Alkalmazott Matematikai Lapok **29** (2012), 67-100.

ÉRZÉKENYSÉGVIZSGÁLATOK A STATISZTIKAI ELJÁRÁSOKBAN

TAKÁCS SZABOLCS

Bizonyos matematikai eljárások fontos, kihagyhatatlan része az úgynevezett érzékenységvizsgálat. E vizsgálat során arra vagyunk elsősorban kíváncsiak, hogy a különböző inputadatok megváltozása következtében feladatunk megoldása (eredménye) milyen mértékben változik – illetve milyen viselkedést mutat. Érdekes kérdés lehet az is, hogy milyen input változások esetén nem módosul a megoldás, ahogyan az is, hogy mely input adatok lesznek nagyobb, mely input adatok pedig kisebb hatással a kimeneti adatok változásaira.

A statisztikai kérdésfelvetések során más és más területeken eltérő fogalmi háttérrel vizsgálhatjuk ezt a jelenséget. Ahogy majd látni fogjuk: mást jelent az érzékenység a becsléselméletben, mást egyes hipotézisvizsgálati módszereknél és megint mást jelent az elsősorban modellezésre használt eljárások esetében.

Cikkünkben nem kívánunk teljes betekintést nyújtani e vizsgálati módszerek széles tárházába és alkalmazásába – pusztán arra vállalkozunk, hogy felvázoljuk e terület széles alkalmazási spektrumát. Szeretnénk továbbá felhívni a figyelmet ezen – általában kiegészítő – eljárások fontosságára.

A cikkben nem célunk új matematikai állítások megfogalmazása – sokkal inkább bizonyos kérdések felvetése, melyekre a cikk megírása során tett kutatómunkánk kapcsán nem találtunk megnyugtató válaszokat.

1. Bevezető

A statisztika az egyik leginkább alkalmazott területe a matematikának: számtalan területen jelen van kutatási eszközként, alkalmazói pedig nem feltétlenül matematikusok. Például Prékopa [37] műszaki alkalmazásokat tartalmazó könyve is segédanyagként szolgálhat azok számára, akik nem matematikusként, de műszaki területeken kívánják a statisztikát alkalmazni. Azonban a könyv nem tartalmazza (mert nem is tartalmazhatja) a tudományterület néhány olyan sajátosságát, melyek az utóbbi évtizedekben kezdtek teret nyerni, hiszen jellemzően mind számításigényes eljárások.

Számos tudományterület foglalkozik azzal a kérdéssel, hogy egyes kísérletek végeredménye milyen mértékben, illetve milyen módon függ a bemeneti adatoktól.

Mely bemeneti adatok azok, melyekre nézve a kísérlet stabilitást mutat és melyek azok, amelyek esetleg az egész kísérlet érvényességét veszélyeztetni tudják?

A kísérletek érvényessége, eredményessége – ha úgy tetszik, a kimeneti adatok bemeneti adatoktól való érzékenysége – fontos kutatási sarokpont, melyre nem minden kutatási folyamat során jut elég figyelem, vagy ha úgy tetszik, nem is feltétlenül vizsgálat tárgya egyes kísérletekben.

Egyre gyakrabban olvasni olyan tudományos, vagy tudományt népszerűsítő cikkeket, ahol a bemeneti adatokkal való, nem eléggé körültekintő bánásmód téves, vagy legalábbis nem igazolható következtetések levonására adott okot. Erre lehet példa LeVay, a Science folyóiratban megjelent tanulmánya [31] – melyet azóta többen is megkérdőjeleztek, illetve eredményeit cáfolták. A szerző e cikkében HIV-fertőzött homoszexuális és nem HIV-fertőzött, heteroszexuális férfiakat vizsgált haláluk után, és agyi struktúrájukban markáns eltérésekre bukkant. Azonban a halál közvetlen okaként szolgáló betegséget „elfelejtette" vizsgálata tárgyává tenni – később kiderült, hogy az eltérésekért nem a szexuális beállítottság, hanem maga a HIV-vírus a felelős (lásd pl. Bayne és társai tanulmányát, melyben kifejtik, hogy többek között a HIV-vírus okozta elváltozások kiszűrése után semmifajta hatását nem tudták kimutatni a szexuális orientációnak).

A kérdés persze úgy is felvethető, hogy ebben az esetben a figyelmetlenség okozta-e az adatokban való különbségek hibás értelmezését – vagy egy olyan szokásjog esetleges megléte, mely a bemeneti adatok különbségeiben való alaposabb vizsgálódás hiányát eredményezhette?

Ugyanis statisztikai oldalról persze úgy értelmezhető a kérdés, hogy a HIV-státusz figyelmen kívül hagyása, vagy ha úgy tetszik, nem megfelelő kezelése olyan különbségeket eredményezett a kimeneti adatokban, melyekből az azóta megjelent tanulmányok szerint, téves következtetés sikerült levonni.

Így persze felvetődik a kérdés: a statisztikai eljárásoknál az érzékenység (a bemeneti adatok változékonyságának, vagy változásának a kimeneti adatok vizsgálatának fényében) maguknak a módszereknek sajátja, vagy külön is érdemes rájuk kitérni?

Cikkünkben megpróbáljuk néhány statisztikai terület esetén az „érzékenységvizsgálat" analóg fogalmait bemutatni, illetve kitérni a fenti kérdésre: a statisztikai eljárásoknak e vizsgálat sajátja kellene, hogy legyen? Vagy netán a különböző eljárásoknál – a bemeneti adatok bizonyos anomáliái vagy tulajdonságai esetén – kiegészítő vizsgálatokra lenne szükség?

A cikkben három nagyobb egységet különíthetünk el. Az első nagyobb fejezetben az egész cikk során használt statisztikai módszerek rövid, áttekintő bemutatását olvashatjuk. Külön kitérünk a becsléselmélet és a hipotézisvizsgálatok főbb pontjaira. A második rész az érzékenységvizsgálatokról szól a statisztikai módszerek alkalmazása esetében. 3 nagyobb részfejezetre bontottuk a kérdést: érzékenységvizsgálatok a becsléselméletben, ahol a módszereket részint a mintanagyság, részint pedig a vizsgált paraméterek esetére osztályoztuk. A második részfejezetben a hipotézisvizsgálatok esetét tárgyaljuk, külön kitérve bizonyos speciális módszerekre, nem hagyományos statisztikai eljárásokra. A harmadik részfejezetben

egy biostatisztikai módszert mutatunk be – egy konkrét példán is végigvezetve az olvasót.

2. Statisztikai bevezető

E fejezetben bemutatjuk azokat a statisztikában használt definíciókat, illetve fogalmakat, melyekre a cikk olvasása során szükségünk lehet. Alapvetően három területre koncentrálva gyűjtöttük össze ezeket a formulákat: egyik oldalról a becsléselmélethez kapcsolódó eljárásokra és elnevezésekre koncentrálunk, másik oldalról pedig az ezzel erősen összekapcsolható hipotézisvizsgálati fogalmakat is szeretnénk bemutatni.

A harmadik terület valójában algoritmusok gyűjteménye: szimulációs technikák, melyeket statisztikai eljárások során alkalmazhatunk. Egy szimulációs módszert mi is bemutatunk e fejezet végén.

2.1. Becsléselmélet

Az alább található bevezető definíciók lényegében bármely, bevezető statisztikai könyvben, jegyzetben megtalálhatók. Angol nyelven Lehmann pontbecslésekről szóló könyve [29], magyarul akár Borovkov [9], akár Bolla és Krámli [6] frissebb kiadású könyvei említhetők, illetve egyetemi jegyzetek formájában szintén magyar nyelven Prékopa [37] vagy Mogyoródi [33] munkái lelhetők fel.

A becsléselmélet alkalmazása során az alábbi statisztikai kérdésekre keressük a választ.

Legyen adott egy X véletlen változó és egy $(\Omega, \mathcal{A}, P)$ valószínűségi mező.

Az $X : \Omega \to \mathbb{R}$ valószínűségi változónk adott θ paraméterét szeretnénk megbecsülni. E kérdésfelvetésre azért is szükség lehet, mert a becslési eljárások számos módon függhetnek vizsgálatunk tárgyát képező paramétereinktől.

Amiben minden becslési eljárás megegyezik: veszünk egy $X_1, \ldots, X_n$, n elemű mintát, mely minta segítségével:

$$T(X_1, \ldots, X_n) : \mathbb{R}^n \to \Theta$$

statisztika alapján becslést készítünk $\theta \in \Theta$ paraméterre.

Becslésünk jóságát általánosságban a

$$d(T(X_1, \ldots, X_n); \theta),$$

megfelelő d metrikában mért eltéréssel mérhetjük.

Megjegyzés. Gyakori a $d(a, b) = (a - b)^2$ négyzetes eltérés használata, alkalmazása.

Legyen $E(T(X_1, \ldots, X_n)) = \theta^*$ és jelölje $u = \theta^* - \theta$ a statisztikai eljárásunk torzításának mértékét.

2.1. Definíció. Amennyiben $u = 0$, úgy a $T(X_1, \ldots, X_n)$ becslést torzítatlan becslésnek szokás hívni.

Megjegyzés. Általában nem ad félreértésre okot, de érdemes megjegyezni, hogy θ^* elméleti paraméter (pl. elméleti átlag, elméleti szórás, elméleti ferdeség, elméleti csúcsosság).

A $T(X_1, \ldots, X_n)$ statisztika konkrét értékére a tapasztalati paraméter (tapasztalati átlag, tapasztalati szórás stb.) elnevezéssel szokás élni.

Azaz, a véletlen változó eloszlásának elméleti jellemzőjét szeretnénk a tapasztalati, mintából számított paraméterek segítségével megbecsülni.

Legyen $\delta(T(X_1, \ldots, X_n))$ a $T(X_1, \ldots, X_n)$ becslés valamely szóródási mutatója.

Többnyire a szórást[1] választjuk szóródási mutatónak, de érdemes azt is figyelembe venni, hogy a δ szóródási mutatót a d metrikával összhangba hozzuk, illetve akár vizsgálat tárgya is lehet a metrika és a szóródási mutató egymáshoz való viszonya. Például ha $d(a, b) = |a - b|$ választással élünk, akkor δ-ra az átlagos abszolút eltérés bizonyos szempontból jobb (indokoltabb) választásnak látszik az átlagos négyzetes eltérés (szórás) helyett.

A standard hiba így például az alábbi

$$H(X_1, \ldots, X_n) = u + \delta(X_1, \ldots, X_n)$$

összegként definiálható. Ez felfogható úgy is, hogy az eljárás hibája nem más, mint a becslés torzításának és – pusztán mert véletlen jelenségeket vizsgálunk – az eredendő eltéréseknek az együttese.

2.2. Definíció. Amennyiben a becslés torzítatlan (tehát $u = 0$), úgy ha teljesül, hogy

$$\lim_{n\to\infty} H(X_1, \ldots, X_n) = 0,$$

a becslést konzisztens becslésnek nevezzük. Tehát a konzisztens becslés egy olyan torzítatlan becslés, melynek standard hibája a mintaelemszám növelésével tetszőlegesen csökkenthető.

Megjegyzés. A két metrika, d és δ szerepe igen eltérő. Vegyük azt a példát, hogy attól függetlenül hogy mit is szeretnénk becsülni, mi mindenképpen egy konstans értéket mondunk: legyen ez 42. Így a $\delta = 0$ esettel állunk szemben – azaz, hacsak nem 42 a valódi paraméter, amit becsülni szeretnénk, úgy az eljárásunk „véletlen vizsgálatából fakadó" hibáját kiiktattuk, csak a torzítás marad.

[1] A szórás a variancia négyzetgyöke, azaz az átlagtól való átlagos négyzetes eltérés négyzetgyöke. Azonban ennek viselkedése és így megbízhatósága erősen függ a vizsgált változónk eloszlásától, ahogy ezt Lee és munkatársai dolgozatukban [28] megállapítják – erről a későbbiekben, részminták szórásának tesztelésekor részletesebben szót ejtünk.

Így a statisztikánk jóságát mérő d metrikában a véletlen szerepét kiiktattuk – de az eljárásunk valódi paramétertől való eltérését ettől még mérni fogjuk.

Amennyiben egyes paraméterekre több becslési eljárás is létezik (és általában létezik), akkor a lehetséges becslések közül az alábbi módon szokás választani:

2.3. Definíció. Két becslés közül azt nevezzük hatékonyabbnak, melynek kisebb a hibája adott mintanagyság mellett.

A fenti definíciók értelmében egy adott paraméterre az elérhető leghatékonyabb becslést érdemes választanunk (amennyiben az létezik).

Léteznek más megközelítések is egy-egy becslés elkészítésének vizsgálatakor. Világos, hogy az eddigiekben azt tekintettük alapnak, hogy a becslésünkből számított tapasztalati paraméter és elméleti paraméter várhatóan milyen távol lesznek egymástól.

Becslést alkothatunk úgy is, ha mintában rejlő információnk vizsgálatából indulunk ki:

2.4. Definíció. Legyen az X $(X_1, \ldots, X_n)$ független azonos eloszlású minta az X háttérváltozó eloszlásából, amely tehát a θ paramétertől függ, $\theta \in \Theta$. Feltesszük azt is, hogy $\dim(\theta) = 1$, és hogy Θ konvex. Ekkor a minta úgynevezett Fisher-féle információja:

$$I_n(\theta) = E\left[\left(\frac{\partial}{\partial\theta} l_\theta(x)\right)^2\right] > 0,$$

ahol az $l_\theta(x)$ az úgynevezett loglikelihood függvény, azaz a tapasztalati sűrűségfüggvény[2] logaritmusa.

Ez vezet az úgynevezett maximum-likelihood becslésekhez, amikor is lényegében arról van szó, hogy a minta alapján leginkább valószínű θ paramétert (eloszlást) választjuk a Θ paramétertérből.

Megjegyzés. Megjegyezzük, hogy másfajta becslési eljárásokat találhatunk, ha a

$$H(X_1, \ldots, X_n) = u + \delta(X_1, \ldots, X_n)$$

hibából elindulva úgy gondolkodunk, hogy az eltéréseket – annak mértékétől függően – más és más módokon büntetjük. Ezt a veszteséget nevezhetjük akár rizikónak is (bizonyos határig nem érdekel minket az eltérés vagy a torzítás, míg egy

[2]A tapasztalati sűrűségfüggvény lényegében egy oszlopdiagramként fogható fel (vagy annak simításaként). Technikailag úgy kell elképzelni, hogy a valószínűségi változó értékkészletét ekvidisztáns módon felosztjuk (a változót diszkretizáljuk) – majd az adott intervallumok relatív gyakoriságait ábrázoljuk. Az értékkészletet felosztó intervallumok számára általában $\sqrt{n}$ értéket választanak, ha $n < 100$, míg $1 + \log_2(n)$ értéket, amennyiben $n \geq 100$.

adott határt átlépve az eltérésekért például exponenciális módon fizetnünk kell). Ilyenkor értelemszerűen azt a $\theta \in \Theta$ paramétert fogjuk választani, ahol a veszteségünk (vagy rizikónk) minimális.

A becsléseinket sokszor az alábbi megközelítésben érdemes tárgyalni: tegyük fel, hogy most rendelkezünk két, $T_1(X_1, \ldots, X_n)$ és $T_2(X_1, \ldots, X_n)$ statisztikával (becsléssel) a $\theta^* \in \Theta$ paraméterre.

2.5. Definíció. Ekkor a $(T_1(\underline{X}), T_2(\underline{X}))$ intervallum legalább $1-\varepsilon$ szintű konfidenciaintervallum a θ^* paraméterre, ha

$$P(T_1(\underline{X}) < \theta^* < T_2(\underline{X})) \geq 1 - \varepsilon,$$

ahol $\varepsilon > 0$. A $1-\varepsilon$ az úgynevezett konfidenciaszint.

Megjegyezzük, hogy általánosítható bármely $f(\theta^*)$ függvényére a paraméternek e fenti felírása, ilyen esetben a

$$P(T_1(\underline{X}) < f(\theta^*) < T_2(\underline{X})) \geq 1 - \varepsilon$$

egyenlőtlenségnek kell fennállnia.

2.2. Hipotézisvizsgálat

Az előző fejezetben megalkottuk a konfidenciaintervallumokat, melyek azzal a tulajdonsággal bírtak, hogy vagy a $\theta^* \in \Theta$ paramétert, vagy annak valamely függvényét tartalmazták adott valószínűséggel. Ilyenkor azonban döntéseket is tudunk hozni – mely döntések átvezetnek minket a hipotézisvizsgálatok területére.

A hipotézisvizsgálatok során – igazodva most a becsléselméletben alkalmazott jelöléseinkhez – az alábbi módon járunk el általában: legyen $H_0 : \theta \in \Theta_0$ és $H_1 : \theta \in \Theta_1$, ahol $\Theta_0 \bigcap \Theta_1 = \emptyset$ és $\Theta_0 \bigcup \Theta_1 = \Theta$.

Fontos feltétele a hipotézisvizsgálatoknak, hogy a $T(\underline{X})$ statisztikánk eloszlását H_0 esetén ismernünk kell. A döntéshozatal felfogható olymódon, hogy e H_0 feltételezés mellett megalkotunk egy – a korábbi fejezetben már ismertetett, ε szintű konfidenciaintervallumot.

Ez az intervallum az alábbi módon interpretálható: amennyiben H_0 feltételezés igaz, úgy bármely, adott eloszlásból származó minta esetén számított $T(\underline{X})$ statisztika értékének legalább $1-\varepsilon$ valószínűséggel az adott intervallumba kell esnie.

A konfidenciaintervallumot elfogadási tartománynak nevezzük, míg annak komplementer halmazát kritikus tartománynak. Amennyiben a mintánkból számított $T(\underline{X})$ az elfogadási tartományba esik, úgy a H_0 nullhipotézis mellett döntünk és azt mondhatjuk, hogy a minta nem mond ellent e feltételezésnek (adott ε szinten). Míg ha a $T(\underline{X})$ a kritikus tartományból vesz fel értéket, úgy a H_1 ellenhipotézist választjuk és azt mondhatjuk, hogy az adott minta alapján a H_0 nullhipotézis teljesülése valószínűtlen (adott ε szint mellett), így e hipotézist elvetjük.

Világos, hogy ilyen esetekben két hibát[3] követhetünk el: ha a nullhipotézist nem tartjuk meg, pedig igaz, akkor az úgynevezett elsőfajú hibát követjük el – ennek valószínűsége legfeljebb ε, így elmondható, hogy a konfidenciaszint segítségével az elsőfajú hiba becsülhető. Szokás mind ε-t, mind $1-\varepsilon$ mennyiséget szignifikanciának, vagy szignifikanciaszintnek nevezni – általában nem okoz félreértést egyik vagy másik használata. Jelölésben hagyományosan α használatos a szignifikanciaszintre (nem a becsléselméletben használt ε).

A másikfajta hibát akkor követjük el, ha a nullhipotézist elfogadjuk, holott az nem teljesül. Ezt a hibát másodfajú hibának nevezzük és a statisztikai eljárás erejével becsülhető. E hiba mértékét β-val szokás jelölni, és a próba erejét β vagy $1-\beta$ jelöli (és a szignifikanciához hasonlóan itt sem szokott félreértést eredményezni egyik vagy másik mennyiség használata).

Megjegyzés. Fontos kiemelnünk: míg az elsőfajú hiba felülről becsülhető a szignifikanciaszinttel, addig a másodfajú hibát nem tudjuk becsülni. A hipotézisvizsgálati eljárások (próbák) ereje így általában csak adott helyzetben, tapasztalati úton az adott problémára vonatkoztatva kimérhető mennyiségek.[4]

2.3. Egy szimulációs módszer

A szimulációs technikák általában nem találhatók meg bármely bevezető statisztikai könyvben, azonban széles körben használtak, így a statisztikai bevezető fejezetben ezeket az eljárásokat is ismertetjük vázlatosan. A bevezetőben – miután a továbbiakban is csak erre koncentrálunk – az úgynevezett bootstrap eljárást ismertetjük, mely részletesen megtalálható például Efron e témában klasszikusnak számító cikkében [15].

A becsléselméletben már definiált hibát explicit formában a legritkább esetben lehet megadni, így például Monte-Carlo-módszer segítségével, szimulációval becsülhetjük.

– *(1.)* $\widehat{\theta}(x_1,\ldots,x_n)$ a statisztika értéke. (Egy adott $X_1=x_1,\ldots,X_n=x_n$ realizáció mellett.)

  Ekkor $\sigma(\theta)=\sqrt{Var_\theta(X_1,\ldots,X_n)}$ a statisztika valódi hibája.

  Ezt többnyire lehetetlen zárt formában felírni.

– *(2.)* Miután F eloszlást nem ismerjük, ezért $\widehat{F}$-pal, a tapasztali eloszlásfügg-

[3] Gondoljunk a farkast kiáltó pásztorfiú esetére. A *farkaskiáltás* tekinthető az úgynevezett elsőfajú hibának: nincsen gond a vizsgált rendszerben, mégis hibáról, problémáról teszünk jelentést. A másodfajú hiba ennek ellentéte, nevezhetjük *struccpolitikának* – a mesében a harmadik farkaskiáltás után a falusiak viselkedése: gond van a rendszerben, és mégsem veszünk róla tudomást.

[4] Az elsőfajú hiba mindig azt jelenti, hogy az adott, fix eloszlás mellett sikerült egy valószínűtlen mintát vennünk, melyből elutasítottuk a nullhipotézisben feltett eloszlásunkat. Azonban a másodfajú hiba azt jelenti, hogy a nullhipotézis nem az, aminek gondoljuk – viszont ez számtalan módon bekövetkezhet, ezért nem tudjuk egzakt módon megmondani e hiba valószínűségét, csak például szimulációkat készíteni az adott, konkrét minta ismeretében. Úgy is fogalmazhatunk, hogy a döntéshozatalunkhoz minden esetben az adott szignifikancia-szinten döntő, legerősebb próbára van szükségünk.

vénnyel becsüljük. Ekkor $\widehat{\sigma}_B = \sigma\left(\widehat{F}\right)$[5] becsüli $\sigma(F)$-et.

Itt csak approximációról van szó, hiszen ezt sem tudjuk zárt alakban felírni.

Megjegyzés. A tapasztalati eloszlásfüggvény nem más, mint hogy a lehetséges realizációkból megmondjuk, hogy a véletlen változónak adott értékei milyen valószínűséggel vétetnek fel (folytonos változó esetén adott értéknél nem nagyobb értékeket, vagy milyen valószínűséggel vesz fel a véletlen változó). E technikával tehát egy lépcsős függvényt nyerünk, mely a mintaelemek értékei esetén $\frac{1}{n}$ függvényértéket emelkedik.

A bootstrap eljárás ezek után egy független, azonos eloszlású, egyszerű, visszatevéses mintavételezés a tapasztalati eloszlásfüggvény alapján.

Ez tehát nem más, mint egy $U(X_1, \ldots, X_n)$, $X_1, \ldots, X_n$ pontokra koncentrált diszkrét egyenletes eloszlás szerint vett újabb és újabb véletlen mintavételezés.

Így tehát egy approximációs eljárást kell végrehajtanunk, mely a következő lépésekből áll.

- **(i)** $\widehat{F}$ meghatározása.
- **(ii)** $\widehat{F}$-ból független mintavétel segítségével $X_1^*, \ldots, X_k^*$ úgynevezett bootstrap minta létrehozása. Itt be kell tartanunk, hogy $\forall i: \; P(X_i^* = x_j) = \frac{1}{n}$. (Minden mintaelem ugyanolyan valószínűséggel veheti fel a realizációban szereplő különböző értékeket). *Azaz: a mintából független módon választunk, visszatevéses mintavételezéssel k darabot.*
- **(iii)** $\widehat{\theta}^* = \theta(X_1^*, \ldots, X_k^*)$ bootstrap másolatból származó statisztika kiszámítása.
- **(iv)** az **(ii)** és **(iii)** lépések B számú ismétlése. Így előállítunk egy $\widehat{\theta}_1^*, \ldots, \widehat{\theta}_B^*$ független boostrap másolatból származó statisztika-becslés mintát.
- **(v)** $\widehat{\sigma}_B$ approximáció kiszámítása az alábbi formula segítségével:

$$\widehat{\sigma}_B = \sqrt{\sum_{b=1}^{B} \frac{\left(\widehat{\theta}_b^* - \widehat{\theta}_\bullet^*\right)^2}{B-1}},$$

ahol

$$\widehat{\theta}_\bullet^* = \frac{1}{B} \sum_{b=1}^{B} \left(\widehat{\theta}_b^*\right).$$

[5] $\widehat{\sigma}_B$ az approximációs eljárás utolsó lépésében formalizálásra kerül, mely tehát nem más, mint a tapasztalati eloszlás szórása.

Megjegyzés. Ekkor, ha $B \to \infty$, úgy a $\widehat{\sigma}_B$ közelíti $\sigma(F)$-et. B optimális megválasztásáról nincsenek különösebb viták: általában elegendő 100 és 500 közötti bootstrap minta kiszámítása.

Más filozófia alapján folytatható addig az **(ii)**-**(v)** lépések egymásutánja, ameddig a lépésenként kiszámított és korrigált $\widehat{\theta}_\bullet$ valamilyen, előre meghatározott, minőséget előíró korlátnál kevesebbet változik 1 lépés alatt.

Fontos megjegyezni, hogy az approximáció konvergenciájához elégséges feltétel a véges szórásnégyzet (közös eloszlást feltételezzünk fel itt is az X_i változókra nézve), mely – mint azt már láttuk, a centrális határeloszlás tétel teljesülése miatt szükséges.

A bootstrap algoritmus egyik előnye az, hogy a tapasztalati eloszlásfüggvényből táplálkozva lehetőséget biztosít számunkra, hogy pl. a tapasztalati kvantilisek becslésével tapasztalati konfidenciaintervallumokat is meghatározzunk.

3. Érzékenységvizsgálatok

A bevezető fejezetek után rátérhetünk az érzékenységvizsgálatok kérdésére. Már a két bevezető fejezetből is érzékelhető, hogy a statisztikában igen fontos – de gyakran nem elég hangsúlyos – terület az adatok érzékenységének vizsgálata.

Más megközelítésben: a hagyományos eljárások sok helyen, sok formában elérhetők, megtalálhatók – ezek alkalmazása azonban feltételekhez kötött. Annak ismerete, vizsgálata, hogy e feltételek sérülése esetén mi történik a vizsgálatunk kimeneti adataival, nem teljesen kidolgozott. Értjük ezalatt azt, hogy bár a módszerek gondosan felsorolják az alkalmazhatóság feltételeit, nem szólnak arról, hogy mit kellene tenni, ha egyes feltételek sérülnek. A könnyen elérhető programcsomagok nem feltétlenül tartalmazzák a feltételek vizsgálatait, ennek következtében az alternatív eljárások végképp nem kerülnek bemutatásra.

3.1. Becslések érzékenységvizsgálata

Becsléseink elkészítésekor három olyan pont is megemlíthető, mely garantáltan befolyásolja a becslésünk minőségét, jóságát.

1. A d metrika: különböző metrikákban a becslésünk jóságát más és más eltérések fogják befolyásolni – így azt is megállapíthatjuk, hogy attól függően, hogy mely eltérésekre vagyunk érzékenyebbek, esetleg eltérő becsléseket kell majd alkalmaznunk.

2. A n mintanagyság: általánosan megfogalmazható az az elvárás, hogy egy véletlen jelenséget vizsgálva a mintanagyság növelésével egyre jobb becsléseket nyerjünk – de legalábbis ne romoljon a becslésünk minősége.

3. X véletlen változó eloszlása: e harmadik tulajdonság nem biztos, hogy elsőre szembetűnő, de viszonylag könnyen elfogadható, ha arra gondolunk, hogy egy olyan véletlen változó, mely pl. sűrűbben vesz fel extrém nagy, vagy éppen extrém kicsi értékeket, ugyanazon T statisztikára nézve merőben más viselkedést tud mutatni, mint pl. egy dichotóm véletlen változó.

Ezek után felmerül a kérdés: a becsléselméletben, egyes becslések alkalmazása során e három kritérium közül melyekre rendelkezik a statisztika érzékenységvizsgálatra vonatkozó válaszokkal, illetve mely területekre kell még esetleg válaszokat keresni?

E kérdéskört első megközelítésben az úgynevezett standard hibák meghatározása jelenti. A standard hibát általában négyzetes módon határozzák meg – mi ennél általánosabban, a becslési eljárás hibájáról fogunk szólni.

3.1.1. A metrikák

Jól felfogott érdekünkben használunk többesszámot e részfejezet címében: nem mindegy ugyanis, hogy a becslési eljárás véletlentől való függését mérő δ-t szeretnénk vizsgálni – vagy pedig a valódi paraméter és a becsült paraméter d-vel jelölt várható eltérését.

Megjegyzés. Általánosságban az úgynevezett standard hibát szokás a becslések esetén meghatározni, mely az elméleti és a tapasztalati paraméter eltéréséből származtatott átlagos eltérés.

A soron következő példákhoz tartozó vizsgálatokat megtalálhatjuk például Jones és Gill 1998-as cikkében [24].

Megjegyzés. Többször fogunk élni az alábbi jelöléssel: $f\,(\alpha; df)$. Ez azt jelenti, hogy az adott f típusú eloszlás, df szabadsági fokhoz tartozó, α szignifikanciaszintjének úgynevezett kvantilise.

Például $1,89 = t(0,05; 7)$ azt jelenti, hogy a 7 szabadsági fokhoz, $\alpha = 0,05$ szignifikancia-szinthez tartozó kvantilise[6] az úgynevezett t-eloszlásnak (vagy Student-féle t-eloszlásnak).

3.1. Példa. Az első négy tapasztalati momentum konfidenciaintervallumát az alábbi módokon határozhatjuk meg:

– Átlag:

$$\overline{X} \pm t_{(\frac{\alpha}{2}, n-1)} \frac{s^2}{\sqrt{n}},$$

azaz az átlag esetén kis mintáknál (például $n \leq 100$) a megfelelő szabadságfokú és megbízhatósági szintet használó t-eloszlás kvantilisével dolgozunk,

[6]Ez a kvantilis az eloszlásnak az a pontja, melyre igaz, hogy a 7 szabadságfokú t-eloszlásból származó véletlen változó 1,89-nél kisebb értéket 95, tehát ennél nagyobb értéket 5%-os valószínűséggel vesz fel.

nagy mintáknál a standard normális eloszlás is használható a t-eloszlás helyett.

Megjegyzés. Megfigyelhető, hogy az átlag becslése így konzisztens: a standard hibája a minta végtelenbe tartása mellett 0-hoz konvergál – amennyiben véges a szórása a vizsgált véletlen változónknak.

– Szórás:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\left(\frac{\alpha}{2},n-1\right)}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{\left(1-\frac{\alpha}{2},n-1\right)}}}.$$

– Ferdeség:

$$g_1 = \frac{\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^3}{n}}{\left(\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n}\right)^{\frac{3}{2}}}, \qquad G_1 = \frac{\sqrt{n(n-1)}}{n-2}g_1,$$

$$SES = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}.$$

Innen a ferdeség konfidenciaintervalluma:

$$G_1 \pm z_{\left(\frac{\alpha}{2}\right)} SES,$$

ahol $z_{\left(\frac{\alpha}{2}\right)}$ nem más, mint a standard normális eloszlás eloszlásfüggvénye inverzének értéke az $\frac{\alpha}{2}$ helyen. Ez utóbbi az alábbi módon is írható:

$$\frac{G_1}{SES} \sim Z,$$

azaz $\frac{G_1}{SES}$ eloszlása standard normális[7].

– Csúcsosság:

$$a_4 = \frac{\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^4}{n}}{\left(\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n}\right)^2}, \qquad g_2 = a_4 - 3,$$

$$G_2 = \frac{n-1}{(n-2)(n-3)}\left((n+1)g_2 + 6\right).$$

[7] A standard normális eloszlást szokás Z-vel jelölni, az eloszlásban való viselkedést pedig $\sim$ segítségével.

A csúcsosság standard hibája:

$$SEK = 2SES\sqrt{\frac{n^2-1}{(n-3)(n+5)}}.$$

Így a csúcsosság konfidenciaintervalluma meghatározható, hiszen

$$\frac{G_2}{SEK} \sim Z.$$

E fenti konfidenciaintervallumok meghatározásakor felmerülhet a kérdés, hogy az átlagra vonatkozó konfidenciaintervallum leggyakoribb alkalmazása, nevezetesen az egymintás t-próba miként viselkedik abban az esetben, ha a normalitás feltételét nem tudjuk garantálni.

Megjegyzés. Egy fontos megjegyzést kell itt tennünk. Majd a későbbiekben még látni fogjuk, hogy a normalitás esetén nem feltétlenül az a legnagyobb problémánk, hogy az átlagot miként tesztelhetjük, hanem már azon is el kell gondolkodnunk, hogy az átlagot teszteljük-e egyáltalán?

Gondoljunk itt arra, hogy az átlagnak van egy olyan, szükségszerű háttérjelentése, melyet az elméleti paraméter okán hordoz: nevezetesen a várható érték miatt az átlag interpretációjához hozzá tartozik, hogy „ezt az értéket várjuk". Azonban ha például társasjátékot játszunk egy hatoldalú dobókockával, akkor egészen biztosan lehetünk abban, hogy − bár a várható értéke a dobásainknak $3,5$ − a játékot játszók közül senki sem várja, hogy $3,5$-et dobjon. Azt azonban mindenki elfogadja, hogy a dobások fele $3,5$ alatt, míg másik fele $3,5$ felett lesz. Ez azonban a medián, tehát ilyen esetben indokoltabbnak látszik ezt tesztelni − még ha meg is egyezik az értéke szimmetrikus eloszlások esetén az átlaggal.

Ebben a témában számos publikáció látott napvilágot, a teljesség igénye nélkül: a közelmúltban jelent meg magyar nyelven Vargha összefoglaló cikke [45] a Statisztikai Szemlében, illetve idézhető két klasszikusnak számító, t-próba próbastatisztikáján módosítást javasló cikk: Johnson 1978-as cikke [23], illetve egy korábbi, 1949-es cikk Gayentől [17]. E két utóbbi cikkben az alábbi módosításokat javasolják a t-próba[8] próbastatisztikáján:

$$t_{JOHNSON} = t + G_1\sqrt{n}\left(\frac{1}{6n} + \frac{\left(\overline{X}-\mu_0\right)^2}{3s^2}\right),$$

[8]A t-próba (vagy student-próba) egy ismert, klasszikus statisztikai próba. Ennek során a vizsgált nullhipotézisünk: $H_0: E(X) = \mu_0$, próbafüggvénye $t = \frac{\overline{X}-\mu_0}{\frac{s}{\sqrt{n}}}$, ahol $\overline{X}$ és s megegyezik a korábbi jelölésekkel. A t próbastatisztika tehát a mintából számított próbastatisztika (így maga is véletlen), melynek eloszlása az úgynevezett t-eloszlás − amennyiben X véletlen változó eloszlása normális, illetve teljesül a nullhipotézis.

míg Gayen azt mondja, hogy a szokásos $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ helyett az alábbi függvényt[9] használjuk:

$$f(x) = \phi(x) - \frac{G_1}{3!}\phi^{(3)}(x) + \frac{G_2}{4!}\phi^{(4)} + \frac{G_1^2}{72}\phi^{(6)}(x),$$

ahol $\phi^{(r)}$ az r-edik deriváltat jelenti, míg G_1 és G_2 a fent már definiált tapasztalati ferdeség és csúcsosság.

Innen azt is láthatjuk, hogy Johnson módosítása a ferde eloszlások esetén nyújt segítséget számunkra, míg Gayen mind a ferdeséget, mind a csúcsosságot korrigálja módosításában.

E fenti paraméterek viselkedéséről és tulajdonságairól, illetve a standard hibák viselkedéséről széles körben lehet még további szakirodalmat találni, többek között:

- A különböző változók, véletlen jelenségek bizonyos paramétereinek (általában átlag) standard hibáinak összefoglaló táblázata több helyen is megtalálható, erre példa lehet [49]. E táblázatokból arra vonatkozóan kaphatunk információkat, hogy jól specifikált véletlen jelenségek esetén, azok elméleti paraméterét milyen pontossággal lehetett megbecsülni – adott mintanagyság mellett.

- Efron és Tibshirani Statistical Science folyóiratban megjelent cikkükben [15] empirikus és elméleti eredményeket foglalnak össze a bootstrap metódus kapcsán. Ezt az eljárást alkalmazhatjuk különböző paraméterekre vonatkozó standard hibák és konfidenciaintervallumok meghatározására, illetve vizsgálják e módszer általános statisztikai tulajdonságait is (például különböző becslési eljárásokban való viselkedését).

- Belia és munkatársai cikkükben [2] felhívják a figyelmet az általunk is feltett egyik kérdésre, illetve tapasztalatra. E témakörben ugyanis számos anomália van jelen: rosszul interpretált adatokkal és következtetésekkel találkozhatunk e szerzők szerint (tételesen megneveznek idézett cikkükben tanulmányokat), és az általuk idézett tanulmányokban a tanulmányt jegyzők konfidenciaintervallumok és/vagy standard hibák helytelen meghatározása, ábrázolása vagy értelmezése után vonnak le hibás vagy megkérdőjelezhető következtetéseket.

- Végül – egyáltalán nem utolsó sorban, átvezetendő a mintanagyság problémájához e kérdéskört – a Judkins által vizsgált, Fay-féle eljárásban [25] arról van szó, hogy becslésünk megbízhatósága drasztikus mértékben romlik, ha a mintavételezési eljárásunk során nem tudtuk a mintaelemeink függetlenségét

[9] A hagyományos t-próbába kisebb elemszámok esetén a t-eloszlást használjuk, míg nagyobb elemszám esetén (gyakorlatban például 150-nél nagyobb mintáknál) a standard normális eloszlást. Gayen azt javasolja, hogy a normalitás sérülése esetén e két, általánosan használt eloszlás helyett e módosítottat alkalmazzuk inkább. Hangsúlyozzuk, hogy Johnson és Gayen módosításait akkor használjuk, ha szakmailag még mindig indokolt az átlag bárminemű tesztelése a normalitás sérülése esetén. Ellenkező esetben – ahogy már említettük – más középértékek tesztelése indokolt.

garantálni (ez könnyedén előfordulhat többek között szociológiai vizsgálatoknál, hiszen például az egy munkahelyen dolgozók, vagy az egy iskolában tanulók semmiképpen sem tekinthetők függetlennek). Ennek hatásvizsgálatát egy korábbi cikkünkben [44] mutatjuk be esettanulmányként, ahol az OECD által szervezett oktatáspolitikai felmérés adatainak elemzésén a különböző módszerek hatásmechanizmusát elemezzük. A Fay-féle eljárás egy másik aspektusát – a már említett, Efronék [15] által is vizsgált szimulációs eljárással való kapcsolatát – taglalja Saavedra egy előadásában [40].

Megjegyzés. Ez utóbbi tanulmánnyal rá is világíthatunk e kérdéskör egy újabb problémájára: ha úgy találjuk, hogy valamely eljárás biztonságát szimulációs technikák segítségével szeretnénk vagy tudjuk vizsgálni, még akkor sem egyértelmű, hogy mely szimulációs eljárást válasszuk.

Felmerülhet e felsorolás után a kérdés: a hibás döntések e kérdéskör (az érzékenységvizsgálat) elhanyagoltsága, nem kellően fontosnak tartott mivolta miatt keletkeznek – vagy valójában az alkalmazott eljárásoknak kellene olyan biztonsági hálót tartalmazniuk, melyek a hibás döntéseket is kellően megszűrik?

Ez alatt érthetjük például azt, hogy az alkalmazók számára könnyen elérhető statisztikai programcsomagokban az eljárások nem feltétlenül tartalmazzák az adott eljárások feltételeinek teljes vizsgálatát – és ha bizonyosakat tartalmaznak, úgy nem feltétlenül azokat, melyek miatt a tapasztalatok szerint leginkább instabillá válhatnak az eljárások. Egészen pontosan: a programcsomagok általában képesek a feltételek ellenőrzésére – csak azok nem feltétlenül képezik egy-egy eljárás szerves részét. Ne felejtsük el megemlíteni, hogy akár így is előfordulhat a már idézett LeVay féle fiaskó [31].

3.1.2. A mintanagyság

Bizton állíthatjuk, hogy e kérdés szakirodalma és e kérdésben elvégzett vizsgálatok kellő támpontot tudnak nyújtani bárki számára azon kérdés eldöntésében, hogy egyes paraméterek vizsgálata során az adott paraméter és a kiválasztott minta esetszáma között milyen jellegű összefüggések adódnak.

Első feltételezésünk az lehet, hogy a populációnk, melyet vizsgálunk végtelen. (Egészen más a helyzet ugyanis, ha véges populációkkal dolgozunk, erről is lesz még szó.)

A végtelen populációk esetén az elmélet a konzisztens becslések biztonságára hívja fel a figyelmet, illetve azokat a becsléseket részesíthetjük előnyben, melyekről összefoglalóan azt mondhatjuk el: a mintaelemszám növelésével csökken a korábban már definiált hibájuk.

3.2. Példa. A mintanagyság döntően befolyásolja a becsléseink pontosságát és így a belőlük levonható következtetéseket is. Tegyük fel, hogy az általunk vizsgált populációban a két nem magasságát szeretnénk összehasonlítani. A férfiak (1) és nők (2) testmagasságának átlagára és korrigált tapasztalati szórására az alábbi eredményeket kapjuk:

$$\overline{X}_1 = 180,001\,cm \quad s_1 = 10\,cm,$$
$$\overline{X}_2 = 180\,cm \quad s_2 = 10\,cm.$$

A fenti adatok természetesen kitaláltak a probléma érzékeltetése érdekében.

Tegyük fel, hogy első esetben a két minta nagysága $n_1 = n_2 = 100$. Ebben az esetben a kétmintás t-próba próbastatisztikája:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}} \sqrt{\frac{n_1 n_2\,(n_1 + n_2 - 2)}{n_1 + n_2}} = 0,0007,$$

azaz nincsen szignifikáns különbség a két változó között, hiszen a szokásos 5%-os szignifikancia-szint melletti kritikus érték $1,96$ lenne.

Azonban ha a mintanagyságot drasztikusan megnöveljük, $n_1 = n_2 = 10^9$ értékekre, úgy $t = 2,236$ adódik, ami már szignifikáns eltérést jelez. Azaz: a mintanagyság növekedése – minden más paraméter fixen tartása mellett – automatikusan csökkenti az elsőfajú hiba valószínűségét, ennek következtében viszont anomáliák adódhatnak. Egy ilyen anomália a fenti: $0,001$ cm-es eltérés tehát szignifikáns különbségként jelentkezik e próbában – amit igen nehéz komoly eltérésként értelmezni.

Cohen azt javasolja [12] könyvében, hogy az ilyen helyzetekre alkalmazzuk kiegészítő mutatóként az átlagok standardizált különbségét, mely nem más, mint

$$\Delta_{Cohen} = \frac{\overline{X}_1 - \overline{X}_2}{s} = 0,0001,$$

ahol $s = 10$, a teljes minta korrigált tapasztalati szórása.

Amennyiben ez az érték 0,3 alatti, úgy azt mondhatjuk, hogy (bár lehet szignifikáns az eltérés), az szakmailag gyenge hatást mutat. Amennyiben 0,7 feletti Δ értéket tapasztalunk, úgy szakmailag jelentős eltérésre bukkantunk – a köztes értékek szakmailag közepes hatást jeleznek. Azaz: a becslésünk pontosságának javulása automatikusan eredményezi a stabilabb, pontosabb döntéshozatalt – ám ez nem feltétlenül jelent szakmailag is releváns eltéréseket.[10]

Megjegyzés. Fontos kiemelni: a testmagasságokat ilyen módon összehasonlító példánkban a statisztikai döntéshozatal addig terjed, hogy megállapítsuk a szignifikáns eltérések jelenlétét. A döntésünk szakmai utóélete már nem a statisztika, hanem az adott, statisztikát alkalmazó tudományterület feladata és felelőssége.

[10] A fenti példával élve: azért, mert van egy teljes földkerekséget felölelő becslésünk a férfiak és nők testmagasságáról, melyből azt tapasztaljuk, hogy a férfiak magassága szignifikánsan nagyobb $0,001$ cm-rel, nem fogjuk minden építészeti főiskolán és egyetemen azt tanítani, hogy az új tudományos eredményeinknek köszönhetően minden újonnan építendő sportlétesítmény férfiaknak szánt öltözőjébe tegyenek egy kicsivel keskenyebb linóleumot, hogy a magasságbéli különbségeket mostantól korrigáljuk.

Lehmann egyik, becsléselmélettel foglalkozó könyvében [29] számos tételt találhatunk arra vonatkozóan, hogy a véletlen változó bizonyos tulajdonságai mellett milyen hibahatárok érhetők el[11]. E könyv második fejezetében egzisztencia állításokat találhatunk, továbbá olyan feladatokat, problémákat tárgyal, melyben konkrét becslésekre (pl. átlag, szórás, kovariancia) hol az úgynevezett rizikó, hol pedig a Fisher-információ segítségével vizsgálja a becslések jóságát, illetve elemzi a kívánt mintanagyságot.

Hasonlóan ide köthetők elméleti megközelítések alapján a különböző „nagy számok törvényei", illetve a különböző, becslésekre vonatkozó egyenlőtlenségek (Markov, Csebisev).

Annak megválaszolására, hogy adott bizonytalanság eléréséhez milyen mintanagyságra van szükségünk többféle módon is választ kaphatunk, többek között:

– Amennyiben ismerjük a becslésünk eloszlását, úgy meghatározható segítségével a becslésünk úgynevezett konfidenciaintervalluma. Erre közismert példa a mintaátlag és annak standard hibája [29], de ismert a szórás (mely Cochran tétele értelmében az átlagtól független módon becsülhető) konfidenciaintervalluma is (pl. Cochran cikkében [11] megtalálható). Ezeket a formulákat már korábban bemutattuk.

  Fletcher és Webster cikkükben [16] a ferdeség hatását vizsgálták különböző becslésekben, míg szintén a ferdeséggel, illetve az eloszlás csúcsosságával összefüggésben, ezen két paraméter becslésének jóságát vizsgálták Wright és Herrington [47] tanulmányukban, akik azt tapasztalták, hogy már kisebb minták esetén is stabilabb becslés mondható e két paraméterre szimulációs eljárásokkal (ők a bootstrap eljárást használták), mint a paraméterek ismert standard hibájának felhasználásával.

  Mameli és munkatársai tovább is mennek alkalmazásaikban ennél: 2012-ben írt cikkükben nagy mintás[12] elemzéseken, orvosi alkalmazásokkal is kiegészítve (illetve valós adatokon tesztelve), összehasonlítják módszerüket a hagyományos, illetve egy paraméteres bootstrap eljárás eredményeivel.

– Kis minták esetén felmerülő anomáliák feloldására adnak támpontot az úgynevezett „breakdown point" elemzések (lásd alább). E témakör kutatásai arról adnak számot, hogy egyes becslések, illetve belőlük származtatott hipotézisvizsgálati eljárások miként viselkednek a minta egyes elemeinek torzulásakor.

[11] Gondoljunk itt arra az egyszerű feladatra, hogy például az átlag standard hibája a megismert $\frac{s}{\sqrt{n}}$ formulával határozható meg. Ha előírjuk a hibahatárt és ismert a szórás, akkor meg tudjuk mondani, hogy adott szórás mellett mekkora mintára van szükségünk annak érdekében, hogy várhatóan az előre megadott hibahatáron belül tudjuk tartani a becslésünket.

[12] A kis és nagy minták általában nem egzakt megfogalmazások. Egy 20-30 elemszámú mintát még kis mintának szokás nevezni, míg egy 80-100 esetet vizsgáló realizáció már tekinthető nagy mintának. A mintánk elemszáma, annak „nagysága" általában attól függ, hogy mit is vizsgálunk, vizsgálatunkban használt próbastatisztika mennyire érzékeny. Lásd például e fejezet következő pontjában található „breakdown point" analízist.

Megjegyzés. Gondoljunk itt arra, hogy például az átlag számítását egyetlen mintaelem megváltoztatása is tetszőlegesen módosíthatja – más megközelítésben a mintaátlag instabil paraméternek tekinthető e fent nevezett elmélet értelmében.

Ezzel szemben például a medián lényegesen nagyobb tűréshatárral bír akár még egészen kis minták esetén is (például egyetlen mintaelem akár végtelenbe tartása esetén sem fog nagyfokú ingadozást mutatni).

A „breakdown point" elemzés tehát az adott paraméterekre vonatkozóan a becslés egy olyan értéket adja meg, hogy az adott mintanagyságok mellett a minta mekkora hányada módosítható úgy, hogy a minta egésze a becslésre vonatkozóan ne váljon használhatatlanná.

– Átlag: miután az átlagot $\overline{X} = \frac{X_1+\cdots+X_n}{n}$ formulával határozhatjuk meg, világos, hogy ha az első $n-1$ értéket fixnek tekintjük és $X_n \to \infty$ feltételt nézzük, úgy az egész átlagra is teljesül, hogy $\overline{X} \to \infty$.

  Így a véges „breakdown point" $\frac{1}{n}$, míg asszimptotikusan 0.

– Medián: az átlaggal szemben ha elképzeljük, hogy az sorba rendezett minta $\left\lfloor \frac{n-1}{2} \right\rfloor$ legkisebb elemet fixáljuk, úgy látható, hogy a felső, ugyanennyi elem (mediánnál nagyobbak) szabadon növelhetőek, a medián értékét nem módosítják.

  Így a véges „breakdown point" $\left\lfloor \frac{n-1}{2n} \right\rfloor$, míg aszimptotikusan $\frac{1}{2}$.

Azaz érzékenység szempontjából a medián lényegesen jobban viselkedik, mint az átlag – hiszen az adataink közel felét megváltoztatva is stabilitást mutat ez az paramétere az eloszlásnak. Erre vonatkozóan a következőkben egy példával is érzékeltetni fogjuk a két középérték közötti különbséget egy versenyhelyzet értékelése kapcsán.

A „breakdown point" elemzésekről egy speciális esetben értekezik Camponovo és Otsu 2012-ben megjelent cikkükben [10], ahol a szerzők a későbbiekben még szintén tárgyalt bootstrap eljárás viselkedését figyelték az extrém értékek megjelenésének fényében.

Az ezen téma iránt érdeklődő Olvasó számára egy összefoglaló, a fenti példát is tartalmazó, az egymintás t-próba esetét taglaló jegyzetet ajánlhatunk kiindulópontnak, melyet 2006-ban publikált Geyer [18] – és mely jegyzetben e téma néhány alap eredményét foglalja össze, illetve ad támpontot további kutatásokhoz, számításokhoz.

Elmondható tehát, hogy bizonyos paraméterek esetén ismerjük azok becslésének eloszlását – így tudjuk, hogy várhatóan milyen hibát vétünk a becslési eljárás alkalmazásával. Azonban kis minták esetén, vagy olyan paraméterekre, melyek eloszlása nem ismert, ilyen információval nem rendelkezünk.

E helyzetek feloldására tűnik elfogadható empirikus megoldásnak a korábbiakban már említetteken túl a különböző szimulációs technikák alkalmazása.

3.1.3. Véges sokaságok esete

A véges sokaságokról több helyen is szerezhetünk információkat, pl. Lehmann becsléselméleti, továbbiakban is még idézett könyvében részint a mintavételezési problémákról (3. fejezet 6. alfejezet), részint például M-becslésekről (5. fejezet, 6. alfejezet), melyekre vonatkozóan tapasztalati eredményeket is találhatunk az idézett műben.

E fejezetben külön találhatunk számos információt a Huber-féle robusztus becslési eljárásról (Huber-féle simított becslésnek is nevezik). A Huber-féle eljárás során lényegében kombináljuk a medián és az átlag információit, ennek segítségével alkothatunk robusztus becslést az átlagra – azonban feltétele az eljárásnak az eloszlás szimmetriája.

3.3. Példa. Legyen $X_1, \ldots, X_n$ független, azonosan $P_{\mu,\sigma}$ eloszlásból származó minta, ahol

$$f_{\mu,\sigma} := \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

alakú[13]. Ekkor az M-becslés a μ eltolás paraméterre azon t érték, melyre:

$$\sum_{i=1}^{n} \phi\left(\frac{X_i - t}{\sigma}\right) \to \min_t,$$

vagy más megközelítésben:

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - t}{\sigma}\right) = 0,$$

ahol $\psi = \phi'$. Megjegyezzük, hogy a maximum-likelihood becslés esetén $\phi_f = -log(f)$, míg $\psi_f = -\frac{f'}{f}$.

Speciális esetben a fenti simítási eljárás a következőképpen módosítható, alkalmazható. Adott k konstans mellett az úgynevezett Huber-féle becslés vagy transzformáció az alábbi:

$$\psi_k(x) = \begin{cases} k & x > k, \\ x & -k \le x \le k, \\ -k & x < k. \end{cases}$$

Megjegyzés. A fenti függvény egyfajta trimmelésként[14] is felfogható. Azonban míg a trimmelés esetén a kiugró értékektől megszabadulunk – ezzel a minta

[13] Feltesszük, hogy az F eloszlás szimmetrikus, továbbá feltehető, hogy $\sigma = 1$. Az eljárásban tehát az eloszlásfüggvényt ismertnek tekintjük, a két fenti paramétert pontbecslés segítségével becsüljük.

[14] A trimmelés azon statisztikai eljárás, melyben a kiugró vagy extrém értékeket levágjuk, kihagyjuk a mintából – centralizálva így a mintánkat, illetve csökkentve annak szabadásfokát.

szabadságfokát, esetszámát is csökkentve – addig itt az esetszám megmarad, csak egy adott értéken túl a számunkra megválasztott szint (k) kerül az adott szintnél nagyobb és kisebb esetek helyére. Más megközelítésben a kiugró értékeket egy számunkra beállított toleranciaszintre kényszerítjük vissza, ha úgy tetszik centrálunk.

Ezt az eljárást vizsgálta, illetve módosította Hampel munkatársaival, melyet 2011-ben publikáltak. E simítás azért is lehet fontos számunkra, mert a simítás a mintanagyság figyelembe vételével történik. Tanulmányukban kitérnek arra is, hogy az eljárást mind a Huber-féle transzformációra, mind a maximum-likelihood becslésre, mind pedig egyéb M-becslésekre alkalmazzák – ráadásul a simítási eljárásukat minden esetben össze is hasonlítják az eredeti eljárásokkal. Tapasztalataik szerint a simított eljárás minden esetben jobb (vagy legalábbis nem rosszabb) eredményeket hozott, mint nem simított változatuk.

3.1.4. A véletlen változó eloszlása

A korábban, a bootstrap szimuláció kapcsán már említettük, hogy a becslés eloszlásának ismerete segítségével a bizonytalanság, az eljárásunk érzékenysége vizsgálható. Azonban azt is tudnunk kell, hogy a véletlen jelenség eloszlása nagyban befolyásolja a becslési eljárásunkat (egyáltalán, már azt is befolyásolja, hogy mely paraméterekre szeretnénk becslést mondani és mely paraméterek nem érdekesek számunkra).

Lehmann [29] több eloszlás esetén is tárgyalja különböző paraméterek becslési tulajdonságait, azok viselkedését konzisztencia, torzítatlanság szempontjából, illetve hatékonyságukat is vizsgálja. Sak és munkatársai ennél tovább is mennek egészen friss kutatási riportjuk [41] tanúsága szerint, melyben azt vizsgálják, hogy különböző eloszlások ferdeségi mutatója miként hat az átlag konfidenciaintervallumára, illetve ezt milyen empirikus módszerekkel lehet korrigálni. Azt tapasztalták, hogy Hall 1992-ben publikált transzformációja [20] hatékony eszköznek bizonyul annak érdekében, hogy az átlagra vonatkozó konfidenciaintervallumot továbbra is zárt formula segítségével, szimulációk nélkül határozhassuk meg.

3.4. Példa. Míg az eredeti t-próba próbastatisztikája

$$t = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}},$$

addig a Hall-féle transzformáció:

$$g_1(t) = t + \frac{1}{\sqrt{n}} G_1 \left(\frac{1}{3} t^2 + \frac{1}{6} \right) + \frac{1}{3n} \left(\frac{1}{3} G_1 \right)^2 t^3,$$

ahol G_1 a tapasztalati ferdeség, tehát ilyen szempontból Hall transzformációja a Johnson-féle, ferdeséget korrigáló eljárással rokon[15].

[15] A t-próbának, mint már említettük, feltétele, hogy a vizsgált változó normális eloszlású legyen. Ennek egyik lehetséges ellenőrzése is lehet, hogy a normális eloszlás – szimmetrikus

A t-próba korrekciójának akkor van csak ilyen jellegű jelentősége, ha a normalitás sérülése mellett továbbra is a várható értéket (átlagot) szeretnénk tesztelni. A felmerülő probléma érzékeltetésére képzeljük el a következő esetet.

3.5. Példa. Adott két középiskolai osztály, akik futóversenyt szeretnének rendezni. Az összehasonlítás alapja a két osztály átlagos futásteljesítménye lesz. Az egyik osztályban csupa élsportolót találunk: 27 atlétát és 3 szumóbirkózót. A másik osztályban sok átlagos diák mellett (29 fő) egyetlen nagyon túlsúlyos diák is tanul. Azonban e túlsúlyos diák – megismerve az ellenfél adottságait – cselt eszel ki. A futóversenyt a Margitszigeten rendezik meg, egyetlen kört kell futni. A diákok nekikezdenek – de túlsúlyos egyedünk csak sétál, mellette a 3 szumóbirkózóval. Az atléták természetesen gond nélkül gyorsabbak az átlagos középiskolás diákoknál – de a csel még nem teljesedett ki. Hősünk beszélgetést kezdeményez a birkózókkal és a beszélgetést a gasztronómia irányába tereli. Majd a sziget egy céltól és rajttól egyaránt távoli pontján lévő talponálló büféhez vezeti a gyanútlan birkózókat. Ott aztán pénzt nem kímélve etetni kezdi őket. A trükk ugyanis a következő: a 3 birkózó – még akár az utolsó pár méteren le is hajrázhatják majd a továbbra is sétáló, velük tartó egyetlen túlsúlyost – eredményei már úgyis olyannyira fogja az egész osztályuk átlagát rontani, hogy bármely, átlagot összehasonlító eljárásban toronymagas győztesként kerül majd ki a teljesen átlagos középiskolai osztályunk. Ez azonban nyilván amiatt alakulhat ki, hogy az eloszlásaink, melyek az osztályokat jellemzik ferdék (például túl sok jó/átlagos és aránylag kevés rossz futó van), továbbá az átlagot egyetlen extrém érték is bármilyen irányba el tudja mozgatni. Így az átlag helyett más mutatóval, eljárással kellene döntenünk a két osztály összehasonlításában (ahogy ezt a „breakdown point" elemzésben már megállapíthattuk). Ha azt a kísérletet végeznénk el, hogy páronként futtatjuk őket, mely párokat véletlenszerűen válogattuk ki egyik és másik osztályból, úgy érzékelhető, hogy a sporttagozatos osztály esetén csak minden 10. választás lesz olyan, ahol az átlagos középiskolából választott diáknak lenne valami esélye – feltéve, ha onnan nem a túlsúlyos egyedet választjuk. Ez utóbbi kísérletet sztochasztikus egyenlőség vizsgálatnak nevezzük és Wilcox már korábban is idézett könyve [46] tartalmaz ilyen – vagy hasonló – helyzetekre alkalmazható próbákat, eljárásokat.

3.1.5. Összefoglaló megállapítások a becslésekhez

Megállapíthatjuk tehát az alábbiakat:

– Amennyiben ismert az eljárásunkból származó becslés eloszlása (pl. a mintaátlag alkalmazása ilyen), akkor zárt formulák segítségével meghatározható az eljárás standard hibája (vagy általánosságban hibája), melynek segítségével a becslésünk pontossága, konfidenciaintervalluma meghatározható. Ennek segítségével tehát képet kaphatunk arról, hogy a valószínűségi változó adott

lévén – $G_1 = 0$ értékkel rendelkezik, azaz a ferdesége 0. Magyarán, ha azt tapasztaljuk, hogy az eloszlásunk ferde – pozitív vagy negatív irányba „eldől", akkor a ferdeségi együttható segítségével korrigáljuk a próbastatisztikánk értékét.

paraméterének becslése esetén milyen hibákat követhetünk el: a véletlen jelenségre mennyire érzékeny a becslésünk.

- Amennyiben nem ismert az eljárásunk eloszlása, úgy szimulációs eljárások bevetésével tudunk képet kapni arról, hogy az adott minta sajátosságaiból következően milyen várható hibákat követünk el az adott paraméter vagy paraméterek becslése során.
- A szimulációk helyett – a kezdeti tapasztalatok sikeressége okán – említhetjük például Hampel és munkatársai, 2011-ben publikált simítási eljárását is [21], mely szintén alkalmazható lehet annak érdekében, hogy a becsléseink bizonytalanságát pontosabban meghatározhassuk.

Megjegyzés. Fontos kiemelni, hogy a fenti felsorolás messze nem teljes. Például nem szóltunk a Bayes-becslések problématikájáról, illetve azok érzékenységéről, ezen keresztül nem adtunk számot azokról az esetekről, amikor rizikó vagy információ (és nem közveltenül az eltérés) alapján akarjuk vázolni a becslés jóságát. Bayes-becslések érzékenységéről, annak vizsgálatáról és függéséről pl. az a-priori eloszlások[16] befolyásáról olvashatunk Lavine 1991-es cikkében [27].

Nem beszéltünk a hiányzó értékek problémájáról vagy arról, ha az adott változóval összefüggő más változókról is rendelkezünk információkról. Erről például Robins és munkatársai értekeznek könyvükben [39], ahol hiányzó értékek esetén való becslések érzékenységvizsgálatára találhatunk módszereket, lehetőségeket.

A témák szerteágazó volta miatt célunk nem is lehetett mindenre kiterjedő – továbbra is a kérdések felvetését tartjuk inkább fontosnak.

3.2. Hipotézisvizsgálatok

A hipotézisvizsgálatok nyilván jelentős mértékben összefüggnek az előző kérdéskörrel: amennyiben van becslésünk és tudjuk annak megbízhatóságát (konfidenciaintervallumát), akkor lényegében hipotézisekről is tudunk döntéseket hozni.

Azonban a hipotézisvizsgálat során több, egymástól funkciójában is igen eltérő hibát tudunk elkövetni.

3.6. Példa. Tegyük fel, hogy egy betegséget szeretnénk diagnosztizálni, melynél az is gondot jelent, ha valakit betegnek mondunk a vizsgálatok alapján – pedig nem az, illetve akkor is gondban vagyunk, ha kiengedjük kezelés nélkül, pedig szüksége lenne rá.

Gondolhatunk itt egy rákos megbetegedésre, aminél a hibás diagnózis bármely kimenetele veszélyeket rejt: ha nem kezeljük, akkor esetleg menthetetlenné válik a beteg, míg ha kezelünk egy egészséges pácienset például kemoterápiával, úgy könnyen megbetegíthetjük.

[16] A Bayes-féle becslésekben azt feltételezzük, hogy maga a vizsgált paraméter is egy véletlen változó, melynek az úgynevezett a-priori (tapasztalás előtti) eloszlása adott. A vizsgált paraméter a-posteriori (tapasztalás utáni) eloszlása nem más, mint az a-priori eloszlás minta esetén vizsgált feltételes eloszlása. A Bayes-becslés pedig az a-posteriori eloszlásból számított paraméterbecslés.

Nyilvánvalóan vannak betegségek, melyeknél valamely kimenetel nem hordoz ekkora kockázatot: ha megszúrom a mutatóujjamat egy tűvel és a baleseti sebész nem hajlandó egy teljes műtőstábot összehívni a problémám elhárítására, majd hazaküld – nagy valószínűséggel nem követ el végzetes hibát. Másik oldalról, ha egy egészséges embernek C-vitamint írok elő, várhatóan nem fog neki ártani, így nagyobb gondot sem fogok vele okozni.

A statisztikai érzékenységvizsgálatokra a hipotézisvizsgálatok során két területet fogunk bemutatni.

3.2.1. A próba erejének és szignifikanciájának vizsgálata

A próba ereje, illetve a szignifikancia minden esetben az eljárás érzékenységeként kezelhető.

- A szignifikancia és a korábban már tárgyalt konfidencia, (megbízhatóság) egymással lényegében megegyező fogalmak.

- A próba ereje egy bonyolultabb módon számolható paramétere a kiválasztott hipotézisvizsgálati eljárásnak. A próba ereje a vizsgálat úgynevezett másodfajú hibájával analóg fogalmak. A fenti példával élve, ha a nullhipotézisünk az, hogy a vizsgált páciensünk egészséges, úgy a másodfajú hibát akkor követjük el, amikor a betegeket nem részesítjük kezelésben.

3.7. Példa. A hibák kummulálódására az alábbi, általában ismert példát említjük. Több átlag összehasonlítását végezzük a varianciaanalízis során. Ekkor hagyományosan azt teszteljük, hogy több csoport átlaga egyezik-e egymással vagy sem. Világos, hogy a több átlag egyidejű, páronkénti összehasonlítása nem végezhető el független módon – és ilyen esetben az elsőfajú hibák valószínűségének viselkedéséről keveset tudunk.

A páros összehasonlítások úgynevezett „Post Hoc" tesztjeinek számos változata ismert, ezekből a teljesség igénye nélkül felsorolunk néhányat. A képletekben minden esetben szerepelni fog az MSE-érték, ami nem más, mint a csoportokon belüli átlagos négyzetes eltérés[17].

Továbbá általában feltételezzük, hogy ha k darab csoport van, akkor minden csoportban azonos, n esetszámmal dolgozunk (mutatunk egy olyan formulát is, ahol e feltételtől eltérhetünk).

Értelemszerűen két átlagot akkor nem fogunk szigkifikánsan különbözőnek tekinteni, ha a különbségük konfidenciaintervalluma tartalmazza a 0-t.

- Bonferroni-eljárás: Bonferroni azt javasolta, hogy ha α megbízhatósági szintű döntést szeretnénk hozni, de egymás után m darab tesztet kell végeznünk, me-

[17] Azaz a csoportok átlagaitól vesszük a csoportban lévő egyedek, részminták eltéréseinek négyzetösszegét és átlagoljuk – belső variancia, vagy hibavariancia néven is ismert. Szabadságfoka a fenti jelölésekkel $n-k$, azaz a teljes létszám és a csoportok számának különbsége.

lyek egymástól nem függetlenek, akkor α szint helyett $\frac{\alpha}{m}$ szinten döntsünk[18]. Fontos azonban megjegyeznünk, hogy ez általában feleslegesen szigorú eljárást jelent, így ezt általában finomítani szokás.

– Átlagok Bonferroni-összehasonlítása:

$$\overline{X}_{i,\bullet} - \overline{X}_{j,\bullet} \pm t_{(1-\frac{\alpha'}{2},\nu)}\sqrt{\frac{2MSE}{n}},$$

ahol $\frac{\alpha'}{2}$ tehát $\frac{\alpha}{2(m-1)}$, ahol m az összehasonlítások száma. $\overline{X}_{i,\bullet}$ az i-edik csoport átlagát jelöli, míg ν az MSE szabadságfoka.

– Átlagok Bonferroni-összehasonlításának Sidák-féle módosítása:

$$\overline{X}_{i,\bullet} - \overline{X}_{j,\bullet} \pm t_{(1-\frac{\alpha_m}{2},\nu)}\sqrt{\frac{2MSE}{n}},$$

ahol $\frac{\alpha_m}{2} = \frac{1-(1-\alpha)^{\frac{1}{m}}}{2}$.

– Átlagok Dunnett-féle összehasonlítása:

$$\overline{X}_{i,\bullet} - \overline{X}_{j,\bullet} \pm D_{(1-\alpha,k-1,\nu)}\sqrt{\frac{2MSE}{n}},$$

ahol D az úgynevezett Dunnett-eloszlás[19], k a csoportok száma, míg ν továbbra is MSE szabadságfoka.

– Átlagok Hsu-féle (MCB) összehasonlítása:

$$\overline{X}_{i,\bullet} - \max_{i\neq j} \overline{X}_{j,\bullet} \pm OD_{(1-\alpha,k-1,\nu)}\sqrt{\frac{2MSE}{n}},$$

ahol OD az egyoldali Dunnett-eloszlás.

– Átlagok Fisher-féle (LSD) összehasonlítása:

$$\overline{X}_{i,\bullet} - \overline{X}_{j,\bullet} \pm t_{(1-\frac{\alpha}{2},\nu)}\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)},$$

mely eljárás tehát alkalmazható különböző csoportlétszámok esetén is.

[18] Ilyenkor tehát a korábban már tárgyalt elsőfajú hiba valószínűségét drasztikusan lecsökkentjük.

[19] A Dunnett-eloszlásról általában táblázat segítségével döntenek [50]. A standard normális eloszlás esetén is az eloszlásfüggvény inverzének táblázatát használják a statisztikai számításoknál, hiszen az inverznek zárt alakja nincsen – így ez a táblázatos eljárás nem nevezhető szokatlannak. A táblázathoz használt, a standard normális eloszlás eloszlásfüggvényénél lényegesen bonyolultabb formula megtalálható például Dunlap és munkatársai cikkében [14], mely cikkben ráadásul több példát is bemutatnak ezen eloszlás alkalmazására.

Megfigyelhető volt, hogy az átlagok egyenlőségének tesztelésekor a részmintáink szórásának egyenlősége is feltételként szabható – vannak eljárások, ahol ez nem feltétel. Azonban a korábban már említett Lee és munkatársai is megfogalmazzák 2010-ben publikált anyagukban [28], hogy a szórások összehasonlítása – helyesebben a részminták szóródási mutatóinak egyezése vagy különbözősége – egyáltalán nem triviális kérdés. Ráadásul több tesztet is összehasonlítanak egymással szimulációk segítségével, így a különböző tesztek numerikus eredményeit is áttekinthetjük dolgozatukban.

3.8. Példa. Az alábbiakban összefoglalunk néhány tesztet Lee és munkatársainak cikkéből. Mindezt azért tesszük, hogy jobban rávilágíthasunk: amennyiben a vizsgált változónk normalitása sérül, úgy a már korábban elmondottak alapján nem csak a középértékek megválasztása lehet problematikus (átlag helyett például medián, átlagok összehasonlítása helyett sztochasztikus egyenlőség vizsgálat, lásd Wilcox könyvét [46]), hanem a szóródási mutatók megválasztása, vagy azok tesztelése sem egyértelmű.

Két szórás összehasonlítására a hagyományos eljárás az úgynevezett F-próba (a két variancia hányadosa alapján tesztel), melynek feltétele a normalitás és melynek megsértésre kifejezetten érzékeny (lásd például Klotz és Johnson dolgozatát, [26] akik – ahogyan a most idézett dolgozat is – az először ismertetendő tesztet, mint alternatívát ajánlják helyette).

Az alábbi tesztek tehát mind a $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ nullhipotézis eldöntésére szolgálnak.

– **Levene-teszt**: A Levene-teszt próbastatisztikája:

$$W = \frac{(N-k)\sum_{i=1}^{k} n_i \left(\overline{Z}_i - \overline{Z}\right)^2}{(k-1)\sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\overline{Z}_{ij} - \overline{Z}_i\right)^2},$$

ahol N a teljes mintanagyság, n_i az i-edik részminta nagysága, $Z_{i,j} = \left|Y_{ij} - \overline{Y}_i\right|$, $\overline{Y}_i$ az i-edik részminta átlaga, $\overline{Z}_i$ a Z_{ij}-k csoportjainak egyenkénti átlaga, míg $\overline{Z}$ a Z_{ij}-k főátlaga, azaz a Levene-teszt az átlagos abszolút eltéréssel számol az átlagos négyzetes eltérés helyett.

A fenti W-próbastatisztika H_0 fennállása esetén F-eloszlást követ $k-1$ és $N-k$ szabadságfokkal.

– **Módosított Levene-teszt**: lényegében azonos a fenti teszttel, csak átlagok helyett mindenhol a mediánt kell használni.

– **Z-variancia teszt**: Az Overall és Woodward által 1974-ben publikált [35] eljárás a következő alakot ölti. A próbastatisztika:

$$F = \frac{\sum_{i=1}^{k} Z_i^2}{k-1},$$

$$Z_i = \sqrt{\frac{c_i\,(n_i-1)\,s_i^2}{MSE}} - \sqrt{c_i\,(n_i-1) - \frac{c_i}{2}},$$

ahol $c_i = 2 + \frac{1}{n_i}$, s_i^2 a korrigált tapasztalati variancia, n_i az adott részminta mintanagysága, MSE pedig a már korábban ismertetett négyzetes eltérés.

Ekkor H_0 fennállása esetén Z_i eloszlása standard normális, tehát a fenti F-próbastatisztika eloszlása F-eloszlás, $k-1$ és ∞ szabadságfokkal.

– **Az Overall–Woodward-féle módosított Z-variancia teszt:** 1976-ban a már hivatkozott Overall és Woodward szerzőpáros újabb dolgozatukban [36] módosították az eredeti c_i értékeket az alábbira:

$$c_i = 2\left(\frac{2,9 + \frac{0,2}{n_i}}{K}\right)^{\frac{1,6\left(n_i - 1,8K + 14,7\right)}{n_i}},$$

ahol n_i továbbra is az i-edik részcsoport mintanagysága, továbbá:

$$Z_{i,j} = \frac{X_{i,j} - \overline{X}_i}{\sqrt{\frac{n_i-1}{n_i}}\,s_i^2},$$

$$K = \frac{\sum_{i,j} Z_{i,j}^4}{n_i - 2}.$$

– **O'Brien-teszt**: Az O'Brien által publikált próba [34] azt mondja, hogy a hagyományos F-próbát módosítsuk olymódon, hogy az eredeti próbában használt $Y_{i,j}$ értékeket módosítjuk az alábbi módszerrel:

$$V_{ij} = \frac{(n_i - 1,5)\,n_i\left(Y_{ij} - \overline{Y}_2\right)^2 - 0,5 s_i^2\,(n_i - 1)}{(n_i - 1)\,(n_i - 2)},$$

ahol az alábbi jelöléseket alkalmaztuk:

$$\overline{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i},$$

$$s_i^2 = \frac{\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y_i}\right)^2}{n_i - 1},$$

a megfelelő részcsoportátlagok és részcsoportvarianciák, tehát lényegében Y_{ij}-ket a fenti V_{ij} értékekre cseréljük, és úgy alkalmazzuk az eredeti F-próbát.

Megjegyzés. Megjegyezzük, hogy ilyenkor fennáll az alábbi egyenlőség:

$$s_i^2 = \overline{V}_i = \frac{\sum V_{i,j}}{n_i}.$$

Megállapíthatjuk, hogy amennyiben a normalitás nem teljesül, úgy a szóródási mutatóknál sem feltétlenül a szórást kell választani, hiszen látható, hogy a szórás nem feltétlenül a lehetséges legjobb, valamely középértéktől való átlagos eltérést mérő, jól interpretálható mennyiség.

3.3. Egy biostatisztikai megközelítés: ROC-görbék alkalmazása

A másik megközelítés a hipotézisvizsgálatok esetén a biostatisztika[20] egyik bevett eljárása. A továbbiakban a következő jelöléseket fogjuk alkalmazni:

Er	Érzékenység
Fa	Fajlagosság
$N_{+,v}$	Nem hibás, pozitív tesztek száma
$N_{+,h}$	Hibás pozitív tesztek száma
$N_{-,v}$	Nem hibás, negatív tesztek száma
$N_{-,h}$	Hibás negatív tesztek száma

Megjegyzés. Tehát $N_{+,v} + N_{-,h}$ a betegek, míg $N_{+,h} + N_{-,v}$ az egészséges egyedek száma. Érzékenységnek (sensitivity) nevezik annak valószínűségét, hogy egy beteget a teszt valóban betegnek mutat. Más megközelítésben:

$$Er = \frac{N_{+,v}}{N_{+,v} + N_{-,h}}.$$

Megjegyzés. Megjegyezzük, hogy egy másik, ezzel analóg fogalom is gyakran használatos a biostatisztikában. A fajlagosság (specificity) megmutatja, hogy mi a valószínűsége annak, hogy negatív tesztet kapunk abban az esetben, ha az illető tényleg egészséges. A fenti jelölésekkel:

$$Fa = \frac{N_{-,v}}{N_{-,v} + N_{+,h}}.$$

[20]Fontos megjegyezni, hogy az úgynevezett „túlélési statisztikák" e bevett grafikus elemzési eszközét számos területen – így nem csak a biostatisztikában – alkalmazzák. Így például a pénzügyi statisztikai eljárásokban is számos felhasználása ismert: egy betegségben való elhalálozás a cégek számára a csődeljárásként fogható fel. A modellek ilyen szempontból tehát rokonságban állnak egymással.

Az érzékenység és fajlagosság témájában is fontos mérnünk, hogy e két mennyiség milyen hibahatáron belül mozoghat. Ez lényegében nem más, mint annak mérése, hogy bizonyos statisztikai próbák első és másodfajú hibája miként alakul.

E biostatisztikai témakörben számos publikáció készült – melyek olykor pont e terület érzékenységvizsgálatát nem érintik. Erre hozható példaként Bender és munkatársainak elemzése [3] Brenner és Gefeller dolgozatáról [5], ahol a számításokat reprodukálva mutattak arra rá, hogy a becslésekben, melyeket a szerzők tettek, számos megkérdőjelezhető pont van.

Az orvoslásban persze adott egy igen egyszerű – bár nem feltétlenül költséghatékony ellenszere a téves diagnózisok szűrésének. Ez pedig nem más, mint amit Diepgen és Coenraads feszeget cikkükben [13]: több tesztet futtatnak egy-egy diagnózis felállítására. A több teszt futtatása, összefüggéseinek matematika sajátosságaira, statisztikai hibáinak kummulálódására vagy éppen szűkítésére hívják fel munkájukban a figyelmet egy igen konkrét diagnosztikai eljárás kapcsán.

Az orvosi alkalmazások során nyilván nem csak ilyen helyzetek adódnak. Egyes betegségek esetén a döntést és a becsléseket általában logisztikus regresszió[21] alkalmazásával és úgynevezett ROC-görbék elemzésével szokták megoldani.

3.1. Definíció. Tegyük fel, hogy adott k darab, $X_1, \ldots, X_k$ véletlen változó, melyek segítségével az Y bináris változó lehetséges értékeinek bekövetkezési valószínűségét szeretnénk meghatározni adott $x_1, \ldots, x_k$ realizáció esetén.

Világos, hogy $P(Y = 1)$ meghatározása elegendő, hiszen

$$P(Y = 1) + P(Y = 0) = 1.$$

A logisztikus regresszió modellje azt mondja, hogy

$$P\left(Y = 1 | X_1, \ldots, X_k\right) = \frac{e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k}}$$

alakban keresendő.

Innen is világosan látszik, hogy a logisztikus regresszió egyfajta lehetséges modellje a bekövetkezési valószínűség meghatározásának, adott realizáció mellett. A lábjegyzetben is olvasható, Boros Endre és munkatársai által jegyzett [8] cikk éppen e helyzetek másfajta megközelítésére ajánl alternatívát egy, a logiszitkus regresszió modelljétől teljesen más megközelítés alkalmazásával.

[21] Jegyezzük meg, hogy nem csak logisztikus regressziót lehetne alkalmazni egy-egy ilyen osztályozási eljárás során. Például Boros és munkatársainál könyvet is olvashatunk [7] a Logical Analysis of Data (LAD) eljárásról, mely szintén egy bináris osztályozás, ahol azonban nem statisztikai, hanem optimalizálási technikák segítségével dolgoznak. Konkrét implementációit is adják Boros és szerzőtársai dolgozatukban [8], ahol pszichometriai, műszaki és gazdasági adatokon egyaránt bemutatják eljárásukat, numerikus eredményekkel alátámasztva. Érdemes tehát arról is tudnunk, hogy a logisztikus regresszió nem feltétlenül az egyetlen olyan eljárás, melynek segítségével bináris változók eloszlásáról szerezhetünk információt – sőt.

Megjegyzés. A ROC-görbékkel az egységnégyzetben ábrázolják a érzékenység (sensitivity) és fajlagosság (specificity) közötti összefüggéseket. Míg az x tengelyen az $1 - Fa$, addig az y tengelyen az Er érték (arány) helyezkedik el.

Bár számos helyen fellelhető e módszer (lásd például [43]), egy egyszerű példán keresztül könnyen bemutatható mind az alkalmazás, mind pedig a görbe elkészítésnek metódusa. A ROC-görbéhez e feladat Buza Krisztián jegyzete [48] alapján készült.

3.9. Példa. Tegyük fel, hogy lázat szeretnénk mérni, láz alapján pedig valamely betegséget diagnosztizálni, mely betegség általában lázzal jár – de persze nem minden esetben, illetve nem minden lázas szenved ebben a betegségben.

A mintánkat már testhőmérséklet szerint sorrendbe rendeztük a jobb átláthatóság kedvéért.

V	-	-	-	-	-	-	-	-	+	-	+	+
M	36,4	36,4	36,5	36,6	36,6	36,6	36,7	36,8	37,5	37,6	39	39,2

Azaz: a valóságban (V) a „-” jel azt mondja, hogy egészséges, nem szenved e specifikus betegségben, míg a „+” azt mondja, hogy beteg. A modellben (M) pedig a testhőmérsékletekkel modellezünk, tehát azzal szeretnénk mérni, diagnosztizálni.

A ROC-görbéhez ki kell számolnunk az igazi pozitív ($N_{+,v}$), a hamis pozitív ($N_{+,h}$), igazi negatív ($N_{-,v}$) és hamis negatív ($N_{-,h}$) értékeket. Szükségünk lesz az igazi pozitívok (Er) és a fals pozitívok ($1 - Fa$) arányára a betegek és az egészségesek között a testhőmérséklet különböző, értelmes értékei esetén, hiszen az y és x tengelyek rendre ezeket az arányokat mutatják.

A testhőmérséklet különböző szintjein kell hát eldönteni, hogy hány helyes és hány helytelen diagnózis lenne a fent adott modellel a betegséget illetően (tehát a táblázat első 4 sorában az adott módon besorolt betegek számát jelöljük, az alsó két sorban pedig az arányokat). A sorok elején a már korábban definiált jelöléseket használjuk.

A táblázat első sorában az értelmes testhőmérséklet vágópontokat tüntettük fel. Amely értékből több is volt, azt zárójelben szerepeltetjük.

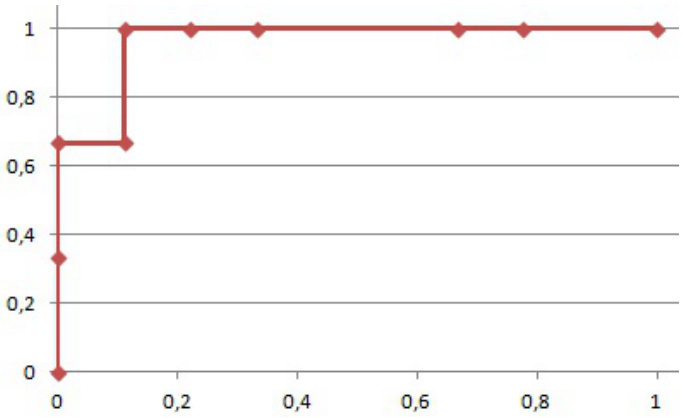
Hőmérséklet	36,4 (2)	36,5	36,6 (3)	36,7	36,8	37,5	37,6	39	39,2	FIN
$N_{+,v}$	3	3	3	3	3	3	2	2	1	0
$N_{+,h}$	9	7	6	3	2	1	1	0	0	0
$N_{-,v}$	0	2	3	6	7	8	8	9	9	9
$N_{-,h}$	0	0	0	0	0	0	1	1	2	3
Er	1	1	1	1	1	1	2/3	2/3	1/3	0
$1 - Fa$	1	7/9	6/9	3/9	2/9	1/9	1/9	0	0	0

A táblázat kitöltésének módjára vegyünk egy konkrét cellát. Az $N_{+,h}$ sorban tehát azt vizsgáljuk, hogy a testhőmérséklet adott értékének vágópontként való

definiálásával hány darab fals, pozitív eredményt kapnánk. Így például ha $36,5C^\circ$-os testhőmérésékletet vágópontként kezelve, a $36,4C^\circ$-os pácienset nem tekintenénk betegnek, azonban továbbra is maradna 7 darab fals, valóságban egészséges páciensünk, akiket betegnek jeleztünk (és lenne természetesen 3 darab, valóságban beteg páciensünk helyesen azonosítva).

Általánosságban: ha a görbe átmegy az egységnégyzet bal felső sarkán, akkor téves diagnózis nélküli eljárást sikerült alkotni. Minden görbe esetén fontos tehát annak alakja, hiszen minél jobban közelíti a görbe a bal felső sarkot, annál precízebb, pontosabb diagnózist lehet az eljárással felállítani. Azonban a görbe alakján kívül a görbe alatti területnek is jelentése van: lényegében a tesztünk hatékonyságának mérőszáma (a bal első sarkon átmenő esetben a terület 1, tehát ilyenkor a leghatékonyabb, míg egy olyan görbe esetén, ami a négyzet bal alsó sarkát a jobb felső sarokkal összekötő átlóját mutatja lényegében pénzt is dobálhatnánk döntéshozatal helyett).

A példánkhoz tartozó ábrát az utolsó két sor alapján elkészítettük, tehát az alsó két sorban található értékek a görbe koordinátái:



1. ábra. ROC-görbe az igazi pozitív és fals pozitív arányok szerint

Az ábra elég jól közelíti a bal felső sarkot, tehát azt mondhatjuk, hogy a fenti példában egy kellően jól viselkedő modellt tudtunk alkotni: a görbe alatti terület $\frac{26}{27}$, tehát a helyes diagnózisok valószínűsége magasnak mondható.

A döntéshozatalra, illetve alkalmazásukra számos példa hozható fel – pusztán a módszert és annak értelmezését láthatjuk Goldstein és munkatársai, 1906 öngyilkosságot túlélteken elvégzett pszichiátria kutatásában, illetve annak dokumentációjában [19].

Egy elméleti, a ROC-görbék elemzésében alkalmazott mennyiségek χ^2 statisztikák segítségével vizsgáló cikk olvasható Bennettől [4], aki teljesen elméleti megkö-

zelítésben tárgyalja – majd saját vizsgálati eredményein teszteli is a diagnosztikai eljárások ilyedtén való becslését, illetve becslésének jóságát.

3.4. Megjegyzések a hipotézisvizsgálatokhoz

Nem érintettük itt a hipotézisvizsgálatok során az összes létező lehetőséget a próbák lehetséges hibáinak tesztelésére. Világos, hogy minden statisztikai próba csak bizonyos – szigorúbb vagy kevésbé szigorú – feltételek mellett viselkedik optimálisan. E feltételek sérülése esetében különböző robusztus eljárások választhatók – azonban e választások során sem elhanyagolható, hogy a „hagyományos" eljárás feltételei, mely eljárás helyett most e robusztusat választottuk, milyen mértékben sérülnek.

A sérülés mértékének, minőségének következményeire ritkán találhatunk egzakt módon is igazolható, megbízható és kalkulálható eljárásokat – azaz, amit például a student-féle t-próba esetén jól körüljárható területnek gondolunk.

A t-próba esetén a ferdeség, csúcsosság – vagy általánosabban a normalitás hiánya esetén választható robusztus tesztek megbízhatóságára Vargha 2003-as cikke [45], vagy a próba erejének vizsgálatára a normalitás sérülése esetén Srivastava 1958-as dolgozata [42] lehet példa. Ez más hipotézisvizsgálati módszerek esetén messze nem tűnik kérdések nélküli területnek, illetve elméleti háttere – a fellelhető szakirodalmak alapján – nem látszik ennyire körüljártnak.

4. Összefoglalás

E témában több összefoglaló mű is született, melyek támpontot, kiindulási alapot adhatnak a különböző statisztikai tesztek, illetve azok robusztus változatainak megismeréséhez (példaként említhetjük összefoglaló anyagként Wilcox könyvét [46], melyből számos hagyományos módszert, és azok több robusztus változatát is megismerhetjük).

Megállapítható, hogy a statisztikai vizsgálatok jelentős hányada a bemeneti adatok változásait vagy változékonyságát – amiatt, hogy eleve valószínűségi változókkal dolgozik, melyek szükségszerűen változékonyak kisebb-nagyobb mértékben – kezelik valamilyen formában. A leggyakrabban ez olymódon jelenik meg, hogy az eljárások megfelelő biztonsági szinten való alkalmazását feltételekhez kötik (a véletlen változó eloszlásának pl. normális volta, csoportok szórásának homogenitása, stb). Amennyiben e feltételek sérülnek, úgy az adott eljárás valamely korrekciós – robusztus – változatát javasolják. Ezen esetben az eljárásokban mindenképpen jelen lévő hibákat (hiszen véletlen jelenségek alapján hozunk döntéseket) általában megfelelő szinten lehet tartani.

Más esetekben viszont nem ismertek azok a matematikai alapok és vizsgálatok, melyek biztosítanák az eljárást alkalmazók számára azokat a stabilitási kritériumokat, melyekkel a hibás döntések valószínűsége meghatározható, uralható. Így pl.

empirikus eszközök segítségével – szimulációs eljárások – az adott tapasztalati eloszlások vizsgálatával kimérhetők az alkalmazott eljárások hibái. Ha a hibákat e módszerekkel nem is tudjuk kiküszöbölni, azok mértékével tisztában lehetünk – és így továbbra is megalapozott döntések hozhatók.

Szintén empirikusak, de nem feltétlenül igényelnek nagyobb gépigényt – illetve a kezdeti tapasztalatok alapján kis minták esetén is működőképes alternatívát jelenthetnek – a simítási eljárások. Segítségükkel robusztus becslések készíthetők – stabilabbá, kevésbé érzékennyé téve így az eljárásunkat, illetve a segítségükkel meghozott döntéseinket.

Természetesen – ahogy jeleztük, nem törekedtünk cikkünkben a statisztika minden területének lefedésére. Nem beszéltünk például a különböző regressziós technikák megbízhatóságáról, a Bayes-becslések érzékenységéről vagy az idősorok esetén alkalmazható különböző technikákról és felmerülő problémákról. Célunk pusztán az volt, hogy két, egyszerűbb területet kiragadva, azok segítségével vázoljuk a probléma általános mivoltát, nagyságát és fontosságát.

Hivatkozások

[1] Bayne, W.; Tobet, S.; Mattiace, L. A.; Lasco, M. S.; Kemether, E.; Edgar, M. A.; Morgello, S.; Buchsbaum, M. S.; Jones, L. B.: *The interestitial Nuclei of the Human Anterior Hypothalamus: An Inverstigation of Variation with Sex, Sexual Orientation, and HIV Status*, Hormones and Behavior, Vol. **40/2**, pp.: 86–92, 2001.

[2] Belia, S.; Fidler, F.; Williams, J.; Cumming, G.: *Researchers Misunderstand Confidence Intervals and Standard Error Bars*, Psychological Methods, Vol.: **10/4**, pp.: 389–396, 2005.

[3] Bender, R.; Langue, S.; Freitag, G.; Trampisch, H. J.: *Letters to the Editor on „Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence*, Statistics in Medicine, Vol. **16**, pp.: 981–991, 1997.", Statistics in Medicine, Vol. **17**, pp.: 945–950, 1998.

[4] Bennett, B. M.: *On comparisons of sensitivity, specificity and predictive value of a number of diagnostic procedures*, Biometrics, Vol. **28**, pp.: 793–800, 1972.

[5] Brenner, H.; Gefeller, O.: *Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence*, Statistics in Medicine, Vol. **16**, pp.: 981–991, 1997.

[6] Bolla, M.; Krámli, A.: *Statisztikai következtetések elmélete*, Typotex, 2005.

[7] Boros, E.; Hammer, P. L.; Ibaraki, T.: *Logical Analysis of Data*, IGI-Global, 2005.

[8] Boros, E.; Hammer, P. L.; Ibaraki, T.; Kogan, A.; Mayoraz, E.; Muchnik, I.: *An implementation of logical analysis of data*, Knowledge and Data Engineering, Vol. **12/2**, pp.: 292–306, 2000.

[9] Borovkov, A. A.: *Matematikai Statisztika*, Typotex, 1999.

[10] Camponovo, L.; Otsu, T.: *Breakdown pont theory for implied probability bootstrap*, The Econometrics Journal, Vol. **15/1**, pp.: 32–55, 2012.

[11] Cochran, W. G.: *The distribution of quadratic forms in a normal system, with applications to the analysis of covariance*, Mathematical Proceedings of the Cambridge Philisophical Society, Vol. **30/2**, pp.: 178–191, 1934.

[12] Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences*, New York, 1988.

[13] Diepgen, T. L.; Coenraads, P. J.: *Sensitivity, specificity and positive predictive value of patch testing: the more you test, the more you get?*, Contact Dermatitis, Vol. **42/6**, pp.: 315–317, 2000.

[14] Dunlap, W. P.; Marx, M. S.; Agamy, G,J.: *Fortain IV functions for calculating probabilities associated with Dunnett's test*, Behavior Research Methods and Instrumentation, Vol. **13/3**, pp.: 363–366, 1981.

[15] Efron, B.; Tibshirani, R.: *Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy*, Statistical Science, Vol. **1/1**, pp.: 54–75, 1986.

[16] Fletcher, D.; Webster, R.: *Skewness-Adjusted confidence Intervals on Stratified Biological Surveys*, Journal of Agricultural, Biological and Environment Statistics, Vol. **1/1**, pp.: 120–130, 1996.

[17] Gayen, A. K.: *The distribution of „Student's" t in random samples of any size drawn from non-normal universes*, Biometrika, Vol. **36**, pp.: 353–369, 1949.

[18] Geyer, C. J.: *Breakdown Point Theory Notes*, http://www.stat.umn.edu/geyer/5601/notes/break.pdf, (letöltés: 2012. 10. 16.)

[19] Goldstein, R. B.; Black, D. W.; Nasrallah, A.; Winkour, G.: *The Prediction of Suicide, Archives of General Psychiatry*, Vol. **48/5**, pp.: 418–422, 1991.

[20] Hall, P.: *On the removal of skewness by tranformation*, Journal of the Royal Statistics Society, Vol. **54**, pp.: 221–228, 1992.

[21] Hampel, F.; Hennig, C.; Ronchetti, E.: *A smoothing principle for the Huber and other location M-estimators*, Computational Statistics and Data Analysis, Vol. **55**, pp.: 324–337, 2011.

[22] Huber, P. J.: *Robust Estimation of a Location Parameter*, Annals of Matematical Statistics, Vol. **35/1**, pp.: 73–101, 1964.

[23] Johnson, N. J.: *Modified t tests and confidence intervals for asymmetrical distributions*, Journal of the American Statistical Association, Vol. **73/363**, pp.: 536–544, 1978.

[24] Jones, D. N.; Gill, C. A.: *Comparing Measures of Sample Skewness and Kurtosis*, The Statistician, Vol. **47/1**, pp.: 183–189, 1998.

[25] Judkins, D. R.: *Fay's method for variance estimation*, Journal of Official Statistics, Vol. **6**, pp.: 223–239, 1990.

[26] Klotz, S.; Johnson, N. L.: *Breakthroughs in Statistics*, Foundations and Basic Theory, Vol. **1**, pp.:680, 1993.

[27] Lavine, M.: *Sensitivity in Bayesian Statistics: The Prior and the Likelihood*, Journal of the American Statistics Association, Vol. **86/414**, pp.: 396–399, 1991.

[28] Lee, H. B.; Katz, G. S.; Restori, A. F.: *A Monte Carlo Study of Seven Homogeneity of Variance Tests*, Journal of Mathematics and Statistics, Vol. **6/3**, pp.: 359–366, 2010.

[29] Lehmann, E. L.: *Theory of Point Estimation*, John Wiley and Sons, New York, 1983.

[30] Lehmann, E. L.: *Testing Statistical Hypotheses*, John Wiley and Sons, New York, 1959.

[31] LeVay, S.: *A difference in Hypothalamic Structure between Heterosexual and Homosexual Man*, Science, New Series, Vol. **253**, No. **5023**, pp.: 1034–1037, 1991.

[32] Mameli, V.; Music, M.; Sauleau, E.; Biggeri, A.: *Large sample confidence intervals for the skewness parameter of the skew-normal distribution based on Fisher's information*, Journal of Applied Statistics, Vol. **39/8**, pp.: 1693–1702, 2012.

[33] Mogyoródi J.; Michaletzky Gy.: *Matematikai statisztika*, ELTE, TTK, Nemzeti Tankönyvkiadó, Budapest, 1995.

[34] O'Brien, R. G.: *Robust tschniques for testing heterogeneity of variance effects in factorial designs*, Psychometrika, Vol. **43**, pp.: 327–342, 1978.

[35] Overall, J. E.; Woodward, J. A.: *A simple test for homogeneity of variance in complex factorial design*, Psychometrika, Vol. **39**, pp.: 311–318, 1974.

[36] Overall, J. E.; Woodward, J. A.: *A robust and powerfull test for heterogeneity of variance*, University of Texas, Medical Branch Psychometric Laboratory, 1976.

[37] Prékopa, A.: *Valószínűségelmélet műszaki alkalmazásokkal*, Műszaki Könyvkiadó, Budapest, 1962.

[38] Rényi A.: *Valószínűségszámítás*, Tankönyvkiadó, Budapest, 1968.

[39] Robins, J. M.; Rotniczky, A.; Scharfstein, D. O.: *Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models*, SZERK: Holloran, M. E.; Berry, D.: *Statistical Models in Epidemiology, The Environment, and Clinical Trials*, Vol. **116**, Springer, 2000.

[40] Saavedra, P. J.: *An extension of Fay's method for Variance Estimation to the Bootstrap*, Proceeding to the Annual Meeting of the American Statistical Association, August 5–9, 2001.

[41] Sak, H.; Hörman, W.; Leydold, J.: *Better Confidence Intervals for Importance Sampling*, Research Report Series, Rep. 106, Institute for Statistics And Mathematics, Wirtschafts Universitat Wien (Vienna University of Economics and Business), 2010.

[42] Srivastava, A. B. L.: *Effect of non-normality on the power function of t-test*, Biometrika, Vol. **45/3/4**, pp.:421–430, 1958.

[43] Takahashi, K.; Uchiyama, H.; Yanagisawa, S.; Kamae, I.: *The Logistic Regression and ROC Analysis of Group-based Screening for Predicting Diabetes Incidence in Four Years*, Kobe J. Med. Sci., Vol. **52** (6), pp.: 171–180, 2006.

[44] Takács Sz.: *Egy nem hagyományos statisztikai eljárás bemutatása az OECD PISA adatbázison – esettanulmány*, Alkalmazott Matematikai Lapok, Vol. **27**., 157–174, 2010.

[45] Vargha, A.: *Robusztussági vizsgálatok az egymintás t-próbával*, Statisztikai Szemle, Vol. **81/10**, pp.: 872–890, 2003.

[46] Wilxoc, R. R.: *Applying Contemporary Statistical Techniques*, Academic Press, 2003.

[47] Wright, D. B.; Herrington, J. A.: *Problematic standard errors and confidence intervals for skewness and kurtosis*, Behavior Research Methods, Vol. **43/1**, pp.: 8–17, 2011.

[48] http://cs.bme.hu/ buza/edu/dm techn/dm feladatok.pdf (letöltve: 2012. 11. 16.).

[49] http://www.nsf.gov/statistics/nsf03302/pdf/setables.pdf (letöltés ideje: 2012. 10. 02.).

[50] http://www.watpon.com/table/dunnetttest.pdf (letöltés ideje: 2012. 11. 16.).

(Beérkezett: 2012. november 30.)

TAKÁCS SZABOLCS
Károli Gáspár Református Egyetem
Bölcsésztudományi Kar, Pszichológiai Intézet, Általános lélektani és módszertani tanszék
1037, Budapest, Bécsi út 324, 5. épület, fszt.
e-mail: tretarkhon@gmail.com

SENSITIVITY ANALYSIS IN A STATISTICAL PROCESSES

Szabolcs Takács

An important aspect of many mathematical process is sensitivity analysis. In these analysis we investigate the change of output data – result and behavior – when changes are made to the input. It is of interest what type of changes in the input doesn't affect the results – or which type of modifications in the inputs results in larger or smaller scale changes to the output.

In the various fields of statistical processes, sensitivity has different a meaning. As an example, it has different meaning in estimation or in hypothesis theory – or in the different modelling processes.

In this paper we are not aiming to address all the various questions about sensitivity in the fields of statistics – instead we embark on providing an insight to the wide spectrum of the applications involved with sensitivity analysis, while also drawing attention to the importance of these analysis.

The paper will not state new theorems – but rather it raises several open questions of interest which have arisen in recent statistical research projects.