztikának a leíró statisztikával¹⁹ szembeni fölénye azon az előítéleten alapul, miszerint a tudományos módszereknek létezik a módszerek alkalmazási körülményeitől független hierarchiája. Ezzel az elképzeléssel és a negatív hatásaival A területi elemzések alapjai című könyvemben annak első mondatától²⁰ kezdve hosszan foglalkoztam.

A hatodik érvek, amely szerint a szignifikanciavizsgálatok használhatók az egyes kapcsolatok jelentőségének megítélésére, alapvetően az a gyenge pontja, hogy a statisztikai szignifikancia és a tudományos szignifikancia közé nem lehet egyenlőséget tenni. Emellett a szignifikanciaszint nagyságának objektív eldöntéséhez ismerni kellene a nullhipotézis hibás visszautasításának költségét. Ez a kritérium általában elérhető a változatlan körülmények között könnyen megismételhető hipotézisvizsgálatoknál, így a minőségvizsgálatoknál és a kísérleti kutatások egy részénél, azonban a területi adatok esetén nem. Az is közismert, hogy önmagában a minta méretének növekedésével növekszik a statisztikailag szignifikáns kapcsolatok aránya; az pedig meglehetősen furcsának tűnne, ha a mintaméret határozná meg egyes kapcsolatok tudományos jelentőségét.

A mérési hiba érv nem veszi figyelembe a természettudományos mérési eredmények és a társadalmi-gazdasági adatok alapvetően különböző jellegét. A természettudományos méréseknél a mérési hiba megadása hozzátartozik az eredmények közléséhez; ez akkor maradhat el, ha mértéke közismert. Az ugyanazon objektumra vonatkozó mérések nagy részének normális eloszlással történő közelítése nem csupán feltevés, hanem tapasztalati-
lag bizonyítható az azonos körülményű mérések ismételt elvégzésével. Ezzel szemben a

alkalmazással kapcsolatos tartalmi elemekkel. Az ebben a könyvben a szignifikanciavizsgálatokról írtak megerősítik és számos példával támasztják alá az általam leírtakat. A területi kutatásokkal kapcsolatban az értelmetlen szignifikanciapróbákról lásd elsősorban Gould többször idézett alapvető tanulmányát (Gould 1970), valamint Meyer (1972) rövid, lényegre törő kommentárját és Summerfield (1983) tanulmányát. A következtető statisztika területi adatokra történő alkalmazása mellett érvel Court (1972), Gudgin-Thornes (1974), Fotheringham-Brunsdon (2004).

¹⁹ Olykor a statisztika szót kimondva kizárólag a következtető statisztikára gondolnak: „Ha viszont mindenkit, aki az egyetem hallgatója, a minta elemének tekintünk, azzal nem mintát, hanem – úgy látszik – a populációt vizsgáltuk meg! A feladat ekkor már nem statisztikai, hiszen annak lényegéhez tartozik a *mintából való következtetés* a populációra” (Hajtmann, 1968, 214–215 o.). Ezt a szóhasználatot indokolatlannak és zavart keltőnek tartom.

²⁰ „A kutatás során alkalmazott módszereknek minden tudományos vizsgálódás során a kutatás tárgyához kell igazodniuk” (Dusek, 2004, 5. o.).

társadalmi és gazdasági adatok nem természettudományos értelemben vett mérés eredményei, hanem vagy adminisztratív folyamatok melléktermékei vagy statisztikai adatgyűjtés, felmérés eredményei. Amikor a munkanélküliség vagy az árszínvonal változásának „méréséről” és „mérési hibájáról” beszélünk adatgyűjtés (vagy felmérés) és adatgyűjtési hiba helyett, akkor tisztában kell lenni azzal, hogy csak egy metaforát használunk, mégpedig szemmel láthatóan megtévesztő metaforát.

A felmérés tárgyra, a társadalomra és gazdaságra állandó változás jellemző, lehetetlen az egyes felméréseket újra és újra változatlan körülmények között elvégezni. A társadalmi-gazdasági adatok tartalmaznak az adatfelvétel módjából fakadó sokoldalú hibákat, azonban ezek nem véletlen, független eloszlású hibák. A jelentős adatoknak csak kis része keletkezik egyszerű megszámlálással, a mutatók nagy részének előállítására súlyozási, definíciós, klasszifikációs és egyéb módszertani döntéseket igényel (valamennyi korábbi példám, amit a 2. táblázat előtt soroltam fel, ilyen volt).

A mérési mód önkényességének érvéről hasonlóak mondhatóak el. A különböző felmérési lehetőségekből, például a munkanélküliek számának és arányának megállapításakor, aligha hihető, hogy a statisztikusok véletlenszerűen választanak ki a ténylegesen alkalmazott módszert. A felmérés módját a probléma és a megismerés tárgya, valamint egyéb tudatos megfontolások határozzák meg.²¹

A területi adatok jelentős része (mikrocenzus, háztartásstatisztikai adatok, valamint számos gazdaság- és társadalomstatisztikai jellemző) a megfigyelési egységekből vett mintavétel aggregálással, átlagolással nyert eredménye. Ezért esetleg lehetne amellett érvelni, hogy mivel az eredeti megfigyelési egységekből mintát vettünk, így a területi adatok is tekinthetők mintának. Ennek a gondolatmenetnek azonban több gyenge pontja létezik. A statisztikai kiadványokban megjelenő adatok egyedi megfigyelései nem elérhetők a kutatók számára, gyakran még azt az információt sem tudják meg, hogy az adat mintavételből származott-e, és ha igen, milyen típusú mintavételből. A gazdasági adatok nagy része nem véletlen mintavétel eredménye, hanem többnyire a legnagyobb szervezetek teljes körének és a kisebb szervezetek töredékének a megfigyelésével keletkezik. Amennyiben ilyen adatokkal vizsgálatokat kíván végezni egy kutató, az egyes terület-egységekre vonatkozó adatokat az adott területegység populációjára vonatkozó akkurátus becslésnek foghatja fel, és az eredmények értelmezésekor figyelembe veszi, végkövetkeztetéseiben óvatosabbá teszi, hogy az alapadatok mintavétel eredményei. Magukat a területegységekre vonatkozó átlagadatokat azonban már nem tekintheti véletlen mintavétel egyedi eredményeinek és így az átlagadatokat összességét véletlen mintának, és nem végezhet ennek megfelelő következtető statisztikai számításokat.

A módosítható területi egység problémája és a mintavétel közé vont párhuzam azon alapul, hogy a tényleges adatok a gyakorlatilag végtelennek tekinthető, potenciálisan szóba jöhető lehatárolások közül egyetlen egy lehatárolás keretében ismertek, amely így véletlen mintának tekinthető. Az analógia azonban megint nélkülöz minden tartalmi hasonlóságot, mivel a lehatárolás módjában ismét nincs véletlen elem.

²¹ A populációk mintaként kezelése melletti érvként a mérési hibára és módra hivatkozik (mindenféle indoklás nélkül) Gudgin-Thomes (1974).

A tizenegyedik és tizenkettedik pontban megfogalmazott érvekről azt gondolom, hogy ellentétben a többiekkel, elfogadhatóak. Bár tudjuk, hogy adataink nem minták és nem sztochasztikus folyamatok eredményei, a szignifikanciavizsgálat egyfajta (a grafikus ábrázoláshoz, kilógó adatok vizsgálatához hasonló) választ képes adni arra, hogy az eredmények bizonyos értelemben rendkívüli adatokból származnak-e, a rendkívüliséget az adatok szórásával megragadva. Azt is tudjuk, hogy ugyanaz a mutatószám eltérő értelmezést igényelhet az adatok elemszámának és szórásának függvényében. A randomizációs eljárásokkal pedig valamennyi mutatószámot egységesebb (bár teljes mértékben nem azonos) skálán tudjuk elhelyezni, és nagyobb mértékben tudjuk összehasonlítani egymással az eltérő nagyságú és területi konfigurációjú adatokból számított azonos mutatószámokat. A szignifikanciavizsgálatok ilyen értelmű használata azonban a leíró statisztikák részének, kiegészítésének tekinthető, és nem az alapsokaságra következtetés klasszikus eszközeinek.

Az utolsó érvelés valójában nem gondoltam komolyan, csupán egy hibás mentalitásra kívántam utalni vele. Egyetlen tudományos eljárást sem emelhet általános érvényűvé és alkalmazását nem teheti megkérdőjelezhetetlenné az a tény, hogy léteznek olyan problémák, amelyek vizsgálatában sikeresen használható.

Összegzés

Tanulmányomban általános képet kívántam nyújtani a valószínűségszámítás és következtető statisztika területi alkalmazásaival kapcsolatos, a szakirodalomban elvéve megjelenő kérdésekről. Igyekeztem bemutatni, hogy a területi adatok véletlen mintaként, illetve sztochasztikus folyamatok eredményeként kezelése indokolatlan, igazolhatatlan, csak az elemzési gyakorlat által szentesített konvenció. A tárgyalt kérdések nem az alkalmazott kutatásoktól elszigetelt elméleti érdekességek, hanem a tényleges elemzési gyakorlat legégetőbb problémáit jelentik.²² A valószínűségszámítás és a következtető statisztika rendkívül hatékony eszköztárat ad a kezünkbe azon tudományos problémák megoldásához, amelyeknél alkalmazásuk feltételei biztosítottak. A területi gazdasági-társadalmi statisztikai adatok nem tartoznak ebbe a kategóriába, esetükben a leíró statisztikai módszerek használata indokolt.

IRODALOM

- Court, A.* (1972): All statistical populations are estimates from sampling. *Professional Geographer*, 24. évf. 2.
- Dévényi Dezső–Gulyás Ottó* (1988): *Matematikai statisztikai módszerek a meteorológiában*. Tankönyvkiadó, Budapest
- Dusek Tamás* (2004): A területi elemzések alapjai. *Regionális Tudományi Tanulmányok 10.* ELTE Regionális Földrajzi Tanszék – MTA–ELTE Regionális Tudományi Kutatócsoport, Budapest

²² Jól látszik ez a területi autokorrelációval kapcsolatos széles körű félreértéseken, amelyek szerint a területi minta elemei közötti autokorreláció bármilyen problémát jelentene. Problémát csak akkor jelentene, ha a területi elemzési egységekhez kötődő adatokat véletlen mintavétel eredményének tekinthetnénk. Az autokorreláció mint probléma paradox voltáról lásd Gould (1970) kitűnő tanulmányát (az autokorrelációra vonatkozó lényegyet idézem könyvemben (Dusek 2004, 203. o.)).

- Faluvégi Albert* (2002): A területi statisztikai adatbázisok kialakulása, helyzete, fejlesztési lehetőségei. *Területi Statisztika*, 5. (42.) évf. 4.
- Feller, W.* (1978): Bevezetés a valószínűségszámításba és alkalmazásaiba. Műszaki Könyvkiadó, Budapest
- Fotheringham, A. S.–Brunsdon, C.* (2004): Some thoughts on inference in the analysis of spatial data. *International Journal of Geographical Information Science*, 18. évf. 5.
- Freedman, D.–Pisani, R.–Purves, R.* (2005): *Statisztika*. Typotext, Budapest
- Gould, D.* (1970): Is *statistix inferens* the geographical name for a wild goose? *Economic Geography*, 3.
- Gudgin, G.–Thornes, J. B.* (1974): Probability in geographic research: Application and problems. *The Statistician*, 23. évf. 3., 4.
- Hajtmann Béla* (1971): Bevezetés a matematikai statisztikába. Akadémiai Kiadó, Budapest
- Kish, L.* (1989): Kutatások statisztikai tervezése. Statisztikai Kiadó Vállalat, Budapest
- Klonkai László* (2003): Gondolatok a területi statisztikai információs rendszerről 50 évvel a területi szervek létrehozása után. *Területi Statisztika*, 6. (43.) évf. 1.
- Kolmogorov, A. N.* (1982): A valószínűségszámítás alapfogalmai. Gondolat, Budapest
- Meszéna György–Ziermann Margit* (1981): Valószínűségelmélet és matematikai statisztika. Közgazdasági és Jogi Könyvkiadó, Budapest, 1981
- Meyer, D. R.* (1972): Geographical population data: statistical description not statistical inference. *Professional Geographer*, 24. évf. 1.
- Mises, R. von* (1964): *Mathematical theory of probability and statistics*. Academic Press, New York, London
- Morgenstern, O.* (1963): On the accuracy of economic observations. Princeton, Princeton University Press
- Nemes Nagy József* (2001): Másfél évtized után ismét a hazai területi kutatások mennyiségi elemzési kultúrájáról. *Regionális Tudományi Tanulmányok* 6.
- Nemes Nagy József* (2005a): Nemzetközi és hazai tendenciák a területi elemzésben. *Területi Statisztika*, 8. (45.) évf. 1.
- Nemes Nagy József* (szerk.) (2005b): Regionális elemzési módszerek. *Regionális Tudományi Tanulmányok* 11. ELTE Regionális Földrajzi Tanszék – MTA–ELTE Regionális Tudományi Kutatócsoport, Budapest
- Podani János* (1997): Bevezetés a többváltozós biológiai adatfeltárás rejtelmeibe. Scientia Kiadó, Budapest
- Prékopa András* (1980): Valószínűségelmélet műszaki alkalmazásokkal. Műszaki Könyvkiadó, Budapest
- Summerfield, M. A.* (1983) Populations, samples and statistical inference in geography. *Professional Geographers*, 35. évf. 2.
- Waffenschmidt Jánosné* (2002): Minőség a területi statisztikában. *Területi Statisztika*, 5. (42.) évf. 5.

Kulcsszavak: területi statisztika, minta és alapsokaság, mintavétel, statisztikai következtetésemélet, módszertan, valószínűségszámítás.

Resume

The aim of the paper is to examine the questions arising from the application of probability theory and statistical inference to spatial data. After presenting the main characteristics of spatial data, the paper criticizes the relevance of procedures of classical statistical inference to population data in spatial analysis. Population data are perfectly suitable for describing spatial processes but the mechanical procedures of statistical significance tests on such data are not anchored in rigorous statistical theory, it is only a practice based on convention.