SOKASÁGOK ÖSSZEHASONLÍTÁSA ÚJ MÓDSZEREKKEL*

VARGHA ANDRÁS

A társadalomtudományi jelenségek empirikus kutatásaiban komoly erőfeszítéseket tesznek azért, hogy a vizsgált változók értékskálája eleget tegyen minimálisan az ordinalitás, vagyis a rangsorskála kritériumának. Ezzel párhuzamosan az elmúlt években számos új statisztikai eljárást dolgoztak ki az ordinális változókkal történő összehasonlításokra. A jelen tanulmány kitér az ilyen – ún. sztochasztikus – összehasonlítások egyik nehezen értelmezhető fonákságára, a sztochasztikus körbeverés jelenségére, amikor is az összehasonlítás során sérül a nagyság szerinti rendezés tranzitivitása. Tanulmányunk rávilágít arra, hogy az intranzitivitás hátterében gyakran nem normális eloszlások speciális egyenetlenségei rejtőznek, melynek következtében bizonyos diszkriminációs információk egy-egy kritikus skálapontban sűrűsödnek. Az ilyen skálapontok felderítése az eloszlások részletes összehasonlításával végezhető el, amelyre cikkünkben egy új módszert is bemutatunk. Ezután több valódi empirikus vizsgálat adataival demonstráljuk azt az érdekes jelenséget, hogy ha erősen nem normális eloszlású többértékű skálákat binarizálunk a fentebb leírt kritikus pontokban, akkor ezzel nemcsak hogy nem veszítünk feltétlenül információt, hanem esetenként hatékonyabb csoportdiszkriminációt és pontosabb regressziós előrejelzést tehetünk. Bár tanulmányunkban csak pszichológiai alkalmazásokra térünk ki, az itteni módszertan társadalmi és gazdasági problémák leírására is alkalmazható.

TÁRGYSZÓ: Sztochasztikus összehasonlítás. Sztochasztikus rendezés. Eloszlásfüggvények. Binarizálás.

Az empirikus társadalomtudományi kutatások egyik alapkérdése, hogy valamely X változó értékszintje ugyanakkora-e különböző sokaságokban. E kérdés vizsgálatára hagyományosan a várható értékek (elméleti átlagok) egyenlőségének nullhipotézisét szokták megfogalmazni, amely független minták esetén a kétmintás t-próba, illetve a varianciaanalízis ismert módszerével tesztelhető (*Vincze* [1968]). Mivel ezen statisztikai eljárások alkalmazási feltételei (normalitás és szóráshomogenitás) gyakran sérülnek (lásd *Micceri* [1989]; *Wilcox* [1996]), alternatív módszerek lehetőségét is szükséges megvizsgálni.

Korábbi tanulmányainkban (*Vargha* [2002], [2004]; *Vargha–Delaney* [1998], [2000]) részletesen ismertettük a sztochasztikus összehasonlítás modelljét, amely szélsőségesen nem normális eloszlású kvantitatív, vagy rangsorskálát alkotó, azaz ordinális kvalitatív változók esetén egyik jó alternatívája az átlagok összehasonlításának.

* A tanulmány megírásához nagy segítséget nyújtott a T047144 számú OTKA pályázat.

A SZTOCHASZTIKUS ÖSSZEHASONLÍTÁS EGYES JELLEMZŐI

A sztochasztikus összehasonlítás kulcsfogalma a valószínűségi fölény mutatója, amelyet a következőképpen definiálunk. Tételezzük fel, hogy az S_1, S_2, .., S_h ($h > 1$) sokaságokat szeretnénk összehasonlítani egy legalább ordinális skálájú X változó segítségével. Jelölje az X változót az S_i sokaságban X_i. Ekkor bármely $i \neq j$ esetén az S_i sokaság S_j-vel szembeni valószínűségi fölényét vagy sztochasztikus dominanciáját az

$$A_{ij} = P\left(X_i > X_j\right) + 0{,}5 \cdot P\left(X_i = X_j\right) \qquad /1/$$

kifejezéssel definiáljuk. Ez lényegében azt fejezi ki, hogy ha az S_i és az S_j sokaságból egymástól függetlenül kiválasztunk egy-egy véletlen X-értéket, mi lesz a valószínűsége annak, hogy az S_i sokaságból származó megfigyelés nagyobb lesz az S_j-ből származónál (egyenlőség esetén igazságosan felezünk). A_{ij} láthatóan 0 és 1 közötti értéket vehet csak fel, s A_{ji} ellentettjével 1-re egészítik ki egymást: $A_{ij} + A_{ji} = 1$. Ha $A_{ij} = A_{ji} = 0{,}5$, akkor azt mondjuk, hogy az S_i és az S_j sokaság az X változó szempontjából sztochasztikusan egyenlő. Abban az esetben, amikor $A_{ij} < 0{,}5$ (vagy $A_{ij} > 0{,}5$), azt mondjuk, hogy S_i sztochasztikusan kisebb (nagyobb), mint S_j. A továbbiakban a sztochasztikusan kisebb, egyenlő, illetve nagyobb relációk jelölésére rendre a $<_{szt}$, $=_{szt}$, $>_{szt}$ szimbólumokat fogjuk használni (például $S_i <_{szt} S_j$ vagy $X =_{szt} Y$ vagy $X_i >_{szt} X_j$).

Adott S_1, S_2, ..., S_h sokaságok esetén az A_{ij} értékek a sokaságok páronkénti vagy lokális sztochasztikus dominancia viszonyait jelzik. A teljes együttesen belüli, ún. globális dominancia viszonyok mérhetők például a P_i sztochasztikus kezelési hatások segítségével, melyekhez úgy jutunk, hogy rögzített i index esetén az S_i-hez tartozó összes A_{ij} lokális dominancia értéket átlagoljuk:

$$P_i = \frac{1}{h} \sum_{j=1}^{h} A_{ij} \qquad /2/$$

($i = j$ esetén értelemszerűen $A_{ii} = 0{,}5$). P_i azt mutatja meg, hogy ha S_i-ből és az S_j-k egyesítésével létrejövő S összsokaságból véletlenszerűen és egymástól függetlenül kiválasztunk egy-egy X-értéket, mi lesz a valószínűsége annak, hogy az S_i sokaságból származó megfigyelés nagyobb lesz, mint az S-ből származó (egyenlőség esetén igazságosan felezünk). Megjegyezzük, hogy ha az S_i sokaságok mérete eltérő, akkor /2/-t célszerű súlyozott átlagként definiálni, ahol a súlyok a sokaságok méretével arányosak (lásd *Vargha* [2002], /7/ formula).

Tetszőleges $h > 1$ esetén az S_i sokaságok együttesét sztochasztikusan homogénnek nevezzük, ha az egyes sokaságokat jellemző sztochasztikus kezelési hatások egymással mind megegyeznek:

$$P_1 = P_2 = \ldots = P_h = 0{,}5. \qquad /3/$$

A sztochasztikus homogenitás fennállása azt jelenti, hogy az S_i sokaságok között nincs olyan, amelyikben az X változó értékei általában nagyobbak vagy általában kisebbek lennének, mint a többi sokaságban. A sztochasztikus egyenlőség (SZTE) és a sztochasztikus homogenitás (SZTH) nullhipotézisének tesztelésére alkalmas statisztikai próbákkal kapcsolatban lásd *Delaney–Vargha* [2002)], illetve *Vargha* [2002].

$h = 2$ esetén az SZTE és az SZTH egymással ekvivalens, $h > 2$ mellett – a modell újszerűsége miatt – azonban az alábbi érdekes jelenségek figyelhetők meg.

1. Az SZTH nem vonja maga után a sokaságok páronkénti SZTE-jét, vagyis előfordulhat, hogy az $S_1, S_2, ..., S_h$ ($h > 2$) sokaságok sztochasztikusan homogén együttest képeznek, miközben páronként sztochasztikusan eltérnek egymástól ($A_{ij} \neq 0{,}5$, ha $i \neq j$).

2. Általános esetben az is előfordulhat, hogy a $P_i - P_j$ különbségek által jelzett globális sztochasztikus viszonylatok ellentétesek az A_{ij} segítségével definiált lokális viszonylatokkal, vagyis például $P_i > P_j$ esetén $A_{ij} < 0{,}5$.

3. A sztochasztikusan kisebb reláció nem tranzitív, vagyis $A_{ij} < 0{,}5$ és $A_{jk} < 0{,}5$ együttes fennállása nem vonja maga után minden esetben az $A_{ik} < 0{,}5$ reláció fennállását. Olyan eset is előfordul, amikor S_1, S_2 és S_3 sztochasztikusan homogén együttest képez, miközben a három sokaság egymást sztochasztikusan „körbeveri”:

$$A_{12} < 0{,}5,\ \ A_{23} < 0{,}5 \ \text{ és }\ A_{31} < 0{,}5,$$

vagyis

$$S_1 <_{\text{szt}} S_2 <_{\text{szt}} S_3 <_{\text{szt}} S_1 \,.$$

Mindezen körülmények esetenként jelentősen megnehezíthetik a sztochasztikus összehasonlítások szakmai értelmezését. Miként lehet megmagyarázni például azt, hogy két terápiás eljárás, oktatási módszer vagy termelési eljárás lokálisan más nagyságrendi viszonyban van egymással, mint globálisan? Végül is milyen alapon dönthetjük el, hogy melyikük a jobb (hatékonyabb, eredményesebb stb.)? A jelen tanulmányban ezekre a kérdésekre is szeretnénk a gyakorlat számára kielégítő választ találni.

Először bemutatunk egy konkrét példát a sztochasztikus körbeverésre, és rávilágítunk arra, hogy ezen jelenség hátterében az áll, hogy az összehasonlított eloszlások dominancia viszonyai esetenként páronként más-más mérce szerint dőlnek el. Ezután megmutatjuk, hogy mindez szoros kapcsolatban van a valószínűségi fölény *A* mutatójának matematikai jelentésével, amely az összehasonlított két változó eloszlásfüggvényének egyfajta súlyozott különbsége segítségével is definiálható. Ezzel felhívjuk a figyelmet arra, hogy nem normális eloszlású változók, illetve az additív eltolástól számottevően különböző kísérleti hatások esetén fontos az eloszlások teljes vertikumának összehasonlítása. Ezt az elemzést valódi életből származó példák segítségével illusztráljuk, és egyben módszert adunk arra, hogy milyen módon lehet azonosítani az eloszlások olyan karakterisztikus pontjait, amelyek segítségével a függő változót dichotomizálva esetenként statisztikailag nagyobb magyarázó erejű új változókhoz jutunk.

PÉLDA SZTOCHASZTIKUS KÖRBEVERÉSRE

Vegyünk három szabályos dobókockát, melyeken rendre az alábbi számok láthatók:

A kocka:	1	1	4	4	4	4
B kocka:	2	2	2	2	5	5
C kocka:	3	3	3	3	3	3

Ha bármelyik kockával szabályosan dobunk, a dobás eredménye egy véletlen változó lesz. Ha az A és a B kockával egymástól függetlenül dobunk, az esetek 5/9 részében B nagyobb számot mutat. B nyerési esélye C-vel szemben 1:2, tehát C-vel várhatóan kétszer olyan gyakran nyerünk, mint az A-nál nagyobb nyerési esélyű B-vel. Ugyanakkor A-t és C-t összevetve, C az esetek 2/3 részében kikap A-tól, vagyis a sztochasztikus körbeverés esetével állunk szemben. Matematikai jelöléssel:

$$A <_{szt} B <_{szt} C <_{szt} A.$$

Hogyan értelmezhetünk egy ilyen körbeverést, ha a három véletlen változó mondjuk három különböző pszichológiai kezelés hatásosságának valamilyen mérőszáma?

Ha három sakkozó körmérkőzésén tapasztalnánk hasonló jelenséget, hamar kész volnánk a magyarázattal: A, B és C nagyjából azonos tudású sakkozók, de B különleges érzékkel talál rá A gyengéjére, miközben neki valahogy nem fekszik C stílusa, ami ellen viszont A tudja sikeresen felvenni a harcot. Mindez csak azért lehetséges, mert a sakkban valakivel szemben győzedelmeskedni többféle stílussal, játékfelfogással is lehet, és ha eltekintünk a szintén lehetséges formaingadozástól, a körbeverés azt jelzi, hogy a páros csaták esetenként más-más mérce szerint dőlnek el.

Ugyanez érvényes populációk, változók és statisztikai eloszlások sztochasztikus összehasonlítása esetén is. Például a fenti három dobókockának megfelelő eloszlás (két tizedesre kerekítve) az 1. táblában van összefoglalva.

1. tábla

A három kockadobás eredményének eloszlása

Megnevezés	Értékek					*Összesen*
	1	2	3	4	5	
A kocka	0,33	0	0	0,67	0	*1,00*
B kocka	0	0,67	0	0	0,33	*1,00*
C kocka	0	0	1,00	0	0	*1,00*

A körbeverés következménye, hogy mindhárom kockához található az értékskálának egy olyan dichotomizálása, amelynél a „nagy” értékek az adott kockánál fordulnak elő a legnagyobb valószínűséggel (lásd a 2–4. táblát).

2. tábla

Az „A” kocka dominanciája

Megnevezés	„Kis” értékek	„Nagy” értékek	*Összesen*
	1–3	4–5	
A kocka	0,33	**0,67**	*1,00*
B kocka	0,67	0,33	*1,00*
C kocka	1,00	0	*1,00*

3. tábla

A „B” kocka dominanciája

Megnevezés	„Kis” értékek	„Nagy” értékek	*Összesen*
	1–3	4–5	
A kocka	1,00	0	*1,00*
B kocka	0,67	**0,33**	*1,00*
C kocka	1,00	0	*1,00*

4. tábla

A „C" kocka dominanciája

Megnevezés	„Kis" értékek 1–3	„Nagy" értékek 4–5	*Összesen*
A kocka	0,33	0,67	*1,00*
B kocka	0,67	0,33	*1,00*
C kocka	0	**1,00**	*1,00*

Ha a kicsi és a nagy értékeket egy m osztópont segítségével definiáljuk (kicsi: $X < m$; nagy: $X > m$), akkor a nagy értékek valószínűsége az F eloszlásfüggvény segítségével egyszerűen kifejezhető:

$$P(\text{Kis érték}) = F(m) \text{ és } P(\text{Nagy érték}) = 1 - F(m).$$

Úgy látszik, hogy a három eloszlás körbeverő tulajdonsága maga után vonja, hogy mindhárom X_i ($i = 1, 2, 3$) eloszláshoz található olyan m_i osztópont, amelyre vonatkozóan az X_i változó F_i eloszlásfüggvényének $F_i(m_i)$ függvényértéke kisebb, vagyis P(Nagy érték) = $1 - F_i(m_i)$ nagyobb, mint a másik két eloszlás esetében. Hogy ez valóban így is van, azt a következő fejezetben mutatjuk meg.

A VALÓSZÍNŰSÉGI FÖLÉNY A MUTATÓJÁNAK EGY ÚJ ÉRTELMEZÉSI LEHETŐSÉGE

Az /1/ kifejezéssel definiált A_{ij} mutató *Brunner* és *Munzel* [2000] szerint felírható az alábbi alakban is:

$$A_{ij} = P(X_i > X_j) + 0{,}5 \cdot P(X_i = X_j) = \int F_j dF_i = E[F_j(X_i)], \qquad /4/$$

ahol F_i ($i = 1, ..., h$) az X_i változó normalizált eloszlásfüggvénye:

$$F_i(\mathrm{x}) = P(X_i < x) + 0{,}5P(X_i = x) \text{ minden } x\text{-re} \qquad /5/$$

(lásd *Ruymgaart* [1980]). Ha X diszkrét változó, melynek lehetséges értékei az x_1, x_2, ... számok, akkor /4/ így is felírható:

$$A_{ij} = \Sigma_k P(X = x_k) F_j(x_k).$$

Következésképpen /4/ felhasználásával $A_{ii} = 0{,}5$ miatt az X_i és az X_j változó sztochasztikus viszonyát meghatározó $A_{ij} - 0{,}5$ különbség az alábbi módon néz ki:

$$A_{ij} - 0{,}5 = A_{ij} - A_{ii} = \Sigma_k P(X = x_k)[F_j(x_k) - F_i(x_k)]. \qquad /6/$$

Folytonos esetben analóg módon kapjuk, hogy

$$A_{ij} - 0{,}5 = \int f_i(x)[F_j(x) - F_i(x)]dx, \qquad /7/$$

ahol f_i az X_i változó sűrűségfüggvénye.

Akármilyen eloszlású is az X_i és az X_j változó, ha $X_i >_{szt} X_j$, akkor $A_{ij} > 0{,}5$, ami

$$\Sigma_k P(X = x_k) = 1, \text{ illetve } \int f_i(x)dx = 1$$

miatt maga után vonja, hogy az $[F_j(x) - F_i(x)]$ különbségek súlyozott átlaga 0-nál nagyobb. Ez azt jelenti, hogy $X_i >_{szt} X_j$ esetén az F_j eloszlásfüggvény általában (átlagosan) nagyobb, mint F_i, és egyben létezik legalább egy olyan m érték, amelyre $F_j(m) > F_i(m)$. Megjegyezzük, hogy ugyanez nemcsak az /5/-tel definiált normalizált, hanem a valódi – balról folytonos – eloszlásfüggvényre is igaz.

Ennél többet kíván meg az eloszlások *Lehmann* ([1975] 66. old.) szerinti rendezettsége, amelyet erős sztochasztikus rendezésnek nevezünk. Eszerint valamely X és Y változóra X sztochasztikusan nagyobb mint Y, ha $F_Y(c) \geq F_X(c)$ minden c-re úgy, hogy legalább egy c-re $F_Y(c) > F_X(c)$. A Lehmann-i erős sztochasztikus rendezés nyilvánvalóan maga után vonja az általunk bevezetett „gyenge" sztochasztikus rendezést, sőt, még a várható értékek és a mediánok hasonló irányú viszonyát is. Ebből következik, hogy az erős sztochasztikus rendezettség tranzitív, ami kizárja a körbeverés lehetőségét. Sztochasztikus körbeverés természetesen akkor sem állhat fenn, ha teljesül az additivitási modell (például a varianciaanalízis modelljében), mely szerint az összehasonlított eloszlások legfeljebb egy eltolási paraméterben különböznek egymástól.

Végül is mit tudunk meg abból, hogy $A_{ij} > 0{,}5$, vagyis hogy X_i gyenge értelemben sztochasztikusan nagyobb X_j-nél? A /6/ és a /7/ összefüggés alapján azt, hogy létezik legalább egy olyan m érték, amelyre az m-nél nagyobb értékek az X_i változó eloszlásában nagyobb valószínűséggel fordulnak elő, mint az X_j változó eloszlásában, és ha ez az osztópont szakmailag releváns, jól értelmezhető szintet definiál – ilyen lehet például egy gyógyulási kritérium vagy valamilyen teljesítmény minimálisan megkövetelt szintje –, akkor A_{ij} becslése, illetve a vele kapcsolatos hipotézisek vizsgálata fontos feladatnak tekintendő.

Ha pszichológiai kezelések változói sztochasztikusan körbeverik egymást, akkor ez arra utal, hogy az egyik kezelés mondjuk a teljes gyógyulás valószínűségében nő a többi fölé, a másik esetleg abban, hogy a legnagyobb valószínűséggel biztosítja egy bizonyos gyógyulási szint elérését stb. A legjobb persze az lenne, ha mindig létezne olyan terápiás eljárás, amelynek a gyógyítási hatékonyságát mérő változója a Lehmann-i szigorúbb definíció szerint is sztochasztikusan nagyobb lenne bármely kandidáns terápiás eljárásénál. Ez azonban már olyan erős rendezés, amelyet még az átlagok vagy a mediánok szigorúan monoton rendezettsége sem garantál. Megjegyezzük, hogy a valamely X változó szerinti sztochasztikus körbeverés az X változó gyengeségeként is felfogható, és olyan jelzés, amely e változó egydimeziós jellegét vonja kétségbe.

Mivel a gyakorlati példák egy jelentős részében sem az additivitási modell, sem az erős sztochasztikus rendezés fennállására nem számíthatunk, két eloszlás összehasonlítása során szakmailag fontos lehet megkeresni azt a pontot, ahol a két eloszlás a legélesebbben különbözik egymástól, vagyis ahol az $[F_j(x) - F_i(x)]$ különbség a legnagyobb. Ezt a módszert több valódi példa segítségével a következő fejezetben részletezzük.

ELOSZLÁSOK RÉSZLETES ÖSSZEHASONLÍTÁSA

Mindenekelőtt keresnünk kell egy olyan statisztikai módszert, amelynek segítségével két független minta esetén megbízhatóan kideríthető, hogy van-e az X függő változó értékskálájának olyan pontja, amelyben a két elméleti eloszlás közti különbség koncentrálódik.

Ez a probléma hagyományos megközelítés szerint két részfeladatot foglal magában. Elsőként azt kell tisztázni, hogy elvethető-e a két eloszlás azonosságának nullhipotézise, majd pozitív eredmény esetén utóelemzéssel fel kell deríteni, hogy az értékskálának melyek azok a pontjai, ahol a két eloszlásfüggvény szignifikánsan különbözik egymástól.

E probléma megoldásához elsőként a Kolmogorov–Szmirnov-féle kétmintás próbát választottuk, melynek próbastatisztikája a két empirikus eloszlásfüggvény maximális különbségének egyszerű függvénye (lásd *Vincze* [1968] 158. old., vagy *Hollander–Wolfe* [1999] 178–186. old.). A próba azonban nem igazolta a vele kapcsolatos elvárásokat, mert ereje több empirikus elemzés során rendkívül alacsonynak mutatkozott. Számos olyan eset is előfordult, amikor a két eloszlás a Mann–Whitney-próba és annak néhány robusztus változata segítségével 1 százalékos szinten szignifikánsan különbözött, miközben a Kolmogorov-Szmirnov-próba még 10 százalékos szinten sem jelzett különbséget.

A Kolmogorov-Szmirnov-próba hagyományos alternatívájaként szóba jöhet még a χ^2-próba is, amelynél azonban a kis hatékonyság mellett még elemszámproblémák is jelentkeznek. Végül a két eloszlásfüggvény részletes összehasonlítására az alábbi módszert alkalmaztuk:

1. Egyesítve a két független mintát, osszuk a legkisebb és a legnagyobb adat közti tartományt igen sok (mondjuk 100) azonos szélességű érintkező osztályra, majd ezen osztályok közül hagyjuk el azokat, amelyekbe egyetlen adat sem esik.

2. A maradék osztályok felső határán határozzuk meg az egyesített minta empirikus eloszlásfüggvényének értékét, majd tartsunk meg ezek közül maximálisan k darabot, ahol k egy 5 és 10 közötti egész szám, amelyet az X változó értékkészletének számossága, illetve az összelemszám figyelembevételével a statisztikai összehasonlítás megkezdése előtt rögzítünk. A cél az, hogy a k számú eloszlásfüggvény-értékhez tartozó k osztóponttal minél egyenletesebben lefedjük az X-értéktartományt, vagyis hogy az egymást követő osztópontok közé eső adatok relatív gyakoriságai a lehető leghasonlóbbak legyenek. Jelölje az így kapott k osztópontot rendre $x_1, x_2, ..., x_k$.

3. Ezen x_i $(i = 1, 2, ..., k)$ osztópontok mindegyikének a segítségével hasonlítsuk össze a két független minta empirikus eloszlásfüggvényének értékét a

$$H_0\text{: } P(X < x_i \mid 1.\text{ minta}) = P(X < x_i \mid 2.\text{ minta})$$

hipotézis tesztelésével. A H_0 hipotézis vizsgálatára használhatjuk alkalmazási feltételének teljesülése esetén a 2·2-es χ^2-próbát, ellenkező esetben pedig a Fisher–Irwin-féle egzakt-próbát (*Fleiss–Levin–Paik* [2003] 56. old.).

4. Minthogy a két empirikus eloszlásfüggvény összehasonlítását egyidejűleg k osztópontban végezzük el k próbát végrehajtva, a próba szintjének biztosításához a Bonferroni-féle elvet alkalmazzuk, vagyis egy-egy osztópont esetén akkor utasítjuk el a H_0 hipotézist α szignifikanciszinten, ha a 2·2-es χ^2-próba, illetve a Fisher–Irwin-próba α/k szinten szignifikáns.

Kettőnél több (h számú) független minta empirikus eloszlásainak összehasonlítását a fentebb leírttal azonos módon végezhetjük el. Az egyetlen különbség az, hogy itt a pontonkénti összehasonlítást $2 \cdot h$ nagyságú gyakorisági táblák alapján az általános χ^2-próba segítségével hajtjuk végre (*Hajtman* [1968] 299. old.).

Ha k értékét alacsonyra állítjuk be, akkor az eloszlásokat kevesebb ponton hasonlíthatjuk össze, de az eredmény könnyebben lesz szignifikáns. Nagyobb k érték esetén részletesebb összehasonlítást kapunk, de a pontonkénti próbák nehezebben lesznek szignifikánsak. Ha az X változónak sok különböző értéke van, és az a feltételezésünk, hogy a minták csak 1-2 pontban különböznek markánsan, akkor célszerűbb nagy k-t választani, hogy biztosabban rátaláljunk ezekre az osztópontokra. Ez esetben a nagy k érték megvéd egyben a könnyű szignifikanciától is. Ha az X változó erősen diszkrét (például 3-5-értékű skálaváltozó), akkor k beállítása nem hat az eredményre, mert az értéktartomány felosztása során sosem kaphatunk az X változó különböző értékeinek számánál több nem üres osztályt.

A fentebb részletezett módszert beépítettem a Ministat programcsomag (lásd *Vargha–Czigler* [1999], illetve *Vargha* [2000] A. melléklet) legújabb, 3.3. verziójának nemparaméteres csoportösszehasonlító rutinjába (lásd a 5–9. táblákban bemutatott példákat). Ezzel nemcsak az eloszlások összehasonlítására nyertünk egy magas relatív erejű eljárást, hanem egyben olyan eljáráshoz jutottunk, amelynek segítségével egyidejűleg kettő vagy kettőnél több eloszlást az értéktartomány több pontján is összevethetünk az α szint megtartása mellett.

Az alábbiakban konkrét empirikus elemzések segítségével illusztráljuk a fenti elméleti fejtegetés gyakorlati relevanciáját.

Pszichológia szakra jelentkezők feminitása

1981-ben az ELTE pszichológia szakára jelentkezők közül 94 fő vett részt egy ún. előzetes alkalmassági vizsgálaton, közöttük 16 férfi és 78 nő. Ezek közül 12 férfi és 70 női személlyel az S-CPI-t, a Kaliforniai Személyiség Kérdőív 300 kérdéses rövidített magyar változatát is felvették. E személyiségteszt egyik skálája a „Feminitás” (Fem), mely tájékoztat a vizsgált személy érdeklődésének feminin vagy maszkulin jellegéről. A skála magas pontértéke mindkét nem esetében inkább nőies érdeklődést, a nőkre jellemző viselkedésformák preferálását jelzi. Az alacsony értékekből erőteljes viselkedésre és a férfias viselkedési formák előnyben részesítésére következtethetünk (*Oláh* [1985]).

E skála érvényességét tesztelendő, összehasonlítottuk a fenti mintában a férfiak és a nők Fem-adatait. Az elméleti átlagok egyenlőségét (az adott mintában a fiúk átlaga 12,08, a lányoké 14,00 volt) a kétmintás t-próba ($t(80) = 2{,}954$, $p < 0{,}01$) és a Welch-féle d-próba ($d(13) = 2{,}372$, $p < 0{,}05$) segítségével teszteltük, a sztochasztikus egyenlőséget pedig a Mann–Whitney-próba ($z = 2{,}339$, $p < 0{,}05$) és a Brunner–Munzel-próba (BM(12)= = 2,108, $p < 0{,}10$) segítségével (e próbákkal kapcsolatban lásd *Vargha* [2000]). A nők férfiakkal szembeni sztochasztikus dominanciájának jellemzésére kiszámított *A* valószínűségi mutató mintabeli becsült értéke: *A*(nő, férfi) = 0,71 lett, ami arról tájékoztat, hogy egy véletlenszerűen kiválasztott – pszichológia szakra felvételiző – férfi és nő esetében körülbelül 71 százalék az esélye, hogy a nő Fem-értéke nagyobb a férfiénál (az egyenlőség valószínűségét igazságosan felezzük).

A két empirikus eloszlás részletesebb összehasonlításához megvizsgáltuk az eloszlásfüggvények különbségét a Ministat programcsomag segítségével. A program alkalmas osztópontok segítségével igen sok szűk kategóriára osztja a vizsgált változó értékskáláját úgy, hogy 100-nál kevesebb különböző érték esetén minden érték külön osztályba kerül. A program mindezen osztályok felső határaira kiszámítja az empirikus eloszlásfüggvény értékét az egyesített, illetve a két összehasonlított mintában, de ezek különbségének szignifikanciáját csak *k* számú pontban teszteli. Tekintve, hogy a jelen példa esetében az összelemszám viszonylag alacsony ($n = 82$), *k* értékére 5-öt állítottam be. A program az empirikus eloszlásfüggvények értéke, illetve azok különbsége mellett pontonként kiszámítja a φ kontingencia-együttható értékét is (lásd *Vargha* [2000] 444–445. old.), mely arról tájékoztat, hogy az adott osztópont segítségével dichotomizált függő változó (jelen esetben a Fem) milyen szoros kapcsolatban van az ugyancsak kétértékű csoportosító változóval (jelen esetben a személy nemével). Ennek az elemzésnek az eredménye röviden az 5. táblában látható.

5. tábla

Pszichológia szakra felvételiző férfiak (1) és nők (2) Fem-eloszlásának részletes összehasonlítása a Ministat programcsomag segítségével (n = 82)

Eloszlásfüggvények összehasonlítása

x	F(x)	F1(x)	F2(x)	F1(x)-F2(x)	Phi	p	Szign.
8.050	0.012	0.083	0.000	0.083	0.27		
10.050	0.085	0.417	0.029	0.388	0.49	0.003	**
11.050	0.159	0.500	0.100	0.400	0.39		
12.050	0.280	0.583	0.229	0.355	0.28	0.157	
13.050	0.451	0.667	0.414	0.252	0.18	0.630	
14.050	0.646	0.750	0.629	0.121	0.09	1.000	
15.050	0.768	0.833	0.757	0.076	0.06		
16.050	0.915	1.000	0.900	0.100	0.13	1.000	
17.050	0.951	1.000	0.943	0.057	0.09		
18.050	1.000	1.000	1.000				

A két elméleti eloszlás egyenlőségének tesztelése:
Kolmogorov-Szmirnov próba: J* = 1.280 (p = 0.075)

Az 5. tábla adatai szerint a két nem között akkor kapjuk a leginkább szignifikáns különbséget, ha a Fem-skálát az $m = 10,05$ osztópont segítségével dichotomizáljuk. Ezen érték alatti (azaz 0-10 közti) pontot ért el az adott mintában a férfiak 41,7, illetve a nők 2,9 százaléka. A különbség mintegy 39 százalékpont, és ez – a Fisher–Irwin-próbával – erősen szignifikáns ($p = 0,003$). A táblában feltüntetett *p*-érték már figyelembe veszi, hogy az eloszlások különbségét egyidejűleg 5 pontban teszteljük. Enélkül a Fisher–Irwin-próba *p*-értéke 0,0005 lenne. Érdemes megfigyelni, hogy a Kolmogorov–Szmirnov-próba csak tendencia szinten jelez ($p < 0,10$).

Ezt az eredményt szakmailag a következőképpen interpretálhatjuk. A férfiak és a nők feminitása leginkább abban tér el egymástól, hogy létezik egy olyan minimális feminitásszint (a Fem-skálán ez a 10 és a 11 pont között van), amely alatti értéket döntő többségben csak férfiak produkálnak. A nők közül tehát szinte mindenki (mintánkban 70 közül 68) rendelkezik egy minimális feminitásszinttel. Ilyen jellegű eltérést a Fem-skála magasabb régiójában nem tapasztalunk. Például nincs egy olyan magas feminitásszint (elvileg nyugodtan létezhetne), melynél nagyobbat jobbára csak nők érnek el.

További összehasonlítások az S-CPI személyiségteszt skálái segítségével

Az 1. pontban bemutatott statisztikai elemzés szakmailag igen érdekes eredménye nem fogadható el minden fenntartás nélkül, mert

– a minta (pszichológia szakra felvételizők) meglehetősen speciális;
– a mintanagyság (n = 82) viszonylag kicsi;
– a férfi-nő arány (15 százalék, illetve 85 százalék) túlságosan extrém.

Emiatt ugyanezt az elemzést egy nagyobb mintában is elvégeztük. A minta 331 kábítószerezés miatt orvosi kezelés alatt álló, 17–48 éves személyből állt, akiket 97 hasonló korú és iskolázottságú kontroll személy egészített ki (*Demetrovics* [2005])[1].

Ebben a 428 fős mintában 401 személy (249 férfi és 152 nő) rendelkezett érvényes S-CPI adatokkal. Az itt a Fem-skálával végzett, az 1. pontban leírtakkal azonos elemzések eredményét (a nagy elemszám miatt most k = 10 beállításával) a 6. tábla tartalmazza.

6. tábla

Férfiak (1) és nők (2) Fem-eloszlásának részletes összehasonlítása a Ministat programcsomag segítségével (n = 401)

x	F(x)	F1(x)	F2(x)	F1(x)-F2(x)	Phi	p	Szign.
2.075	0.002	0.004	0.000	0.004	0.04		
4.025	0.007	0.012	0.000	0.012	0.07		
5.075	0.020	0.032	0.000	0.032	0.11		
6.125	0.057	0.088	0.007	0.082	0.17		
7.025	0.127	0.189	0.026	0.162	0.24	0.000	***
8.075	0.222	0.325	0.053	0.273	0.32	0.000	***
9.125	0.327	0.458	0.112	0.346	0.36	0.000	***
10.025	0.491	0.679	0.184	0.495	0.48	0.000	***
11.075	0.594	0.775	0.296	0.479	0.47	0.000	***
12.125	0.728	0.876	0.487	0.389	0.42	0.000	***
13.025	0.830	0.932	0.664	0.267	0.35	0.000	***
14.075	0.928	0.980	0.842	0.138	0.26	0.000	***
15.125	0.963	0.992	0.914	0.077	0.20		
16.025	0.998	1.000	0.993	0.007	0.06		
17.075	1.000	1.000	1.000				

A két elméleti eloszlás egyenlőségének tesztelése:
Kolmogorov-Szmirnov próba: J* = 4.804 (p = 0.000)

[1] Ez úton szeretnék köszönetet mondani *Demetrovics Zsolt* kollégámnak az általa rendelkezésre bocsátott adatokért.

A 6. táblából kiolvasható eredmények összhangban vannak az 5. táblában láthatóakkal. Meggyőzőnek tűnik, hogy a két nem Fem-skála szerinti összehasonlításában egy 10-et kis mértékben meghaladó m = 10+ε érték segítségével végzett dichotomizálással kapjuk a két nem között a legélesebb eltérést. Ez esetben a férfiak 68, a nőknek viszont csak a 18 százaléka ad m-nél kisebb Fem-értéket, és ettől távolodva a két nem közti eltérés egyre kevésbé kifejezett, bár a különbség végig erősen szignifikáns. Megjegyezzük, hogy a program k = 10 beállítása ellenére csak 8 ponton tesztelte az eloszlásfüggvények különbségét. Ennek az az oka, hogy a program az eloszlások extrém szélein ($F(x)$ < 0,06, illetve $F(x)$ > 0,94 esetén) nem hajt végre statisztikai próbát).

Az a jelenség, miszerint két eloszlás az értékskála egy bizonyos pontján a többinél kiemelkedően nagyobb mértékben különbözik egymástól, más tesztmutatókkal kapcsolatban esetenként sokkal markánsabban jelenik meg (lásd a 7. és a 8. táblát), de olyan is előfordul, hogy az értékskálán egynél több olyan pont is található, ahol a két eloszlás közti eltérés ugrásszerűen megnő (lásd a 9. táblát). Ezen eloszlásokat megszemlélve azt is észrevehetjük, hogy nagyobb eséllyel sérül a Lehmann-féle erős sztochasztikus rendezés olyan esetekben, amikor a két eloszlás csak néhány speciális pontban különbözik szignifikánsan egymástól (lásd a 7. és a 8. táblát).

Szakmai szempontból azonban éppen ezek az esetek az igazán érdekesek, amikor is jól látható módon sérül a változók értékskálájának folytonos jellege. Olyasfajta jelenséggel állhatunk szemben, mint mondjuk a testhőmérséklet, amelynek skáláján bizonyos értékhatárok (például a hőemelkedés 37 fokos, vagy a láz 37,5 fokos küszöbe) különleges jelentőségűek, és jól látható módon sértik a skála folytonos kvantitatív jellegét.

7. tábla

Férfiak (1) és nők (2) „Státus elérésre való képesség" (CS)[2] eloszlásának részletes összehasonlítása a Ministat programcsomag segítségével (n = 401)

x	F(x)	F1(x)	F2(x)	F1(x)-F2(x)	Phi	p	Szign.
2.075	0.002	0.004	0.000	0.004	0.04		
3.125	0.010	0.008	0.013	-0.005	-0.03		
4.025	0.022	0.028	0.013	0.015	0.05		
5.075	0.045	0.040	0.053	-0.012	-0.03		
6.125	0.092	0.072	0.125	-0.053	-0.09	0.269	
7.025	0.185	0.133	0.270	-0.137	-0.17	0.002	**
8.075	0.287	0.253	0.342	-0.089	-0.10	0.195	
9.125	0.416	0.378	0.480	-0.103	-0.10	0.150	
10.025	0.571	0.522	0.651	-0.129	-0.13	0.039	*
11.075	0.736	0.699	0.796	-0.097	-0.11	0.112	
12.125	0.843	0.831	0.862	-0.031	-0.04	1.000	
13.025	0.938	0.932	0.947	-0.016	-0.03		
14.075	0.980	0.980	0.980	-0.000	-0.00		
15.125	0.993	0.992	0.993	-0.001	-0.01		
16.025	0.995	0.996	0.993	0.003	0.02		
17.075	1.000	1.000	1.000				

[2] A CS-skála azt próbálja megállapítani, hogy a személy rendelkezik-e azokkal a személyiségadottságokkal, tulajdonságokkal, amelyek alapul szolgálnak ahhoz, hogy szociális közösségekben vezető pozícióra tegyen szert. Magas pontérték esetén a kiemelkedni, fejlődni, előrehaladni akarás motivációs alapjait, illetve igényét azonosíthatjuk (*Oláh* [1985]).

8. tábla

Férfiak (1) és nők (2) „Szocializáltság" (SO)[3] *eloszlásának részletes összehasonlítása a Ministat programcsomag segítségével (n = 401)*

x	F(x)	F1(x)	F2(x)	F1(x)-F2(x)	Phi	p	Szign.
8.090	0.020	0.020	0.020	0.000	0.00		
9.190	0.042	0.040	0.046	-0.006	-0.01		
10.070	0.082	0.076	0.092	-0.016	-0.03	1.000	
11.170	0.147	0.149	0.145	0.004	0.01	1.000	
12.050	0.224	0.221	0.230	-0.009	-0.01	1.000	
13.150	0.304	0.301	0.309	-0.008	-0.01	1.000	
14.030	0.384	0.410	0.342	0.068	0.07	0.887	
15.130	0.494	0.522	0.447	0.075	0.07	0.733	
16.010	0.576	0.606	0.526	0.080	0.08	0.576	
17.110	0.636	0.675	0.572	0.102	0.10		
18.210	0.698	0.723	0.658	0.065	0.07	0.845	
19.090	0.796	0.831	0.737	0.094	0.11	0.114	
20.190	0.843	0.884	0.776	0.107	0.14		
21.070	0.893	0.928	0.836	0.092	0.14	0.019	*
22.170	0.933	0.952	0.901	0.050	0.10		
23.050	0.940	0.956	0.914	0.041	0.08		
24.150	0.973	0.976	0.967	0.009	0.03		
25.030	0.985	0.988	0.980	0.008	0.03		
26.130	0.993	0.992	0.993	-0.001	-0.01		
27.010	0.995	0.996	0.993	0.003	0.02		
28.110	1.000	1.000	1.000				

9. tábla

Férfiak (1) és nők (2) „Jó közérzet" (WB)[4] *eloszlásának részletes összehasonlítása a Ministat programcsomag segítségével (n = 401)*

Eloszlásfüggvények összehasonlítása

c	F(c)	F1(c)	F2(c)	F1(c)-F2(c)	Phi	p	Szign.
4.125	0.012	0.012	0.013	-0.001	-0.00		
6.125	0.015	0.012	0.020	-0.008	-0.03		
7.125	0.027	0.020	0.039	-0.019	-0.06		
8.125	0.045	0.040	0.053	-0.012	-0.03		
9.125	0.055	0.040	0.079	-0.039	-0.08		
10.125	0.080	0.056	0.118	-0.062	-0.11	0.129	
11.125	0.115	0.072	0.184	-0.112	-0.17		
12.125	0.147	0.096	0.230	-0.134	-0.18	0.001	***
13.125	0.197	0.157	0.263	-0.107	-0.13		
14.125	0.242	0.189	0.329	-0.140	-0.16	0.007	**
15.125	0.304	0.245	0.401	-0.156	-0.16		
16.125	0.379	0.317	0.480	-0.163	-0.16	0.005	**
17.125	0.464	0.386	0.592	-0.207	-0.20	0.000	***
18.125	0.541	0.462	0.671	-0.209	-0.20	0.000	***
19.125	0.631	0.570	0.730	-0.160	-0.16	0.006	**
20.125	0.718	0.671	0.796	-0.125	-0.14	0.034	*
21.125	0.808	0.775	0.862	-0.087	-0.11	0.162	
22.125	0.888	0.867	0.921	-0.054	-0.08	0.495	
23.125	0.938	0.928	0.954	-0.026	-0.05		
24.125	0.965	0.956	0.980	-0.024	-0.06		
25.125	0.998	0.996	1.000	-0.004	-0.04		
27.125	1.000	1.000	1.000				

[3] Az SO-skála a felettes én funkciók működésének hatékonyságát, a szociális érettség és szociális felelősségérzet mértékét állapítja meg (*Oláh* [1985]).

[4] A WB-skála célja azonosítani azokat a személyeket, akik minimalizálják aggodalmaikat, panaszaikat, magas szinten elaborálják pszichés feszültségeiket, viszonylagosan mentesek az önmagukban való kételkedéstől és elégedettek elért eredményeikkel (*Oláh* [1985]).

CSOPORTDISZKRIMINÁCIÓ NÖVELÉSE BINARIZÁLÁS SEGÍTSÉGÉVEL

Az előző fejezetben elvégzett összehasonlítások alapján logikusnak tűnik, hogy ha az eloszlásokat legjobban diszkrimináló pontokban dichotomizáljuk a vizsgálatba bevont függő változók értékskáláját, akkor ezekkel az újonnan képzett bináris változókkal esetenként statisztikailag előnyösebb eredményekre juthatunk, mint az eredeti változóegyüttes segítségével. E hipotézis ellenőrzésére a fenti S-CPI vizsgálatnak mind a 21 skálájával összehasonlítottuk a két nemet, majd kiemeltünk 9 olyan skálát, amelyek esetében az összehasonlított két empirikus eloszlás legalább egy pontban 5 százalékos szinten szignifikánsan különbözött egymástól.

Az eredeti és a bináris változók diszkriminációs erejének összehasonlítása céljából az SPSS programcsomaggal lépésenkénti diszkriminancia-analízist (DA) hajtottunk végre a nem bejóslására (lásd *Székhelyi–Barna* [2003] 8. fejezet). A személy nemének pszichológiai teszt segítségével történő predikciója nem tűnik gyakorlati szempontból releváns problémának, mert ritkán fordul elő, hogy a vizsgált személy neme ismeretlen. Elméletileg azonban a két nem között talált minden különbség hasznos információval szolgálhat a két nem személyiségének eltérő működésmódját illetően.

A DA a jelen esetben következő eredményekre vezetett:

a) A 21 eredeti S-CPI skála segítségével végzett elemzés során a DA három szignifikáns hatású változót emelt ki (Feminitás, Jó közérzet és Felelősségtudat) és átlagosan 75,6 százalékos megbízhatósággal tudta azonosítani a két nemi csoportot.

b) Ugyanakkor a 9 bináris skála segítségével végzett elemzés során a DA 6 szignifikáns hatású változót emelt ki (köztük a fenti 3 skála bináris formáját), amelyek segítségével átlagosan 74,6 százalékos megbízhatósággal lehetett azonosítani a személy nemét.

Már az is meglepő, hogy a binarizálással nem csökken számottevően a predikció hatékonysága, ami erősíti azt a feltételezést, hogy számos S-CPI skála esetében a két nem közti különbségben nem az eloszlások szintkülönbségei játsszák a fő szerepet, hanem az értékskálák bizonyos kritikus pontjai. Tekintettel azonban arra, hogy bináris változók esetén a független változók folytonosságát és normalitását feltételező DA nem a legadekvátabb osztályozási eljárás, a 9 binarizált változóval elvégeztünk egy lépésenkénti algoritmusú bináris logisztikus regresszió elemzést (LRA – lásd *Székhelyi–Barna* [2003] 9. fejezet) is, melynek során a két nemet már átlagosan 79,3 százalékos megbízhatósággal lehetett azonosítani öt szignifikáns hatású változó (Feminitás, Énerő, Felelősségtudat, Jó benyomás keltés és Jó közérzet) segítségével. Megjegyezzük, hogy az LRA-t az eredeti 21 skálával mint független változóval elvégezve, a DA-hoz hasonló eredményességű, 75,3 százalékos átlagos helyes azonosítású modellt kaptunk három magyarázó változó (Feminitás, Jó közérzet és Tolerancia) segítségével. Ugyanakkor a legjobb diszkriminatív hatékonyságú eredeti és bináris változókat egy csapatba összevonva az LRA-ban nem sikerült a helyes azonosítás százalékát 79,3 fölé vinni.

Mentálisan beteg és egészséges nők diszkriminációja pszichiátriai skálák segítségével

A két nem elméletileg tanulságos összehasonlító elemzése után kerestünk egy olyan adatbázist, ahol valamely kétértékű változó bejóslása igazi gyakorlati relevanciával bír. Ezt az alábbi kutatás adatállománya biztosította.

Pethő Bertalan és munkatársai 1967 és 1974 között 237 pszichésen súlyosan beteg nőt vontak be egy komplex követéses vizsgálatba (*Pethő* [2001]). E betegeket egy 54 fős, mentálisan egészséges személyekből álló kontroll minta egészítette ki. Ezek közül a személyek közül 271 (230 beteg és 41 egészséges) személy esetében rendelkezésre álltak az *Overall* [1968] által kialakított Overall-féle Faktor-szerkezet Becslésskála (Factor Construct Rating Scale – FCRS), valamint a *Rockland* és *Pollin* [1965] nevéhez fűződő Rockland-Pollin Becslésskála (RPS) tételei (e tesztekről bővebben lásd *Pethő* [1972], illetve *Pethő–Szilágyi–Hajtman* [1977]).

Az FCRS-nek 17 elemi tételét (F1, ..., F17) az RPS-nek pedig 34 elemi tételét (R1, ..., R33, R35) vontuk be az elemzésekbe. Ezek a tételek olyan skálák, amelyeken a 0 érték valamely pszichiátriai tünet teljes hiányát, a maximumhoz közeli értékek pedig ezen tünetek markáns jelenlétét jelzik.

A következő statisztikai elemzéseket végeztük el:

a) Változónként teszteltük a két fő diagnosztikai csoport (beteg versus egészséges) sztochasztikus egyenlőségét, és összehasonlítottuk az eloszlásfüggvényeket az előző pontban leírt módon (most is k = 10 beállításával). Ezután a két csoportot szignifikánsan elkülönítő teszttételeket binarizáltuk a két eloszlást legjobban elkülönítő osztópontokban. Így összesen 40 bináris változóhoz jutottunk (F1–F14, F16, F17, R1–R5, R9, R11–R16, R18, R20–R23, R25–R30, R33).

b) Ezután egyrészt az 51 eredeti változóval, másrészt a 40 bináris változóval megpróbáltuk a kétértékű diagnózis (beteg versus egészséges) dichotóm függő változóját a lehető legjobban bejósolni. Ehhez az előző pontban már bemutatott DA és LRA-módszerét használtuk lépésenkénti változó kiválasztással (Forward stepwise módszer). A kiválasztást addig folytattuk, amíg az újonnan beválasztott változó szignifikánsan növelte a függő változó predikcióját. Az elemzések során kapott helyes besorolási arányokat a 10. tábla mutatja be.

10. tábla

A helyes azonosítás arányai a DA és az LRA segítségével történő predikció során a két változócsoportra (eredeti versus binarizált)
(százalék)

Diagnózis	DA az eredeti változókkal	DA a binarizált változókkal	LRA az eredeti változókkal	LRA a binarizált változókkal
Beteg (n = 230)	78,7	87,0	95,7	95,7
Kontroll (n = 41)	92,7	100,0	82,9	85,4
Összesen	80,8	88,9	93,7	94,1
Kiválasztott változók száma (darab)	11	7	10	8

A 10. tábla alapján az alábbi konklúziók vonhatók le:

a) Nem normális eloszlású változók esetén a DA esetenként számottevően gyengébb diszkriminációra képes, mint az LRA.

b) Ha a dichotóm függő változó két értékét maximálisan diszkrimináló skálapontokban binarizáljuk a szignifikáns prediktív erővel rendelkező független változókat, akkor ezzel a DA diszkriminációs hatékonyságát esetenként jelentősen megnövelhetjük. Például jelen adataink esetében 7 bináris változóval 88,9 százalékos helyes azonosítást értünk el, míg 11 eredeti változóval csak 80,8 százalékos azonosítási százalékot lehetett elérni.

c) Ez az előny, ha esetenként kisebb mértékű is, a logisztikus regresszióban is megmarad (8 bináris változóval 94,1 százalék, míg 10 eredeti változóval 93,7 százalék).

KORRELÁCIÓS KAPCSOLATOK ERŐSÍTÉSE ALKALMAS SKÁLAREDUKCIÓ SEGÍTSÉGÉVEL

A társadalomtudományi kutatások statisztikai feldolgozásaiban a csoportok összehasonlítása mellett talán a korrelációs elemzések örvendenek a legnagyobb népszerűségnek. Célszerűnek látszik ezért megnézni, hogy a fentebb bemutatott binarizálási módszerrel nem lehetne-e a változók közti kapcsolatokat esetenként markánsabban kimutatni. E kérdés tisztázására az alábbi empirikus statisztikai elemzést végeztük el.

A szakmai kérdés pszichológiai jellegű, és arra vonatkozik, hogy a Rorschach tesztvizsgálat során a vizsgált személy által adott válaszok tartalmi megoszlása összefügg-e a személy iskolázottsági szintjével. A Rorschach-teszt a klinikai pszichológia egyik legfontosabb diagnosztikai eljárása. A vizsgált személynek tíz táblát mutatnak be, amelyeken tintapacákra emlékeztető foltok vannak. A feladat: jelentést adni ezeknek a foltoknak. A vizsgált személy által adott válaszokat a a teszt elvégzése után összesítik, és különböző szempontok (például az értelmezett folt nagysága, tagoltsága vagy az adott válasz tartalmi kategóriája) szerint minősítik. Ha például a vizsgált személy valamelyik foltban vagy annak részletében egy kígyót vél felfedezni, akkor a válasz az „Állat” tartalmi minősítést kapja, ha pedig az Eiffel-tornyot, akkor az „Architektúra” tartalmi besorolást. Ezekből a válaszokból fontos pszichológiai következtetések vonhatók le.

A jelen elemzésben felhasznált adatok a Magyar Rorschach Standard kialakításánál felhasznált 359 pszichésen egészséges személytől származnak (lásd Vargha [1989a], [1989b]). Az itteni statisztikai elemzésbe a háromértékű iskolázottságot (alsó-, közép- és felsőfokú végzettség) és 44 tartalmi kategória (M = Ember, T = Állat, Myth = Mitológia, Anat = Anatómia, Pfl = Növény, Obj = Tárgy, Szikla, Táj, Tűz, Víz, Robbanás stb.) egész számra kerekített százalékos előfordulási arányát vontam be. Például Anat% = 7,4 arról tájékoztat, hogy a vizsgált személy válaszainak 7,4 százalékában fordult elő az Anatómia tartalmi körbe eső válasz vagy válaszrészlet.

Első lépésben az alsó-, a közép- és a felsőfokú végzettségűek három független mintájának sztochasztikus összehasonlítását végeztem el (*Vargha* [2002]), kiegészítve ezt az eloszlások részletes összehasonlításával (most is k = 10 beállításával). Például a Röntgen% (Rtg) változóval kapcsolatos főbb eredményeket a 11. tábla mutatja be.

A 11. táblából kiolvasható, hogy a három eloszlás között a legélesebb eltérés mindjárt a legkisebb skálaértéknél (x = 0,120) tapasztalható. Ennél kisebb érték csak a 0 volt, ami azt jelenti, hogy a három iskolázottsági csoportból a személyek rendre 85,4, 66,9, illetve 58,8 százaléka esetében Rtg = 0, vagyis ennyien voltak azok, akik a vizsgálat során egyetlen Röntgen-választ sem adtak. Ez a három arány a χ^2-próba szerint szignifikánsan különbözik egymástól (χ^2 = 19,03, p = 0,001). A 0+ε érték (például 0,1) tehát alkalmas osztópontnak tűnik az Rtg változó skáláján a három iskolázottsági csoport elkülönítésére. Ráadásul az ezen érték segítségével képzett bináris változó szakmailag jól értelmezhető, ez ugyanis a Röntgen tartalom indikátorváltozója. Akik adtak ilyen választ a Rorschach-vizsgálat során, azon személyeknél e bináris változó értéke 1, a többi személynél pedig 0.

A 44 megvizsgált tartalmi változó közül 30 tudta szignifikánsan elkülöníteni az iskolázottság egyes szintjeit egymástól. Közülük 29 esetében a 0+ε érték jól diszkrimináló pont volt az értékskálán. Egyetlen változó (T = Állat%) esetében az eloszlások az értékskálának szinte a teljes vertikumában markánsan különböztek egymástól, vagyis a különbség nem összpontosult egy vagy több jól meghatározható skálapontra. Ennek megfelelően a 30 szignifikáns diszkrimináló képességű tartalmi változóból egyet (T-t) meghagytam eredeti formájában, a többit pedig 0+ε osztópont segítségével binarizáltam. E transzformációk hatását a korrelációs kapcsolatokra az alábbi statisztikai elemzésekkel vizsgáltuk meg.

11. tábla

Az iskolai végzettség három szintjének sztochasztikus összehasonlítása
a Rorschach-teszt Röntgen% tartalmi mutatójának segítségével

```
Függő változó: Rtg
Csoportosító változó: Isk
  Csoport      Érvényes                                   Sztoch. dominancia
Index  Név      esetek   Rangátlag   Rangszórás    = súly.   kül.súly.
----------------------------------------------------------------------
  1.   7-11      103       151.96        65.11      0.426**    0.422**
  2.   12-15     142       187.78        90.69      0.526      0.522
  3.   16-18     114       195.64        86.30      0.548*     0.544*
----------------------------------------------------------------------
Sztochasztikus homogenitás tesztelése
    Hagyományos eljárás, amely feltételezi a szóráshomogenitást:
    - Kruskal-Wallis-próba: H(2) = 16.483**
    Szóráshomogenitást nem igénylő robusztus közelítő eljárás:
    - Korrigált rang Welch-próba: rW3(2; 225) = 11.082**
    KULLE-féle aszimptotikusan egzakt próbák
    - Populációk azonos súlyozása: KG2(1.96; 356) = 9.915**
    - Mintaelemszámmal arányos súlyozás: KF2(1.96; 356) = 9.869**

Eloszlásfüggvények a különbségek előjelével és az eltérés szignifikanciájával
      x     F1(x)  F2(x)  F3(x) 12  13  23    Khi2    p    Szign.
----------------------------------------------------------------
     0.120  0.854  0.669  0.588 ++  +++ +   19.03  0.001   ***
     1.080  0.854  0.683  0.596 ++  +++ +   17.82  0.001   ***
     2.040  0.874  0.725  0.728 ++  +   .    8.96  0.079   +
     3.000  0.883  0.754  0.807 +   +   -    6.50  0.272
     4.200  0.942  0.803  0.868 ++  +   -    9.81  0.052   +
     5.160  0.971  0.859  0.904 ++  +   .    8.70  0.091   +
     6.120  0.971  0.880  0.921 +   .   .    6.61  0.257
     7.080  0.971  0.901  0.930 +   .   .    4.45
     8.040  0.981  0.915  0.956 +   .   .    5.32
     9.000  0.981  0.923  0.965 +   .   .    5.02
    10.200  0.981  0.937  0.982 .   .   .    4.97
    12.120  0.990  0.937  0.982 +   .   .    6.63
    13.080  0.990  0.944  0.982 .   .   .    5.33
    14.040  0.990  0.965  0.982 .   .   .    1.95
    15.000  1.000  0.972  0.991 .   .   .    3.77
    17.160  1.000  0.993  1.000 .   .   .    1.53
    24.120  1.000  1.000  1.000 .   .   .    0.00
 ----------------------------------------------
```

A legegyszerűbb ellenőrzési mód, hogy korreláltatjuk a iskolázottságot (a végzett osztályok számával mérve) a különböző tartalmi változók eredeti és binarizált változatával. Tekintve, hogy az iskolázottság csak az ordinalitás kritériumának tesz eleget, vagyis nem igazi kvantitatív változó, a szokásos Pearson-féle *r* mellett esetünkben célszerű rangkorrelációt is számolni. E célból a változók diszkrét jellegét is figyelembe vevő Kendall-féle tau-b monotonitási mérőszámot választottuk (*Dixon* [1990] 556. old.). Az eredmények a 12. és a 13. táblában láthatók.

12. tábla

29 binarizált változó Pearson-féle r korrelációja a végzett osztályok számával eredeti és transzformált alakban (n = 359)

Tartalmi kategória	Eredeti változóalak	Binarizált változóalak	Tartalmi kategória	Eredeti változóalak	Binarizált változóalak
Emberszerű	0,147***	0,229***	Füst	0,094*	0,161***
Mythológia	0,170***	0,284***	Tűz	0,163***	0,171***
Szörny	0,145***	0,183***	Víz	0,128**	0,215***
Anatómia	–0,024	0,225***	Jég	0,049	0,152***
Rtg	0,100*	0,211***	Explózió	0,108**	0,219***
Obj	0,280***	0,304***	Térkép	0,093*	0,194***
Matéria	0,132**	0,248***	Ornamentika	0,125**	0,234***
Jármű	0,135***	0,229***	Festmény	0,042	0,189***
Architektúra	0,209***	0,262***	Illusztráció	0,154***	0,281***
Ruha	0,259***	0,312***	Szobor	0,124**	0,236***
Táj	0,197***	0,249***	Ennivaló	0,049	0,183***
Pfl	–0,040	0,166***	Szem	0,165***	0,250***
Asztronómia	0,043	0,145***	Tőr	0,188***	0,190***
Sacrum	0,093*	0,164***	Barlang	0,089*	0,167***
Felhő	0,064	0,148***			

* $p < 0,1$
** $p < 0,05$
*** $p < 0,01$

13. tábla

29 binarizált változó Kendall-féle tau-b rangkorrelációja a végzett osztályok számával eredeti és transzformált alakban (n = 359)

Tartalmi kategória	Eredeti változóalak	Binarizált változóalak	Tartalmi kategória	Eredeti változóalak	Binarizált változóalak
Emberszerű	0,165***	0,196***	Füst	0,133***	0,141***
Mythológia	0,222***	0,247***	Tűz	0,141***	0,143***
Szörny	0,163***	0,175***	Víz	0,132***	0,189***
Anatómia	0,080**	0,216***	Jég	0,106**	0,135***
Rtg	0,151***	0,186***	Explózió	0,176***	0,209***
Obj	0,235***	0,271***	Térkép	0,142***	0,179***
Matéria	0,176***	0,210***	Ornamentika	0,173***	0,202***
Jármű	0,171***	0,198***	Festmény	0,126***	0,156***
Architektúra	0,202***	0,218***	Illusztráció	0,221***	0,247***
Ruha	0,231***	0,280***	Szobor	0,177***	0,200***
Táj	0,195***	0,215***	Ennivaló	0,136***	0,168***
Pfl	0,005	0,146***	Szem	0,206***	0,234***
Asztronómia	0,098**	0,126***	Tőr	0,165***	0,168***
Sacrum	0,125***	0,150***	Barlang	0,136***	0,153***
Felhő	0,102**	0,132***			

* $p < 0,1$
** $p < 0,05$
*** $p < 0,01$

A 12. és a 13. táblából kiolvasható, hogy a skálaredukció minden esetben növeli az iskolázottsággal való korrelációt és rangkorrelációt – a legtöbb esetben jelentős mértékben. Ha pedig többszörös korrelációs együtthatót (R) számolunk a 29 binarizált változó és a végzett osztályok száma között, akkor R értéke az eredeti alakokkal 0,50, a binarizált alakokkal pedig 0,57 lesz, a korrigált R^2 pedig rendre 0,19, illetve 0,26, ami a binarizálás figyelemre méltó előnyét mutatja.

A korrelációk növekedése azért is meglepő, mert skálaredukció esetén általában a korreláció csökkenését várjuk. Például *Cohen* [1983] említi, hogy ha (X, Y) normális együttes eloszlású változópár, akkor bármelyiküket a medián pontjában binarizálva, a korreláció köztük az eredetinek körülbelül a 80 százalékára csökken:

$$\rho(X_{bin}, Y) = \rho(X, Y_{bin}) = \sqrt{\frac{2}{\pi}}\,\rho(X, Y) \approx 0{,}8\rho(X, Y) \qquad /8/$$

(a bizonyítást illetően lásd *Vargha–Rudas–Delaney–Maxwell* [1996]). *Peters* és *Van Voorhis* ([1940] 394. old.) képletei alapján *Cohen* [1983] úgy gondolta, hogy ha mindkét változót binarizáljuk a medián segítségével, akkor a korreláció az eredeti 0,8·0,8 = 0,64 részére csökken. Ez a feltételezés ugyan téves, mert *Vargha et al.* [1996] levezetése szerint a fenti helyzetre a

$$\rho(X_{bin}, Y_{bin}) = \frac{2}{\pi}\arcsin(\rho(X, Y)) \approx 0{,}64\arcsin(\rho(X, Y)) \qquad /9/$$

összefüggés érvényes, de kisebb (hozzávetőlegesen a 0,5 alatti) korrelációk esetén Cohen javaslata jó közelítésként elfogadható (*Vargha et al.* [1996] 1. tábla). Megjegyezzük, hogy ha kétdimenziós normális eloszlás esetén a binarizálást a mediántól különböző pontban végezzük, a korreláció még nagyobb mértékben csökken.

A 12. és a 13. tábla adatai azonban azt mutatják, hogy a binarizálás nemcsak hogy nem csökkentette a korrelációt, hanem kivétel nélkül minden esetben növelte, sok esetben tetemesen. Ez egyrészt a normális eloszlástól való nagymértékű eltérés következménye, másrészt azt is felveti, hogy a vizsgált változók értékskálája még csak nem is folytonos jellegű. Esetünkben a 29 binarizált változó mindegyikét a 0+ε érték segítségével binarizáltunk. Ezek tehát mind egy-egy Rorschach-tartalom indikátorváltozói. Ha egy változó ilyen formában informatívabb, mint eredeti alakjában, akkor ez azt jelenti, hogy az 1-nél több előfordulásokat egymástól megkülönböztetve – a változók eredeti százalékos formája éppen ezt teszi – romlik az iskolázottsági szintek diszkriminációja. Ezen összefüggésnek fontos pszichometriai következménye van.

*

Végezetül szeretnénk megjegyezni, hogy a binarizálással kapcsolatban fentebb felsorolt pozitív tapasztalatok alapján nem szeretnénk arra buzdítani, hogy ezután mindig, minden esetben térjünk át a többértékű ordinális vagy kvantitatív változókról alkalmas osztópontokkal az egyszerűbb bináris változóformákra. Ilyen döntéshez elengedhetetlen a kellő statisztikai alátámasztás és a szakmai relevancia. Pusztán arra akartuk felhívni a figyelmet, hogy a binarizálással nem veszítünk automatikusan információt, és hogy egyes jellegzetesen nem normális eloszlású változók esetén ez az eljárás akár eredményesebb statisztikai elemzésekre is vezethet, mint az eredeti változóformákkal végzett elemzések. Azt a tanulságot azonban mindenképpen levonhatjuk, hogy a társadalomtudományi kutatásokban a mélyebb statisztikai elemzések végrehajtása előtt mindig célszerű végiggondolni, hogy változóink, amelyeket igen gyakran bizonyos rejtett tulajdonságok mérésére használunk, vajon a legjobban vannak-e definiálva.

IRODALOM

BRUNNER, E. – MUNZEL, U. [2000]: The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal*. 42. évf. 1. sz. 17–25. old.

BRUNNER, E. – PURI, M. L. [2001]: Nonparametric methods in factorial designs. *Statistical Papers*. 42. évf. 1. sz. 1–52. old.

COHEN, J. [1983]: The cost of dichotomization. *Applied Psychological Measurement*. 7. évf. 3. sz. 249–253. old.

DELANEY, H. D. – VARGHA, A. [2002]: Comparing several robust tests of stochastic equality with ordinally scaled variables and small to moderate sized samples. *Psychological Methods*. 7. évf. 4. sz. 485–503. old.

DEMETROVICS ZS. [2005]: *Különböző típusú drogok használatának személyiségpszichológiai és családi háttere*. PhD-értekezés. Budapest. (Kézirat.)

DIXON, W. J. [1990]: *BMDP Statistical Software Manual*. University of California Press. Berkeley.

FLEISS, J. L. – LEVIN, B. – PAIK, M. C. [2003]: *Statistical methods for rates and proportions* (3rd Ed.). Wiley. New York.

HAJTMAN B. [1968]: *Bevezetés a matematikai statisztikába pszichológusok számára*. Akadémiai Kiadó. Budapest.

HOLLANDER, M. – WOLFE, D. A. [1999]: *Nonparametric statistical methods* (2nd Ed.). Wiley. New York.

LEHMANN, E. L. [1975]: *Nonparametrics: Statistical methods based on ranks*. Holden-Day. San Francisco.

MICCERI, T. [1989]: The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*. 105. évf. 1. sz. 156–166. old.

OLÁH A. [1985]. A Californiai Pszichológiai Kérdőív hazai alkalmazásával kapcsolatos tapasztalatok. In: *Pszichológiai Tanulmányok*. XVI. kőtet 53–101. old.

OVERALL J. E. [1968]: Standard psychiatric symptom description: The Factor Construct Rating Scale (FCRS). *Triangle: Sandoz Journal of Medical Sciences*. 8. évf. 6. sz. 178–186. old.

PETERS, C. C. – VANVOORHIS, W. R. [1940]: *Statistical procedures and their mathematical bases*. McGraw-Hill. New York

PETHŐ B. [1972]: *A psychiatriai nozológia értékességéről, a cycloid és a hebephren psychosisok differentialdiagnosztikája kapcsán*. Kandidátusi disszertáció. Magyar Tudományos Akadémia. Budapest.

PETHŐ B. – SZILÁGYI A. – HAJTMAN B. [1977]: A psychiatriai tüneti kép módositott Rockland-Pollin-féle becslésskálával történő felmérése cycloid és hebephren betegeknél. *Ideggyógyászati Szemle*. 30. évf. 4. sz. 155–175. old.

ROCKLAND I. H. – POLLIN W. [1965]: Quantification of psychiatric mental status. *Archives of General Psychiatry*. 12. évf. 1. sz. 23–28. old.

RUYMGAART, F. H. [1980]: A unified approach to the asymptotic distribution theory of certain midrank statistics. In: *J. P. Raoult* (szerk.), Statistic *non Parametrique Asymptotique*, Lecture Notes on Mathematics. 821. sz. Springer Verlag. Berlin. 1–18. old.

SZÉKELYI M. – BARNA I. [2003]: *Túlélőkészlet az SPSS-hez*. Typotex Kiadó. Budapest.

VARGHA A. [1989a]: *A nem, az életkor, az iskolázottság és a diagnózis hatása az egyes Rorschach-jegyekre*. Egységes jegyzet. Tankönyvkiadó. Budapest.

VARGHA A. [1989b]: *A magyar Rorschach Standard táblázatai*. Egységes jegyzet. Tankönyvkiadó. Budapest.

VARGHA A. [2000]: *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. Pólya Kiadó. Budapest.

VARGHA A. [2002]: *Független minták egyszempontos összehasonlítása új rangsorolásos eljárások segítségével*. Statisztikai Szemle. 80. évf. 4. sz. 328–353. old.

VARGHA A. [2004]: A kétszempontos sztochasztikus összehasonlítás modellje. *Statisztikai Szemle*. 82. évf. 1. sz. 67-82. old.

VARGHA A. – CZIGLER B. [1999]: *A MiniStat statisztikai programcsomag, 3.2 verzió.* Pólya Kiadó. Budapest.

VARGHA, A. – DELANEY, H. D. [1998]: The Kruskal-Wallis test and stochastic homogeneity. *Journal of Educational and Behavioral Statistics*. 23. évf. 2. sz. 170–192. old.

VARGHA, A. – DELANEY, H. D. [2000]: A critique and improvement of the CL common language effect size statistic of McGraw and Wong. *Journal of Educational and Behavioral Statistics*. 25. évf. 2. sz. 101–132. old.

VARGHA, A. ET. AL [1996]: Dichotomization, partial correlation, and conditional independence. *Journal of Educational and Behavioral Statistics*. 21. évf. 3. sz. 264–282. old.

VINCZE I. [1968]: *Matematikai statisztika ipari alkalmazásokkal*. Műszaki Könyvkiadó. Budapest.

WILCOX, R. R. [1996]: *Statistics for the social sciences*. Academic Press. San Diego, New York.

SUMMARY

In the empirical research of social sciences serious efforts have been made to define variables that are at least ordinally scaled. Along with these efforts a number of new methods have recently been developed for group comparisons with ordinal dependent variables.

The present article describes some hard to manage cases (such as the no best situation) of stochastic comparisons. It shows also that the probable cause of these strange situations is that under the violation of normality assumption certain types of discriminative information are densed in some special critical points of the scale. The paper presents several empirical studies with highly non-normal distributions. Here it is demonstrated that in dichotomizing these scales using the above mentioned critical points as cut-points, instead of loosing information we often reach more efficient group discrimination and a better prediction in multiple regression.