

A ROPstat statisztikai programcsomag*

Vargha András,

a Károli Gáspár Református Egyetem, valamint az Eötvös Loránd Tudományegyetem Pszichológiai Intézetének egyetemi tanára

E-mail: vargha.andras@kre.hu

A ROPstat egy általános és széles spektrumú statisztikai programcsomag, mely három területen nyújt különleges lehetőségeket: 1. robusztus eljárások, 2. ordinális elemzések, 3. mintázatfeltáró módszerek. Sok közülük semmilyen más ismertebb programcsomagban nem érhető el. Jelen tanulmányban bemutatjuk a ROPstat főbb vonásait, ezután részletesen ismertetjük menürendszerét és a programban futtatható statisztikai eljárásokat, majd egy konkrét példán szemléltetjük a program használatának egyszerűségét.

TÁRGYSZÓ:

Robusztus statisztikák.

Ordinális elemzések.

Mintázatfeltáró elemzések.

DOI: 10.20311/stat2016.11-12.hu1165

* A szerző köszönetet mond *Takács Szabolcs*-nak és *Vita László*-nak, akik számos hasznos megjegyzést fűztek e cikk első változatához.

A ROPstat egy általános és széles spektrumú statisztikai programcsomag, mely három területen nyújt különleges lehetőségeket: robusztus eljárások (R), ordinális elemzések (O), mintázat/pattern feltáró módszerek (P).

Sok közülük semmilyen más ismertebb programcsomagban nem érhető el. A ROPstat két (magyar és angol) nyelven kommunikáló szoftver, melyet e cikk szerzője tervezett és írta különböző statisztikai moduljainak a programjait (Delphi programozási nyelven). *Bánsági Péter* matematikus mérnöknek köszönhető a programok Windows operációs rendszerhez igazítása és a teljesképernyő-kezelés. *Lars Bergman* szakmai konzultánsként segített a mintázatfeltáró-modulok többségének kidolgozásában.

A jelen cikkben először bemutatjuk a ROPstat főbb vonásait, részletesen ismertetjük menürendszerét és a programban futtatható statisztikai eljárásokat, majd egy konkrét kis példán szemléltetjük a program használatának egyszerűségét.

1. A ROPstat installálása

A ROPstat Windows operációs rendszerben futtatható szoftver, melynek demó verziója a ROPstat saját honlapjáról (www.ropstat.com) tölthető le. A demó verzió sikeres installálása során a ROPstat futásához szükséges összes fájl a számítógép „c” meghajtójának „c:_vargha\ropstat” nevű mappájába másolódik be. A ROPstat futásához az itt megtalálható ropstat.exe programot kell elindítani. Maga a demó verzió is már egy tökéletesen működő ROPstat program, mely csupán egy korlátozásnak van alávetve: az elemzendő adatfájl nem tartalmazhat 5-nél több változót és 500-nál több esetet (nagyobb beolvasott fájlok erre a méretre lesznek redukálva). A profi verzió képes 1 000 különböző változó és esetenként 100 000 személy statisztikai elemzésére. Normál és profi verziók elérhetőségéről minden információ megtalálható a szoftver honlapján.

2. Adatimport és adatexport a ROPstatban

A ROPstat saját típusú és kiterjesztésű – msw – adatfájlokkal dolgozik, ilyeneket hoz létre és olvas be. Az msw egyébként egyszerű szövegfájl, amelyben a változójellemzők és az adatok egy speciális formátumban tárolódnak. Ezenkívül van még két

olyan fájl típus, melyet a ROPstat közvetlenül be tud olvasni a Fájl/Megnyit menüpontban: az SPSS portable (*.por) és az Excel tabulátorral tagolt szövegfájl (*.txt) típusát. Megjegyezzük, hogy az SPSS legfrissebb verzióiban (21. verzió és e felett), a portable formátumban való elmentéshez a mentés előtt a „unicode character set” beállítása szükséges a „local character set” opcióban az Edit/Options/General menüpontban, mégpedig az átkonvertálandó SPSS fájl betöltése előtt.¹ Excelből importált szövegfájlok esetén az első sorban kell elhelyezni a változók rövid (maximum 8 karakter hosszú) neveit. Hosszabb nevek esetén a ROPstat hosszú névként őrzi meg őket, és az első 8 karakter alapján képezi a rövid nevet.

A ROPstatban a változók egyaránt lehetnek numerikus és szöveg típusúak, de a ROPstat csak numerikusan kódolt változókkal tud statisztikai elemzéseket végezni. Speciális opcióként azért a ROPstat képes egy szöveges változó automatikus numerikus átkódolására egyszerű szöveges címkék (például férfi és nő) alkalmazása esetén a „Változók deklarációi” ablak „Típus” oszlopában. Ugyanebben az ablakban szerkeszthetők a változók rövid (maximum 8 karaktert tartalmazó) és hosszú (gyakorlatilag akármekkora) nevei, jelölhetők ki kódokkal vagy övezetekkel a változók csoportjai, továbbá adhatók meg változónként a nem értelmes – hiányzó (missing) – adatokat jelölő adatkódok vagy övezetek. A ROPstat emellett képes a beolvasott változók sokféle átalakítására (például átkódolására), valamint segítségével új (például különféle eloszlású random) változók létrehozására.

Ha egy adatfájlt SPSS-be vagy Excelbe szeretnénk konvertálni, akkor a Fájl menü „Mentés SPSS formátumban”, illetve „Mentés tabulált szövegfájl formátumban” menüpontját kell használnunk.

3. A ROPstat statisztikai menürendszere

A ROPstat konstrukciós alapelve, hogy a felhasználó elsősorban saját szakterületén legyen otthonos, ne a programozás és a matematikai statisztika technikai jellegű részleteiben. Teljesen menüvezérelt és használata nem igényel programozói ismereteket. A programcsomag használatának legfontosabb feltétele, hogy a felhasználó ismerje változóinak mérési szintjét (pszichometriai skálatípusát), valamint legyen tudatában annak, hogy milyen szakmai kérdésre szeretne adatai felhasználásával választ kapni.

Mivel a ROPstatnak nincs publikált kézikönyve, a következőkben részletesen ismertetjük a menürendszerét és az egyes modulok által elvégezhető statisztikai eljárásokat. Ezek közös jellemzője, hogy a statisztikák kiszámíthatók külön, egy ún. felté-

¹ Ugyanez a hatás elérhető úgy is, ha az SPSS sima indítás utáni üres állapotában egy „syntax” ablakot nyitunk és ott a „SET UNICODE OFF” parancsot beírjuk és futtatjuk.

teles csoportosító változó meghatározott szintjeire is (például külön férfiakra és nőkre vagy csak azokra a személyekre, akiknek 120 és 150 közötti az IQ-ja).

3.1. Alapok (egymintás elemzések): az adatállományok egyszerű leíró statisztikai jellemzői és eloszlásvizsgálatok

Az e menüpontba tartozó és továbbiakban bemutatott eljárások közös jellemzője, hogy az elemzéseket a kijelölt változókkal egyenként, külön-külön végzi el a program. Valójában mindegyikük a statisztikai eloszlásvizsgálat egy formája.

3.1.1. Alapstatisztikák

Ez a modul változónként alapstatisztikákat számít. Az eredménylista a következő statisztikákat tartalmazza: érvényes értékek (esetek) száma minden kijelölt változóra; átlag és szórás; variációs együttható (relatív szórás); legkisebb és legnagyobb érték; legkisebb és legnagyobb standardizált érték (z -érték).

3.1.2. Részletesebb statisztikák számítása

Ez a modul minden függő változóra kiszámítja a következő statisztikákat: medián, az átlag standard hibája (az elméleti átlagra vonatkozó becslés várható hibája), konfidencia-intervallum az átlagra, ferdeségi együttható (az aszimmetria standardizált mutatója) és csúcsossági együttható (vö. *Vargha* [2007] 3. fejezet és 5.3.5. alpont). A program egyben azt is teszteli, hogy a változó ferdesége és csúcsossága eltér-e a normális eloszlásétól.

Még részletesebb lesz az eredménylista, ha trimmelést is beállítunk (1 és 25 közötti trimmelési százalékkal). Az eredménylista ilyenkor a következő statisztikákat tartalmazza: átlag, szórás, trimmelt átlag, winsorizált szórás, medián; konfidencia-intervallum (intervallumbecslés) az elméleti átlagra; konfidencia-intervallum az elméleti trimmelt átlagra; ferdeségi együttható; csúcsossági együttható; normalitásvizsgálat a ferdeségi és a csúcsossági együttható segítségével. Megjegyezzük, hogy az SPSS a ferdeségi és a csúcsossági együtthatóra standard hibát és konfidencia-intervallumot is megad, azt azonban az SPSS nem említi meg, hogy ez csak a normalitás ritkán fennálló teljesülése esetén érvényes.

3.1.3. Gyakorisági eloszlás, hisztogram készítése

Ez a modul minden kijelölt változóra gyakorisági eloszlást készít.

– Folytonosként definiált változók esetében egy beépített algoritmus segítségével a program határozza meg az osztályokat, és ezekre

nézve határozza meg a gyakoriságokat, százalékos gyakoriságokat és kumulatív százalékos gyakoriságokat. Az osztályok száma az osztályszám szorzó paraméter értékének módosításával növelhető.

– Diszkrétként definiált változók esetében a program a változó minden értékére meghatározza a gyakoriságokat. E két változótípus esetén hisztogram is ábrázolja a mintabeli eloszlást. Diszkrét beállítás esetén lehetőség van minden kiválasztott X változóból speciális új változók képzésére és elmentésére (ezek az új változók az elemzés után az adattáblázat végén jelennek meg):

a) X binarizálása: az X változó minden talált c értékére olyan új bináris változó (X_c) képződik, melynek értéke 1, ha $X = c$ és 0, ha $X \neq c$.

b) X percentilis transzformáltjának képzése: az X változó minden c értékéhez hozzárendeljük a mintabeli kumulatív százalékos gyakoriságot, vagyis a mintában a c -nél kisebb vagy vele egyenlő értékek százalékos arányát.

– Csoport def. változótípus beállítása esetén a program kilistázza az illető változó különböző csoportjait, azok gyakoriságát, továbbá külön kérésre a χ^2 -próba segítségével teszteli, hogy a definiált csoportok/kategóriák elméleti gyakoriságai megegyeznek-e. Ha a változóhoz nincsenek csoportok rendelve, a program a talált különböző értékek gyakoriságát határozza meg, és ezek segítségével teszteli, hogy a változó eloszlása egyenletes-e.

3.1.4. Közéértékekre vonatkozó hipotézisek vizsgálata

Ez a modul minden kijelölt változó esetében végrehajtja az egymintás t -próbát, annak két robusztus változatát (Johnson- és Gayen-próba)², külön kérésre a trimmelt (nyesett) t -próbát, valamint a Wilcoxon- és az előjelpróbát (vö. Vargha [2007] 7. fejezet). A tesztelendő hipotézis: a populáció átlaga (t -próba és robusztus változatai), trimmelt átlaga (trimmelt t -próba), illetve mediánja (Wilcoxon- és előjelpróba) megegyezik a feladatablakban beállítható hipotetikus értékkel (A). Diszkrét változók esetén az előjelpróba nullhipotézise: $P(X < A) = P(X > A)$. A program a Wilcoxon-, illetve előjelpróba esetében egzakt p -értéket számít ki akkor, ha a hipotetikus értéktől (A -tól) való nem nulla eltérések száma nem haladja meg a 20-at (Wilcoxon-próba), illetve 50-et (előjelpróba), egyébként a normális eloszlással való közelítést alkalmazzuk.

² E két próba kiszámítására nem kerül sor, ha a mintaelemszám 500-nál nagyobb, vagy a t -érték meghaladja a 10-et.

3.1.5. Normalitásvizsgálat

Ezt az elemzést legfeljebb 100 elemű minták esetén a Kolmogorov-próba, 100-nál nagyobb elemszámú minták esetén pedig a χ^2 -próba segítségével végzi el a program (vö. Vargha [2007] 17.1.3. alpont). Ebben a menüpontban lehetőség van lineáris transzformációval minden kiválasztott változó két speciális standardizáltjának képzésére és elmentésére is:

- a) Egyszerű standardizálás: ebben az esetben az új változó átlaga 0, szórása 1 lesz.
- b) *T*-skála készítés: ebben az esetben az új változó átlaga 50, szórása 10 lesz.

Ezek az új változók az elemzés után az adattáblázat végén jelennek meg.

3.2. Csoportok és változók összehasonlítása

Az e menüpontba tartozó és a továbbiakban bemutatott eljárások közös jellemzője, hogy mindegyikük a statisztikai homogenitásvizsgálat egy formája.

3.2.1. Független minták egyszempontos összehasonlítása

Intervallum-skálájúnak beállított változókra a program az elméleti szórások és átlagok egyenlőségét vizsgálja. A populációszórások összehasonlítására a ROPstat az O'Brien és a Levene-próbát (ezeknek is a robusztus, Welch-féle változatát) hajtja végre, amelyek a normalitás megsértésére nézve kevésbé érzékenyek, mint a hagyományos eljárások (lásd Vargha [2007] 9.7. és 13.3. alfejezet). Az elméleti átlagok (várható értékek) összehasonlítására a ROPstat kétféle eljárástípust alkalmaz. Először egy hagyományos eljárást, mely érzékeny a szórások különbözőségére. Ez két csoport esetén a kétmintás *t*-próba, több csoportot tekintve pedig az egyszempontos VA (varianciaanalízis) független mintás eljárása. Ennek az elemzésnek az eredményét – különösen számottevően eltérő elemszámoknál – csak akkor tekinthetjük érvényesnek, ha a szórások nem különböznek túlságosan nagymértékben egymástól (vö. Vargha [2007] 9. és 13. fejezet).

Kettőnél több csoport esetén a program kiszámítja a korrelációs hányadost (nemlineáris korrelációs együtthatót) is, mely azt méri, hogy a függő változó milyen mértékben függ a csoportosító változótól. E mutató négyzete a nemlineáris determinációs együttható, mely azt jelzi, hogy a csoportosító változó a függő változó varianciájának hányad részét magyarázza meg.

A hagyományos eljárás mellett az elméleti átlagok összehasonlítására a ROPstat olyan robusztus módszereket is alkalmaz, amelyek érvényességét nem csökkenti jelentősen az elméleti szórások esetleges különbözősége. Két csoport esetén az alkalmazott eljárás a Welch-féle d -próba, több csoportot véve pedig a Welch-, a James-, és a Brown–Forsythe-féle robusztus VA. Ezek közül 20-nál kisebb minták esetén a Welch-, nagyobb mintáknál pedig a James-próba tűnik a legmegfelelőbbnek. A Brown–Forsythe-féle eljárás csak legfeljebb mérsékelt különböző szórások mellett megbízható (vö. *Vargha* [2007] 13.2. alfejezet).

Ha kettőnél több csoport összehasonlításakor az elméleti átlagok azonosságának hipotézise elvethető, a program páronként összehasonlítja az összes mintaátlagot. Szignifikánsan különböző szórások esetén (lásd Levene-próba eredménye) a Games–Howell-féle, egyébként a Tukey–Kramer-féle módszer eredményét célszerű figyelembe venni (vö. *Vargha* [2007] 13.4. alfejezet).

Ha a normalitás erősen sérül amiatt, hogy az adatmintában a többi közül nagyon kilógó extrém adatok is vannak, célszerű a minták szélsőséges elemeit trimmeléssel „levágni” a mintákról. Ezt 0-tól különböző trimmelési százalék beállításával tehetjük meg. Ilyenkor a program kettőnél több csoport összehasonlításakor trimmelt VA-t, két csoport összehasonlításakor pedig trimmelt kétmintás t -próbát (Yuen-próba) is végrehajt.

Ha kovariáns változót jelölünk ki, ennek hatását a program kovarianciaanalízissel szűri ki. Ezen elemzés feltételezi, hogy a kovariáns és a függő változó között ugyanolyan lineáris kapcsolat van a különböző mintákban. Ennek részleges ellenőrzésére e két változó közötti korrelációt a program csoportonként kiszámítja és összehasonlítja.

Ordinális skálájúnak beállított változók esetén a program az elméleti rangszórások megegyezését és a sztochasztikus homogenitást teszteli. Populációk összehasonlításakor egy-egy populáció sztochasztikus dominanciája (az ún. sztochasztikus kezelési hatás) annak a valószínűségét jelzi, hogy egy random megfigyelés ebből a populációból (X_j) nagyobb lesz, mint egy random megfigyelés a populációk egyesítéséből (X), plusz az egyenlőség valószínűségének a fele:

$$P_j = P(X_j > X) + 0,5 \cdot P(X_j = X).$$

A sztochasztikus homogenitás egyik definíciója: $P_1 = P_2 = P_3 = \dots = 0,50$. Egy-egy P_j érték szignifikanciája azt jelenti, hogy a $H_j : P_j = 0,5$ hipotézis elvethető. A sztochasztikus homogenitás egyébként ekvivalens az elméleti rangátlagok egyenlőségével, és köznapi szavakkal azt jelenti, hogy bármelyik populáció esetén egy innen véletlenszerűen kiválasztott adat ugyanolyan valószínűséggel nagyobb a többi populáció egyesítéséből kiválasztott véletlen adatnál, mint kisebb. E populációk egyesítése történhet a mintaelemszámokkal arányos súlyokkal, illetve azonos

súlyokkal. Ez utóbbit akkor célszerű alkalmazni, ha a mintaelemszámok aránya nem felel meg a csoportok összpapulációbéli valódi arányainak.

A sztochasztikus homogenitást hagyományosan a Mann–Whitney-próba (két minta), illetve a Kruskal–Wallis-próba (kettőnél több minta) segítségével lehet tesztelni, amelyek a kétmintás t -próba, illetve az egyszempontos VA rangszámokon elvégzett megfelelői. Számottevően különböző szórások esetén azonban e hipotézis vizsgálatára a Fligner–Policello-próba és a Brunner–Munzel-próba (két minta), illetve a korrigált rang Welch-próba, és a Kulle-féle aszimptotikusan egzakt próbák (kettőnél több minta) a legmegfelelőbbek (vö. Vargha [2002], [2007] 10. és 15. fejezet).

A program a Kruskal–Wallis-próba H mennyiségének szignifikanciáját a χ^2 -eloszlással értékeli ki, a Mann–Whitney-próba U mennyiségének szignifikanciáját kisebb minták esetén egy egzakt eljárással, nagy mintáknál pedig a normális eloszlással való közelítés segítségével állapítja meg.

Ha kettőnél több minta összehasonlításakor a sztochasztikus homogenitás hipotézise elvethető, a program a Brunner–Munzel-próbával páronként összehasonlítja az összes csoportot, az elsőfajú hiba megnövekedése, vagyis az alfa-infláció elkerülésére minden esetben kiszámítva a Bonferroni-féle korrigált p -értéket is. A szignifikancia mértékétől függetlenül a program minden ordinális skálájú változó esetén csoportonként becslést ad a sztochasztikus dominancia nagyságára, s ezeknek a 0,5-től való eltérését is teszteli.

Ha a csoportosító változónak csak két érvényes csoportja van, akkor a program kiszámítja az

$$A(1, 2) = P(X_1 > X_2) + 0,5 \cdot P(X_1 = X_2)$$

valószínűségi főlény mutatónak, valamint a $P(Csop1 > Csop2)$, $P(Csop1 < Csop2)$ valószínűségeknek a pontbecslését, illetve $A(1, 2)$ -re 95 százalékos intervallumbecslést is készít.

Ha a feladatablakban kérjük az eloszlások részletes összehasonlítását, akkor a program a függő változó értékskálájának maximum 10 pontjában (ún. binarizáló osztópontban) összehasonlítja az empirikus eloszlásokat azt tesztelve, hogy az osztópont alatti értékek aránya szignifikánsan eltér-e a különböző csoportokban. Ezen összehasonlítások p -értékeit összevetve megállapítható, hogy a csoportok az értékskála milyen dichotomizálásával különböznek egymástól a legélesebben. Az alfa-infláció elkerülésére a program Bonferroni-féle korrigált p -értékeket számol. Az osztópontelemzéssel kapcsolatban lásd még Vargha [2005], [2008] és Vargha–Bergman [2012].

A ROPstat különleges vonása, hogy ha az összehasonlítandó csoportok száma 2 és 10 közé esik, valamint a kiválasztott függő változók száma legalább kettő, akkor intervallum és ordinális típusú változók esetén a program az eredménylista végén

tömör táblázatban foglalja össze az összehasonlítások eredményeit (alapstatisztikák, hatásmértékek, próbastatisztikák, p -értékek).

Nominális skálájúnak beállított változóknál a program elkészíti a csoportosító változó (G) és a függő változó (X) csoportjai által meghatározott kétdimenziós gyakorisági/kontingencia táblázatot és χ^2 -próbával (kis elemszámok előfordulásakor Fisher-féle egzakt próbával is) teszteli, hogy X függ-e G -től (vö. *Vargha* [2007] 17. fejezet).

3.2.2. Összetartozó minták egyszempontos összehasonlítása

E modul segítségével összetartozó (páros) minták hasonlíthatók össze egymással (például ugyanazon változók különböző időpontokban mért szintjei). Az összehasonlítás statisztikai módszere attól függ, hogy mi a kijelölt függő változók beállított közös skálátípusa (intervallum, ordinális vagy nominális).

Intervallum skálájú változókra a program kettőnél több függő változó esetén az egyszempontos összetartozó mintás VA eljárásával teszteli az elméleti átlagok megegyezését és Friedman-próbával, illetve rangszámokon végrehajtott VA-val a sztochasztikus homogenitást, mely a független mintás esettel analóg módon definiálható (vö. *Vargha* [2007] 14. fejezet). A hagyományos VA mellett a program szabadságfok korrekciós (Greenhouse–Geisser- és Huynh–Feldt-féle) robusztus VA-t is végrehajt. A korrekció annál nagyobb, minél kisebbek az e két módszer szerint kiszámított epszilonok. Tekintettel arra, hogy az összetartozó mintás VA szóráshomogenitási feltétele gyakran nem teljesül, az eredmények értelmezését célszerű ezen robusztus VA-k figyelembevételével végezni. A két robusztus VA közül a Greenhouse–Geisser-féle eljárás általában egy picit konzervatív (a kellenél ritkábban szignifikáns), a Huynh–Feldt-féle pedig általában enyhén liberális (a kellenél gyakrabban szignifikáns).

A program a Friedman-próba G mennyiségének szignifikanciáját a χ^2 -eloszlással közelítve állapítja meg. Ez tíznél nagyobb elemszámú minták esetén megbízhatónak tekinthető. Ha az elméleti átlagok azonosságának, illetve a sztochasztikus homogenitásnak a hipotézise elvethető, a program Tukey módszerével páronként is összehasonlítja az összes mintaátlagot, illetve rangátlagot. A szignifikancia mértékétől függetlenül a program változónként becslést ad a többi változó együttesétől való sztochasztikus eltérés nagyságára $(A(k, u) - 0,5)$,³ s ezen eltérések szignifikanciáját is megvizsgálja.

Két függő változó esetén a program az elméleti átlagok megegyezésére VA helyett a vele ekvivalens egymintás t -próbát, illetve annak robusztus változatait (John-

³ $A(k, u)$ annak a valószínűsége, hogy a k -adik változónak egy random értéke nagyobb, mint a többi változó értékeiből véletlenszerűen kiválasztott érték, plusz az értékegyenlőség valószínűségének a fele.

son- és Gayen-próba), a mediánok egyenlőségének tesztelésére a Wilcoxon-próbát, a sztochasztikus egyenlőség tesztelésére pedig az előjelpróbát hajtja végre (vö. *Vargha* [2007] 8. fejezet). A Wilcoxon-próba érvényességének feltétele ebben az esetben az, hogy a két változó különbsége szimmetrikus eloszlású legyen. Ez teljesül például abban az esetben, ha a két változó legfeljebb az adatok nagyság szintjében különbözik egymástól (az eloszlás alakjában nem). Ilyenkor a Wilcoxon-próba ugyanazt a nullhipotézist teszteli, mint az egymintás t -próba. Folytonos eloszlású változók esetén az előjelpróba nullhipotézise ekvivalens azzal a megállapítással, hogy a két változó különbségének mediánja nulla.

Ha a normalitás erősen sérül amiatt, hogy az adatmintában a többi közül nagyon kilógó extrém adatok is vannak, célszerű a változók szélsőséges elemeit trimmeléssel „levágni”. Ilyenkor a program kettőnél több változó összehasonlítása esetén trimmelt VA-t, két változó összehasonlításakor pedig trimmelt egymintás t -próbát is végrehajt.

Ordinális változókra a program csak a mediánok összehasonlítását és a sztochasztikus egyenlőséget (két változó), illetve csak a sztochasztikus homogenitást (kettőnél több változó) teszteli (vö. *Vargha* [2007] 8. és 15. fejezet).

Nominális változók esetén a változás/eltérés vizsgálatára a program a függő változók és a talált különböző értékek számától függően különböző diszkrét elemzéseket végez (vö. *Vargha* [2007] 17.4. alfejezet, illetve *Vargha* [2001]). Ha két függő változót jelölünk ki, akkor a program az eloszlások összehasonlítására a McNemar-próbát (más néven Bowker-féle szimmetria-próbát) hajtja végre, kettőnél több függő változónál pedig a függő változók minden talált különböző értékére a Cochran-féle Q -próbát. Ha például a függő változók száma három, s ezek mindegyike olyan ötértékű változó, melynek lehetséges értékei az 1, 2, 3, 4, 5 egész számok, akkor a program mind az öt értékre vonatkozóan a Cochran-féle Q -próbával teszteli, hogy az érték ugyanolyan valószínűséggel fordul-e elő a három függő változónál.

3.2.3. Átlagok kétszemponos összehasonlítása

Ezzel a modullal független minták kétszemponos varianciaanalízise végezhető el (vö. *Vargha* [2007] 16.1. és 16.2. alfejezet). A mintaelemszámok eltérők is lehetnek a csoportosító faktorok különböző kombinációiban. A csoportosító faktoroknak kereszttezetteknek kell lenniük (és nem beágyazottnak). A program a VA szóráshomogenitási feltételének fennállását a Levene-próba robusztus Welch-féle változatával teszteli. Ha szóráshomogenitás érezhetően sérül, célszerű a program által szintén elvégzett robusztus VA-k (Welch-Johansen-féle robusztus változatok) eredményét alapul venni. Ha 0-tól különböző trimmelési százalékot állítunk be, a program robusztus kétszemponos trimmelt VA-t is végrehajt, ami extrém értékek jelenlétekor lehet gyógyír a normalitás feltételének a sérülésére. Ha a VA valamely

kettőnél több szintű faktora szignifikáns, akkor a program utóelemzések segítségével könnyíti meg a hatás értelmezését.

Megjegyezzük, hogy az itt elérhető robusztus VA-k nem találhatók meg más közismert programcsomagokban. Az SPSS ugyan felajánlja opcióként a szóráshomogenitás tesztelését, de nem számítja ki alternatív érvényes VA-t, ha ez a feltétel sérül.

3.2.4. Kétszemponos vegyes varianciaanalízis

Ezzel a modullal olyan kétszemponos VA végezhető el, amelynek egyik szempontja független mintás (csoportosító faktor), másik szempontja pedig összetartozó mintás (ismétléses faktor) (vö. *Vargha* [2007] 16.1. és 16.3. alfejezet). A mintaelemszámok eltérők is lehetnek.

A program a hagyományos VA mellett automatikusan elvégző egy robusztus kétszemponos VA-t is, mely a csoporthatást Welch módszerével teszteli, az ismétlés szempontjának hatását, a két szempont interakcióját pedig egy szabadságfok-korrektív VA-val (Greenhouse–Geisser- és Huynh–Feldt-féle robusztus VA).

3.2.5. Kétszemponos rang-varianciaanalízis

Ezzel a modullal független minták kétszemponos sztochasztikus homogenitásvizsgálata (vö. *Vargha* [2004]) végezhető el. Egy (i, j) alpopuláció P_{ij} sztochasztikus dominanciaértéke annak a valószínűsége, hogy egy ebből az alpopulációból véletlenszerűen kiválasztott X_{ij} érték nagyobb lesz, mint az egész populációból véletlenszerűen kiválasztott X érték, plusz annak a fele, hogy egyenlők lesznek:

$$P_{ij} = P(X_{ij} > X) + 0,5 \cdot P(X_{ij} = X).$$

A sztochasztikus dominanciaérték az X függő változó nagyságintjét méri a két szempont (jelölje őket mondjuk A és B) szintjeinek minden kombinációjára. Az A szempont sztochasztikus homogenitása, ha az A szempont szintjeinek számát a -val jelöljük, így néz ki:

$$P_1 = P_2 = \dots = P_a,$$

ami azt jelenti, hogy az A szempont egyik szintjén sem nagyobbak (és nem is kisebbek) az adatok, mint a többi együttesében. Analóg módon definiáljuk a B szempont sztochasztikus homogenitását is.

A két szempont nincs sztochasztikus interakcióban, ha a sztochasztikus dominanciaértékek mintázata a szempontok különböző szintjein ugyanolyan, vagyis, ha szintenkénti grafikonjaik párhuzamosak. A sztochasztikus interakció és a sztochasztikus homogenitás tesztelését mindkét csoportosító változóra a program a Kulle-féle közelítő eljárással hajtja végre.

Megjegyezzük, hogy ilyen elemzés semmilyen más közismert programcsomagban nem található.

3.2.6. Inter-rater reliabilitás vizsgálat kvantitatív skálákkal

Ezzel a modullal kvantitatív skálák inter-rater reliabilitása (a bírálói egyetértés mértéke) vizsgálható az ICC (intraclass correlation coefficient – interklaszter korrelációs együttható) korrelációs együttható segítségével (vö. *Bartko* [1976]). Az elemzéshez kettő vagy több bírálónak ugyanazokat a személyeket kell értékelnie ugyanazon kvantitatív változók (skálák) szerint, de nem szükséges minden bírálónak minden személyt minősítenie. Egy-egy személy egy-egy bírálótól származó értékeléseit mindig külön sorban kell elhelyezni és a sorokban külön változóban kell jelölni a bíráló azonosító kódját is.

4. Változók kapcsolatának a vizsgálata

Az összes ebbe a fő menüpontba tartozó modulra (az itemanalízis és a többszörös lineáris regresszió kivételével) igaz a következő. Ha a feladatablakban csak X vagy csak Y változók vannak kijelölve, a program az e csoporton belüli összes párra vizsgálja a változók kapcsolatát. Amennyiben X és Y változót is kijelölünk és a „Keresztben” opció van érvényben, akkor a program megvizsgálja az összes kijelölt X és Y változó közötti kapcsolatot (ekkor X -eken és Y -okon belül nem). Ha a kijelölt X és Y változók száma megegyezik és a „Párban” opció van érvényben, akkor a program csak az egymással összetartozó (X, Y) párok közötti kapcsolatot vizsgálja.

4.1. Korreláció, egyszerű regresszió

Ez a modul változópárokra számítja ki az egyszerű lineáris regressziós egyenletet és a két változó közötti kapcsolat szorosságának mérésére egy sor korrelációs mérőszámot, intervallumbecslésükkel együtt (vö. *Vargha* [2007] 11. és 12. fejezet). A kijelölt változókból képezett párokra a program az egyszerű (Pearson-féle) lineáris

korreláción kívül még három kapcsolati mutatót számíthat ki. Az első a Pearson-féle korrelációs együttható egyik legjobb robusztus paraméteres változatának tekinthető Wilcox-féle r_{pb} korreláció (percentage bend correlation), mely nagy széli súlyú eloszlások, illetve extrém értékek előfordulása esetén javallt. A program ezt minden esetben automatikusan kiszámítja. Ha beállítottuk a monotonitási mutatók kiszámítását, a Kendall-féle tau, illetve tau-b is kiszámítható. Előbbi csak folytonos változónál javasolt mutató. Folytonos esetben tau és tau-b megegyezik. A Kendall-féle tau, illetve tau-b mutatót tekintve a program mindig kiszámítja az adott mintában a pozitív kapcsolati arányt (konkordancia százalék), illetve a negatív kapcsolati arányt (diszkordancia százalék). A program az összes kiszámított kapcsolati mutató elméleti értékeire (beállítható szintű) konfidencia-intervallumot is megad.

A program az előzőkön kívül regressziót számít (X -ből Y -ra és Y -ből X -re egyaránt), ha ez a lehetőség nincs letiltva. A regressziók jóságának megítéléséhez a program kiszámítja az X - és az Y -hibavariáciát is, vagyis X , illetve Y regressziós becslésének átlagos négyzetes hibáját, megadva négyzetgyöküket is, a regresszió standard hibáját.

E modul feladatablakában lehetőség van egy rövid korrelációs output kérésére is. Ekkor a program minden kijelölt X változó esetén egymás alatt listázza X -nek a kiválasztott Y változókkal való legfontosabb korrelációit, a hagyományos Pearson-féle r , a Wilcox-féle robusztus r_{pb} és a Kendall-féle tau-b mutatót. Ezen opcióhoz az szükséges, hogy legyen mind X , mind Y változó kijelölve. Ilyen esetben a program sem regressziót nem számol, sem intervallumbecslést nem készít, de változópáronként feltünteti a konkordancia százalék és a diszkordancia százalék értékét.

4.2. Korrelációs és parciális korrelációs mátrixok készítése

Ezzel a modullal a kijelölt változópárok korrelációját és parciális korrelációját számíthatjuk ki három különböző korrelációs együtthatótípusra (Pearson, Spearman és Kendall; vö. *Vargha* [2007] 11. és 12. fejezet). Az utóbbi két típust leginkább ordinális változók esetén célszerű használni. A számításokat a modul csak azokra az esetekre végzi el, amelyeknek érvényes értékük van minden kijelölt változót tekintve.

Ha csak X vagy csak Y változók vannak kijelölve, a program az e csoporton belüli összes párra vizsgálja a változók kapcsolatát, valamint a Pearson-korrelációtípus kiválasztása esetén elvégzi a Bartlett-féle függetlenségvizsgálatot. Ez azt teszteli, hogy van-e a véletlennél szorosabb lineáris kapcsolat a kijelölt változók csoportjában.

Ha a feltételes csoportok száma nem haladja meg az 5-öt, akkor a program külön kérésre páronként összehasonlítja az egyazon változópárokhoz tartozó korrelációs együtthatókat.

4.3. Diszkrét változók kapcsolatvizsgálata

Ez a modul diszkrét (például nominális skálájú) változók kapcsolatának vizsgálatára alkalmas (vö. *Vargha* [2007] 17. fejezet). Segítségével a kijelölt változópárok kétszemponos gyakorisági táblázata készíthető el, tesztelhető a változópár függetlenségének nullhipotézise (χ^2 -, illetve Fisher-féle egzakt próbával) és mérhető a változók különféle típusú kapcsolatának szorossága is. Ezen mutatók, valamint a különféle speciális gyakorisági táblázatok (várt gyakoriságok, sor- és oszlopösszegek szerinti százalékok, összelemszám szerinti százalékok) kilistázása opcionálisan beállítható. „No” beállításnál sosincs kilistázás, „Yes” beállításnál mindig van, „If signif.” esetén pedig a program csak akkor listázza ki a szóban forgó táblázatot, illetve mutatókat, ha a függetlenség nullhipotézise elutasítható legalább 10 százalékos szinten.

Ha a gyakorisági táblázat sorainak és oszlopainak a száma megegyezik, akkor a program opcionálisan Cohen-féle bírálói (interrater) kappa reliabilitást is számol.

4.4. Itemanalízis

A program a kijelölt változók, mint tesztételek együttese által definiált ún. additív skála hagyományos megbízhatósági (reliabilitási) vizsgálatát végzi (vö. *Nagybányai* [2006]). Ennek keretében vizsgálja a skála belső egységességét (konzisztenciáját) a Cronbach-alfa segítségével, illetve az egyes tételek (itemek) jóságát az item-totál és item-maradék korrelációkkal.

A skála tartalmazhat pozitív és negatív tételeket (itemeket) is, amelyeket külön ablakban kell kijelölni. Negatív tételek kijelölésekor a tételek átfordításához meg kell adni a skálák közös elvi minimumát és maximumát. Ha ez nem azonos a különböző negatív tételek esetén, az elemzések bármely érték beállítása mellett érvényesek maradnak (Cronbach-alfa, item-maradék korrelációk), csak a skálaértékhez hozzáadódik egy bizonyos pozitív vagy negatív konstans. A modulban definiált skála kiszámítható és elmenthető.

Az eredménylista minden esetben tartalmazza a Cronbach-alfa értékét, valamint itemenként az IT (item-totál), az IM (item-maradék) korrelációt, továbbá a Cronbach-alfa értékét abban az esetben, ha az illető itemet elhagynánk a skálából. IT az item és a teljes skála közötti Pearson-korreláció, IM pedig az item és a többi item összege, vagyis az item és az adott item értékével csökkentett skálaérték korrelációja. Ez utóbbit szokták (például az SPSS-ben) korrigált item-totál korrelációnak is nevezni.

4.5. Többszörös lineáris regresszió

A modulban több kvantitatív független (X) változók segítségével elkészíthetjük egy kvantitatív függő (Y) változó regressziós becslését. Ez a regressziós becslés (az

ún. regressziós függvény) az X változók olyan súlyozott összege (lineáris kombinációja), melynek az Y változótól való átlagos négyzetes eltérése az adott mintában a lehető legkisebb, vagyis amellyel ebben a mintában az Y változó az X változók segítségével, lineáris képlet alkalmazásával a legkisebb hibával, illetve a legpontosabban becsülhető (vö. *Vargha* [2007] 11. alfejezet, *Székelyi–Barna* [2003] 5. fejezet). Ez a minimalizált átlagos négyzetes eltérés a regressziós becslés hibavariációjára, ennek négyzetgyöke pedig a regressziós becslés standard hibája. Az X változók súlyait (szorzótényezőit) regressziós együtthatóknak, a regressziós függvény és az Y változó közötti Pearson-féle korrelációt pedig többszörös korrelációs együtthatónak (szokásos jele: R) nevezzük. A többszörös lineáris regressziós becslésre az is igaz, hogy az X változók olyan súlyozott összege, amely az Y változóval a lehető legerősebben korrelál. Az elemzésben a regressziós becslés külön változóként létrehozható és beilleszthető az adattáblázatba.

Ha a feladatablakban a lépésenkénti (stepwise) módszert választjuk, a regressziós egyenletbe csak azok a változók kerülnek be, amelyeknek önálló szignifikáns lineáris előrejelző információjuk van. A lépésenkénti regresszió eredménylistája három blokkra tagolódik.

1. Az első blokkban a program lépésenként tájékoztat a beválasztott független változóról, az eddig a lépésig beválasztott változók többszörös korrelációjáról (R) és ennek négyzetéről, a többszörös determinációs együtthatóról (R^2), valamint az ezen lépéshez tartozó regressziós becslés standard hibájáról.

2. A második blokkban a program arról ad információt, hogy az egyes lépésekben a regressziós egyenletbe be nem választott független változóknak mi lesz a parciális korrelációja a függő változóval a már beválasztott független változók hatásának kiszűrése után. Ebből minden lépésben megtudhatjuk, hogy melyek azok a független változók, amelyek az éppen beválasztott változóval azonos előrejelző információt tartalmaznak.

3. A harmadik blokkban a program megadja a regressziós egyenletbe került változók regressziós együtthatóit eredeti és standardizált formában (béta-együtthatók), valamint a regressziós együtthatók standard hibáját és szignifikanciáját.

Ha nem a lépésenkénti regressziós módszert választjuk, akkor a regressziós egyenletbe minden olyan független változó bekerül, amelynek nem nulla a szórása és nem fejezhető ki más változók súlyozott összegeként. Ilyenkor az eredménylista megadja a többszörös korrelációs és determinációs együtthatót (R és R^2), a regressziós becslés standard hibáját, valamint a lépésenkénti elemzés harmadik blokkjában közölt információt.

5. Mintázatfeltáró elemzések

Ebben a fő menüpont olyan modulok csokrá, amelyek segítségével személyek klaszszifikációs elemzése végezhető, az ezeket kiegészítő elő- és utóelemzésekkel. Előbbiek közé tartozik a hiányzó adatok struktúrájának feltárása, a hiányzó adatok pótlása és a szélsőséges személyek azonosítása. Az utóelemzéseket a különböző klasszifikációk összehasonlítását lehetővé tevő Centroid, illetve Exacon modul képviseli. Az utolsó két modul speciális lehetőséget kínál az adatállomány alakjának átszerkesztésére.

5.1. Adatleírás

Ebben a modulban a program alapstatisztikákat számol és tájékoztat a hiányzó adatok mintázatáról (vö. *Vargha–Torma–Bergman* [2015]). Az eredménylista tartalmazza a változók alapstatisztikáit, opcionálisan a kiválasztott változók páronkénti korrelációit, valamint a hiányzó adatok számát változónként és változópáronként. Az elemzésből megtudhatjuk, hogy egy-egy változó elhagyásakor mennyivel nő a minden változóra érvényes értékkel rendelkező – komplett – esetek száma. A modulban lehetőség van arra is, hogy az elemzésre kiválasztott változók tetszőleges súlyozott összegét, mint egy új változót a program létrehozza és elhelyezze az adattáblázat utolsó oszlopában. Ha a feltételes csoportok száma nem haladja meg az 5-öt, és a feladatablakban kérjük a kijelölt változók korrelációs mátrixának elkészítését, akkor a program külön kérésre páronként összehasonlítja az egyazon változópárokhoz tartozó korrelációs együtthatókat.

5.2. Hiányzó adatok pótlása

Ennek a modulnak a segítségével hiányzó adatok pótolhatók a következő három módszer valamelyikével (vö. *Bergman–Magnusson–El-Khoury* [2003] 107–111. old., *Vargha–Torma–Bergman* [2015]).

1. Adathelyettesítés az átlaggal. Minden változó esetében minden hiányzó adat helyére az adott változó érvényes értékeinek átlagát teszi a program.

2. Adathelyettesítés a leghasonlóbb eset (iker/legközelebbi szomszéd) adataival. A program minden hiányzó adat esetén megnézi, hogy az érvényes értékek mintázata mely eset értékmintázatára hasonlít a leginkább azok közül, amelyeknek a kijelölt változók közül mindegyik értéke érvényes. A hiányzó adatot a program ennek megfelelő változóér-

tékével helyettesíti. A hasonlóság mértéke lehet a szokásos euklideszi távolság vagy az értékmintázatok Pearson-féle korrelációja. A feladatablakban kérhető a változók standardizálása, mely feltétlenül indokolt akkor, amikor az elemzésre kiválasztott változók különböző dimenziójúak (mértékegységűek). A távolság kiszámításakor lehetőség van arra is, hogy a program a változókat különböző mértékben súlyozza.

3. Adathelyettesítés többszörös lineáris regresszióval. Ezt leginkább akkor célszerű alkalmazni, ha a hiányzó értékekkel rendelkező változók közepes vagy erős lineáris korrelációban vannak egy vagy több más változóval. Ha ez nem teljesül, inkább az előző módszer alkalmazása javasolt. E típus választása esetén a program változónként kiszámítja a többi változó által megmagyarázott varianciaarányt is.

5.3. Reziduálanalízis

Ebben a modulban az adatállomány extrém, kilógó személyei azonosíthatók a hozzájuk leghasonlóbb más esetek (legközelebbi szomszédok) megkeresése segítségével (vö. *Bergman–Magnusson–El-Khoury* [2003] 109–110. old., *Vargha–Torma–Bergman* [2015]). A program azt tekinti kilógó személynek, akinek nincs legalább k számú olyan szomszédja, aki a feladatablakban beállítható küszöbtávolságnál közelebb van hozzá. A k paraméter értéke 1 és 5 között beállítható szám. Az elemzés alapján olyan indikátorváltozó készíthető, amely a kilógó értékeknél 0, a megmaradók esetében 1. Ezt a program – kérésre – hozzáilleszti az adattáblázathoz, s ennek felhasználásával a kilógó értékek más elemzésekből kihagyhatók.

A modulban kérhető sűrűsödési változók létrehozása is, melyek – különböző sugarú szomszédsági köröket alkalmazva – megadják az eset szomszédsági sűrűségét (sok közeli szomszéd előfordulásakor nagy, kevés szomszéd esetén kis értékkel).

A hasonlóság mértéke lehet a szokásos euklideszi távolság (pontosabban a változóértékek átlagos négyzetes eltérése) vagy az értékmintázatok Pearson-féle korrelációja. A feladatablakban kérhető a távolságok kiszámításakor a változók standardizálása, mely feltétlenül indokolt, ha az elemzésre kiválasztott változók különböző dimenziójúak (mértékegységűek). A távolság kiszámításakor lehetőség van arra is, hogy a program a változókat különböző mértékben súlyozza.

5.4. Hierarchikus klaszteranalízis

Ez a modul hierarchikus klaszteranalízist végez a személyeken (eseteken) a kijelölt változók felhasználásával. A klaszteranalízis típusfeltáró többváltozós statisztikai eljárás.

rás, amely személyek, változók vagy más objektumok mintájában olyan csoportokat (ún. klasztereket) hoz létre, amelyek a lehető legjobban különböznek egymástól, miközben a klasztereken belül az elemek viszonylag kis távolságra vannak egymástól. A klaszteranalízis célja tehát nagy mintán belül olyan homogén alcsoportok létrehozása, amelyek egymástól a lehető legjobban különböznek. A hierarchikus klaszterelemzés főbb lépései a következők (vö. *Bergman–Magnusson–El-Khoury* [2003] 4. fejezet és 113–115. old., *Székelyi–Barna* [2003] 3. fejezet, *Vargha–Torma–Bergman* [2015]).

1. A program minden esetpárnak meghatározza a távolságát a megadott távolságmérték felhasználásával. Ez lehet a szokásos euklideszi távolság (ASED (average squared euclidian distance – a változóértékek átlagos négyzetes eltérése⁴), vagy az értékmintázatok Pearson-féle korrelációja (két személy esetén a kijelölt változók értékeire az összetartozó értékpárok közötti Pearson-korreláció). A feladatablakban kérhető a távolságok kiszámításakor a változók standardizálása, mely feltétlenül indokolt, ha az elemzésre kiválasztott változók különböző dimenziójúak. A távolság kiszámításakor lehetőség van arra is, hogy a program a változókat különböző mértékben súlyozza. Ebben a 0. lépésben minden személyt egy egyelemű klaszternek tekintünk.

2. Ezután a program összevonja, vagyis közös klaszterbe egyesíti az egymáshoz legközelebbi két személy klaszterét (1. lépés).

3. Ezután a program lépésről-lépésre összevonja az egymáshoz legközelebbi két klasztert. Ezen összevonás módszerét a feladatablakban állíthatjuk be (alapértelmezés: Ward-féle módszer). Az a két klaszter kerül összevonásra:

- az *átlagos távolság módszernél*, amelybe tartozó személyek átlagos távolsága a legkisebb;
- a *minimális távolság (legközelebbi szomszéd) módszernél*, amely esetében a legkisebb a két klaszter egymáshoz legközelebbi lévő két személyének a távolsága;
- a *maximális távolság (legtávolabbi szomszéd) módszernél*, amelyben a legkisebb a két klaszter egymástól legtávolabbi lévő két személyének a távolsága;
- a *centroid módszernél*, amely esetében a két klaszter centroidjának (átlagvektorának) a távolsága a legkisebb;
- a *medián módszernél*, amelyben a két klaszterbe tartozó személyek távolságainak a mediánja a legkisebb;
- a *Ward-féle módszernél* (csak euklideszi távolsággal), amelyben a két klaszter összevonásával a legkisebb mértékben nő a klaszterek összheterogenitása (az ESS (error sum of square – négyzetes összhiba)). Az összheterogenitást úgy határozzuk meg, hogy minden klaszter esetén kiszámítjuk a klaszterbe tartozó személyek klasztercentroidtól való tá-

⁴ Például az [1, 3, 1] és a [2, 0, 1] adatsor esetén $ASED = (1^2 + 3^2 + 0^2)/3 = 10/3$.

volságainak négyzetösszegét, majd ezeket összeadjuk. Röviden: ennél a módszernél minden lépésben azt a két klasztert vonjuk össze, amelyben az ESS-növekedés (ESSnö.) a legkisebb.

– a *Béta flexibilis módszer*: ez egy viszonylag bonyolult, de néha igen sikeres klasszifikációt eredményező összevonási eljárás (a részletekről lásd *Bergman–Magnusson–El-Khoury* [2003] 67. old.).

4. A program feltünteti az összevonás lépéseit, az adott lépésben létező klaszterek számát (KL#), az ESS-növekedést (ESSnö.), a megmagyarázott varianciaarányt (EESS%), egy sor más klaszter adekvációs mutatót⁵ (vö. *Vargha–Bergman–Takács* [2016], *Takács–Makrai–Vargha* [2015], illetve *Vargha–Borbély* [2016]), valamint az összevonásra kerülő két klaszter kódját és a kód után zárójelben a klaszter elemszámát. EESS% az euklideszi távolságmérték esetén értelmes és jelzi az adott lépésben az osztályozás, vagyis a klaszterstruktúra jóságát. Értékét úgy definiáljuk, hogy összevetjük az adott osztályozás összheterogenitását (ESS) és a lehető legnagyobb összheterogenitást (ESSmax), amelyet nyilván akkor kapunk, amikor minden személyt egyetlen nagy közös klaszterben egyesítünk. Képlettel:

$$EES\% = 100 \cdot (ESS \max - ESS) / ESS \max .$$

EES% értéke függ az elemzésbe bevont változók számától (több változó esetén általában kisebb) és az összelemszámtól (nagyobb minta esetén általában kisebb). Sikeres osztályozás esetén EES% értéke eléri a 0,65-ös szintet. A klaszter kód a klaszterbeli személyek közül a legkisebb sorszámú, vagyis amelyik az adattáblázatban a klaszterbeli személyek közül a legfelül foglal helyet.

5. Az utolsó lépésben a program minden személyt egyetlen nagy közös klaszterben egyesít.

A feladatablakban kérhető, hogy a program listázza ki a klaszterek főbb jellemzőit (klaszternagyság, változónként átlag, szórás, minimum és maximum), továbbá a klaszterbeli személyek átlagos távolságát (a homogenitási együtthatót).

A kapott hierarchikus klaszterstruktúrán a struktúra további optimalizálása céljából relokációs elemzés (*K*-központú klaszteranalízis) is végrehajtható. Ebben az elemzésben a listázásra kért minimális klaszterszámot (alapérték: 5) fixen tartva addig cserélgetjük a különböző klaszterek elemeit egymás között, amíg EES% értéke növelhető. Ez után a program kilistázza a klaszterek alapstatisztikáit, a standardi-

⁵ PB – az esetpárokon az egyazon klaszterbe tartozás és az egymástól való távolság pontbiszerialis korrelációja (fordított előjellel), XBmod – a módosított Xie–Beni-index azt méri, hogy a két legközelebbi klaszter centrumának távolsága mennyivel nagyobb, mint a saját klasztercentrumtól való távolság átlaga, SC – Silhouette együttható azt, hogy az esetek távolsága a legközelebbi idegen klaszter centrumától mennyire van távolabb, mint a saját klaszter centrumától való távolság. HC (homogeneity coefficient – homogenitás együttható) minden klaszter esetében a klaszterbe tartozó elemek átlagos távolsága, HC-átlag ezeknek a klasztermérettel súlyozott átlaga, HC-terjedelem pedig a legkisebb és a legnagyobb HC-érték.

zált átlagok táblázatát, valamint a standardizált átlagok mintázatát, melyből kiolvasható, hogy az egyes klaszterek mely változók szerint térnek el leginkább fölfele (M) vagy lefele (A) az átlagostól.

5.5. Relokáció: nemhierarchikus, k -központú klaszteranalízis

Ebben a modulban relokációs elemzés, vagyis k -központú (k -means) klaszteranalízis végezhető (vö. *Bergman–Magnusson–El-Khouri* [2003] 4. fejezet és 115–117. old., *Székelyi–Barna* [2003] 3. fejezet, *Vargha–Torma–Bergman* [2015]). Több iterációs lépésben a program megpróbálja az esetek egyik klaszterből a másikba való áthelyezésével a klasztereket minél homogénebbé alakítani. Az elemzéshez meg kell adni a klaszterszámot vagy egy olyan változót, amely tartalmazza az esetek induló klaszterbesorolását. Ha nem adunk meg kezdeti klaszterváltozót, akkor a program először a Ward-féle módszerrel hierarchikus elemzést végez, és megáll abban a lépésben, amikor eléri a feladatablakban megadott klaszterszámot. A relokációs elemzés ez után kezdődik. A program hasonló eredményeket listáz ki, mint a nemhierarchikus elemzés relokáció után. Ha olyan kezdeti klaszterváltozót adunk meg, melynek 0 értéke is van (egy vagy több esetre), akkor a program nem végez klaszterelemzést, csak a 0 kódú eseteket szétosztja (besorolja) a többi kód által meghatározott klaszterekbe, minden személyt a hozzá legközelebbi nem 0 kódú személy klaszterébe téve.

Ebben a modulban lehetőség van arra is, hogy a program tesztelje, vajon a kapott klaszterstruktúra jobb-e, mint egy véletlenszerűen összeállított adatállományból létrehozott struktúra. Ez oly módon történik, hogy a program egymás után, egymástól függetlenül többször (a független ismétlések száma beállítható) létrehoz egy random adatállományt, elvégez azon egy k -központú klaszteranalízist, és kiszámítja a kapott megoldáshoz tartozó különböző klaszter adekvációs mutatók (QC-k (qualification coefficient – minősítési együttható)) értékét. A randomizálás történhet úgy, hogy a program egymás után, egymástól függetlenül többször véletlenszerűen összekeveri a változóoszlopok adatait az oszlopokon belül (random permutáció), vagy a program maga generál random egyenletes, illetve normális eloszlású független változókat. Ezután a program egymintás t -próbával teszteli, hogy a véletlenszerű besorolásokkal nyert QC-értékek szignifikánsan kisebbek-e, mint az eredeti adatokon végrehajtott klaszterelemzés QC-értékei. A program minden QC-re kiszámítja a MORI relatív javulási arányt is (vö. *Vargha–Bergman–Takács* [2016], *Vargha–Borbély* [2016]).

5.6. Konfigurációanalízis

A CFA (configuration analysis – konfigurációanalízis) modullal diszkrét változók tipikus és atipikus érték kombinációinak azonosítása végezhető el. E modul vá-

lasztása esetén a program a véletlennél szignifikánsan gyakrabban és ritkábban előforduló értékszekvenciákat azonosítja binomiális próbával. A program minden értékszekvenciának meghatározza a gyakoriságát és teszteli, hogy ez az érték szignifikánsan eltér-e attól a gyakoriságtól, amelyet akkor kapnánk, ha a kijelölt diszkrét változók függetlenek lennének egymástól (vö. *von Eye–Mair–Mun* [2010], *Vargha–Torma–Bergman* [2015]).

5.7. ExaCon

Ezen elemzés célja kétszemponos gyakorisági táblázatok (tipikusan különböző klaszterkód változók) celláinak elemzése egzakt próbákkal. Ebben a modulban a függetlenség esetén elvártnál szignifikánsan nagyobb és kisebb cellagyakoriságok azonosíthatók (vö. *Bergman–Magnusson–El-Khoury* [2003] 5. fejezet és 125–127. old.), amelyek segítségével például megítélhető, hogy különböző klasztermegoldások mennyire hasonlítanak egymásra. Ezt néhány speciális mutató (például a Rand, Arand, Jaccard) is elősegíti.

5.8. Centroid

Ennek a modulnak a segítségével összehasonlítható több különböző klaszterváltozó átlagmintázata (centroidja) vagy egyetlen klaszterváltozó átlagmintázata több különböző csoportban, egy csoportosító változó által definiálva (vö. *Bergman–Magnusson–El-Khoury* [2003] 5. fejezet és 125–126. old.).

5.9. DensePoint: sűrűsödéspont-elemzés

Ezzel a modullal olyan személyek azonosíthatók, akiknek sok szomszédjuk van. A szomszédságot egy beállítható küszöbérték (szomszédság küszöbe) segítségével definiáljuk. Bármely személy szomszédsága azon személyek együttese, akik az illető személytől a megadott küszöbnél nem nagyobb távolságra vannak. A távolságmérték típusa a feladatablakban beállítható. Ez lehet a szokásos euklideszi távolság vagy az értékmintázatok Pearson-féle korrelációja. A feladatablakban kérhető a távolságok kiszámításakor a változók standardizálása. A távolság kiszámításakor lehetőség van arra is, hogy a program a változókat különböző mértékben súlyozza.

A program először is meghatározza, hogy az adott mintában az adott feltételek mellett minden személynek hány szomszédja van, majd ezeknek elkészíti a gyakorisági eloszlását. Ezután a program ezen eloszlás alapján kiválasztja a legtöbb szomszédal rendelkező személyeket úgy, hogy ezek száma összesen ne haladja meg a sűrűsödéspontok maximális számát, mely a feladatablakban állítható be (alapérték:

15). Ezen személyek adatsorát (adatvektorát) nevezzük sűrűsödéspontnak. Ha két sűrűsödéspont nagyon közel van egymáshoz (a sűrűsödéspontok összevonási küszöb szintén a feladatablakban állítható be), akkor a program összevonja őket. Az új sűrűsödéspont ezek átlaga (átlagvektora) lesz. Ha kijelölünk egy kontroll klaszterváltozót, akkor a program megadja, hogy az ezen változó által meghatározott különböző klaszterekbe az azonosított sűrűsödéspontok szomszédsága milyen mértékben esik bele. Ezzel a más elemzésekben létrehozott klaszterek altípusai (prototípusai) azonosíthatók. Megjegyezzük, hogy a sűrűsödéseket a többdimenziós skálázás módszerével is lehet attraktívan ábrázolni (vö. *Takács* [2013]). A sűrűsödéspont-elemzéssel kapcsolatban lásd még *Vargha* [2008].

5.10. Időszeparáció

Ezen modul segítségével az ismételt mérések változóegyüttese feldarabolható az időpont alapján külön rekordokra, mintha azok független megfigyelések lennének. Az egy időpontban felvett adatok ekkor külön rekordba, vagyis az adattáblázat külön soraiba kerülnek (külön esetként lesznek számon tartva). Az elemzés végrehajtásához első lépésben az időpontok számát (T) és az egy időponthoz tartozó változók számát (V) kell megadni. Ezután minden időponthoz megjelenik egy változóablak, amelybe be kell tenni az idetartozó változókat. Ez háromféleképpen hajtható végre.

1. Ha az átteendő változók (összesen $T \cdot V$ darab változó) a listában olyan sorrendben helyezkednek el, hogy időpontonként ugyanolyan sorrendben egymás mellett vannak a megfelelő változók (például testsúly, testmagasság, fejkerület és lábméret 6 éves korban, 8 éves korban, 10 éves korban stb.), akkor ezeket együtt mind kijelölve a „Változók együtt” feliratú ikonablakra kell kattintani ahhoz, hogy mind a helyükre kerüljenek.

2. Ha az átteendő változók a listában olyan sorrendben helyezkednek el, hogy egymás mellett vannak egy-egy változó különböző időpontokhoz tartozó mérései (például testsúly 6, 8 és 10 éves korban, testmagasság 6, 8 és 10 éves korban, fejkerület 6, 8 és 10 éves korban stb.), akkor ezeket együtt mind kijelölve az „Időpontok együtt” feliratú ikonablakra kell kattintani ahhoz, hogy minden kijelölt változó a helyére kerüljön.

3. Minden más esetben a változókat egyenként vagy kisebb csoportokban kell a megfelelő időpont ablakába betenni, a következő szabályok figyelembevételével.

– Először rá kell kattintani arra az időpontablakra, ahova a kijelölt változót (változókat) át akarjuk tenni. E kattintásra az adott időpontablak sárga színnel kijelölődik.

- Ezután ki lehet jelölni maximálisan V számú változót.
- Végül rá kell kattintani az „Egy időpont egyszerre” feliratú ikonablakra.

5.11. Időegyesítés

E modul segítségével az időszeparációval létrehozott új fájlban végzett elemzések új változói (például klaszterváltozók) visszakerülnek egy eredetivel megegyező rekordstruktúrájú fájlba. Ezzel és az előző modullal könnyen végrehajtható az ISOA komplex klasszifikációs elemzés (vö. *Bergman–Nurmi–von Eye* [2012]).

6. A ROPstat használatának szemléltetése egy konkrét példán

A ROPstat egyszerűségét és felhasználóbarát voltát az 1. táblázatban látható hipotetikus adatállomány (20 tanuló 8. osztályos év végi jegye három tantárgyból) mintáján végrehajtott elemzések segítségével szemléltetjük. A ROPstat demó verziójának letöltése és installálása után nyissunk meg a programban egy új adatfájlt 5 változóval (oszloppal) és 20 esettel (sorral). Az adatok (a változónevek nélkül) közvetlenül is beolvashatók egy ilyen táblázatból (a jól ismert másol/beilleszt művelettel) a ROPstatba. Egy másik lehetőség: másoljuk be a teljes 1. táblázatot (változónevekkel együtt) Excelbe, mentjük el innen tabulátorral tagolt típusú szövegfájlba (természetesen csak az adatokat és az első sorban elhelyezett változóneveket), majd nyissuk meg ezt a ROPstatban. Mindkét esetben a változók jellemzőinek (rövid és hosszú nevek, csoportdefiniálások, hiányzó adatkódok stb.) megadása, illetve szerkesztése a ROPstatban a „Változók deklarációi” ablakban lehetséges.

Szakmai problémaként nézzük meg ennek a mintának a segítségével, hogy van-e különbség a nemek között a három tantárgy év végi jegye tekintetében. Ehhez a ROPstatban a „Statisztikai elemzések” menürendszerében válasszuk ki a „Csoportok és változók összehasonlítása” sort, majd ezen belül a „Független minták (csoportok) egyszempontos összehasonlítása” menüpontot. A megjelenő feladatablakban tegyük a Nem változót a „Csoportosító változó” ablakába, a többi változót (Magyar nyelv és irodalom, Matematika, Fizika) pedig a „Függő változók” ablakába. Változtatlanul hagyva a függő változók skálatípusát (intervallum), ha a „Futtat” ikonra kattintunk, a nem összehasonlításokkal kapcsolatos elemzések eredményei pillanatok alatt megjelennek az eredménylista ablakában.

1. táblázat

Egy egyszerű hipotetikus adatállomány 5 változóval és 20 esettel
(8. osztályos tanulók év végi jegyei)

Kódszám	Nem (1: fiú, 2: lány)	Magyar nyelv és irodalom	Matematika	Fizika
1	1	1	2	3
2	1	3	4	4
3	1	4	4	4
4	1	3	4	3
5	1	3	3	4
6	1	3	4	4
7	1	2	4	4
8	1	3	3	3
9	1	3	3	4
10	2	2	2	1
11	2	5	4	4
12	2	5	4	5
13	2	3	3	3
14	2	4	3	3
15	2	5	4	3
16	2	3	3	3
17	2	4	3	3
18	2	3	2	2
19	2	5	5	4
20	2	4	3	3

A ROPstat egyik kedvező tulajdonsága, hogy az „Excelbe átesz” ikonra kattintva a program átküldi a kapott eredménylistát egy Excel fájlba, és egyben meg is nyitja azt. Az eredmények lényegében ugyanazok, mint amiket más jól ismert programcsomagban is megkaphatunk (csoportok leíró statisztikái, normalitásvizsgálat, szóráshomogenitás tesztelése, Cohen-féle d hatásmérték és eta-négyzet megmagyarázott varianciaarány, kétmintás t -próba és Welch-féle d -próba eredménye). Emiatt itt most nem megyünk bele a részletekbe változónként, csak az eredménylista végén található összefoglaló táblázatot mutatjuk be. (Lásd a 2. táblázatot.) Ebből kiolvasható, hogy a két nem jegyeinek szintje csak a Magyar nyelv és irodalom esetében különbözik szignifikánsan ($p < 0,05$). Minthogy a szóráshomogenitást tesztelő Levene-próba messze nem szignifikáns ($p = 0,296$), most a kétmintás t -próba (t -érték) alapján dönthetünk. A szakmai különbség mértékét a Cohen-féle hatásmérték jelzi ($d = -1,184$), mely a Cohen által javasolt 0,80-as küszöböt meghaladva a lányok jelentős fölényét jelzi a fiúkéval szemben (vö. Cohen [1992] 1. táblázat).

2. táblázat

Összefoglaló táblázat a ROPstat „Független minták (csoportok) egyszempontos összehasonlítása” moduljából intervallum skálájú függő változók esetén (részlet)

Változó	Átlag1	Szórás1	Átlag2	Szórás2	Cohen_d	p/Levene	t-érték	p/t	Welch_d	p/d
Magyar nyelv és irodalom	2,78	0,83	3,91	1,04	-1,184	0,296	-2,632	0,017*	-2,694	0,015*
Matematika	3,44	0,73	3,27	0,91	0,207	0,623	0,46	0,651	0,471	0,643
Fizika	3,67	0,50	3,09	1,04	0,682	0,347	1,513	0,148	1,616	0,127

* $p < 0,05$.

3. táblázat

Részlet a ROPstat „Független minták (csoportok) egyszempontos összehasonlítása” moduljából ordinális skálájú függő változó esetén
(Függő változó: Magyar nyelv és irodalom)

Index	Nem	Esetek	Rangátlag	Rangszórás	Minimum	Maximum
1	fiú	9	7,333	3,832	1	4
2	lány	11	13,090	5,594	2	5

A 3. táblázatban látható adatokat az eredménylistán a következő tájékoztató szöveg egészíti ki:⁶

Elméleti rangszórások egyenlőségének tesztelése

– O’Brien-próba (Welch-féle): $F(1,0; 17,6) = 1,551$ ($p = 0,2293$)

– Levene-próba (Welch-féle): $F(1,0; 16,8) = 3,298$ ($p = 0,0872$)*

Sztocasztikus egyenlőség tesztelése

Hagyományos eljárás, amely feltételezi a szóráshomogenitást:

– Mann–Whitney-próba (egzakt): $R1 = 66,0$, $R2 = 144,0$

($p = 0,028$)**

– Mann–Whitney-próba (normális közelítés): $Z = -2,290$

($p = 0,022$)**

Szóráshomogenitást nem igénylő robusztus közelítő eljárások:

– Fligner–Policello-próba Welch-féle szabadságfokkal:

FPW(16,8) = -2,844 ($p = 0,0113$)**

⁶ * $p < 0,10$; ** $p < 0,05$; *** $p < 0,01$.

- Brunner–Munzel-próba: $BM(15,2) = -2,959$ ($p = 0,0097$)***
- Pont- és intervallumbecslés a valószínűségi fölény A mutatójára:
 - $A_{12} = 0,212$ [$P(\text{Csop1} > \text{Csop2}) = 0,101$,
 - $P(\text{Csop1} < \text{Csop2}) = 0,677$]
 - $C(0,95) = (0,005; 0,419)$

Tekintve, hogy egy iskolai osztályzat inkább tekinthető ordinális skálájúnak, semmint igazi kvantitatív változónak, az egyetlen szignifikáns változóra (Magyar nyelv és irodalom) elvégeztük a két minta rangsorolós összehasonlítását is. Ha ugyanezen modul feladatablakában egy függő változó skálátípusát *intervallumról* átváltjuk *ordinálisra*, a tesztelendő nullhipotézis az elméleti átlagok egyenlősége helyett a sztochasztikus egyenlőség lesz, mely ekvivalens az elméleti rangátlagok egyenlőségével. E nullhipotézis itt a Mann–Whitney rangsorolós próbával és két robusztus alternatív eljárással tesztelhető (lásd Vargha [2007] 10. fejezet). Az eredménylista (lásd a 3. táblázatot) ebben az esetben tartalmazza a csoportok rangstatisztikáit, a szóráshomogenitás tesztelését, a sztochasztikus egyenlőség tesztelésének három próbáját, valamint egy ordinális hatásmértéket, a valószínűségi fölény A mutatóját (vö. Vargha [2007] 10.4. alfejezet). A 3. táblázatból kiolvashatjuk, hogy a két nem az ordinális összehasonlítás során is szignifikánsan különbözik (minden próbával ugyanolyan, $p < 0,05$ szinten). A valószínűségi fölény $A = A_{12} = 0,212$ értéke, minthogy 0,50-nél kisebb, az 1. csoport (azaz a fiúk) hátrányát jelzi a „magyar” osztályzat tekintetében a 2. csoporttal (azaz a lányokkal) szemben. A lányok dominanciaértéke a fiúkkal szemben: $A_{21} = 1 - A_{12} = 1 - 0,212 = 0,788$. Ezt az értéket úgy értelmezhetjük, hogy egy-egy véletlenszerűen kiválasztott fiú és lány esetén a lány 0,788 valószínűséggel kap nagyobb jegyet magyarból, mint egy fiú, ha az egyenlőség esélyét 0,5-ös szorzóval számítjuk be. Az $A = 0,788$ érték szakmailag jelentős különbségnek tekinthető a két csoport között (vö. Vargha [2007] 10.3. táblázat). A „tiszta” fiú és lány fölény a 3. táblázat aljáról olvasható ki. Eszerint a fiú-lány összehasonlítások során a fiúk csak az esetek 10,1 százalékában, míg a lányok 67,7 százalékban kaptak jobb magyar jegyet. Ez az elemzés tehát nemcsak azt az általános trendet jelzi, hogy a lányok magyar osztályzata általában jobb, mint a fiúké, hanem azt is, hogy ez az általános trend milyen gyakran teljesül, illetve sérül.

7. Összefoglalás

Cikkünkben részletesen ismertettük a ROPstat statisztikai programcsomag főbb vonásait és menürendszerét, majd egy konkrét példán szemléltettük a program használá-

tának egyszerűségét. Bár korábbi publikációk írtak a ROPstat egyes eljárásairól (például magyar nyelven Vargha [2008], angol nyelven Vargha–Torma–Bergman [2015]), ez az első olyan publikáció, amely a ROPstatot és menürendszerét teljes áttekintésben mutatja be. Vargha [2008] az egy- és a kétszemponos sztochasztikus összehasonlítás, az osztópont-, valamint a sűrűsödéspont-elemzés ROPstat-beli alkalmazását részletezi, Vargha–Torma–Bergman [2015] pedig a mintázatfeltáró modulokat tekinti át. A ROPstatban futtatható egyváltozós statisztikai elemzések technikai részleteivel kapcsolatban lásd Vargha [2007] írását. Jelen cikkünkben felhívtuk a figyelmet azokra a szoftverjellemzőkre és statisztikai eljárásokra (például robusztus technikákra, ordinális elemzésekre, mintázatfeltáró módszerekre) is, amelyek ebben a programban egyediék, más hasonló programcsomagokban nem találkozhatunk velük.

Megjegyezzük még, hogy a ROPstat honlapjáról (www.ropstat.com) elérhetők a program egyes moduljainak paraméterezését megkönnyítő YouTube-videók is, melyeket Takács Szabolcs készített.⁷

Irodalom

- BARTKO, J. J. [1976]: On various intraclass correlation reliability coefficients. *Psychological Bulletin*. Vol. 83. No. 5. pp. 762–765. <http://dx.doi.org/10.1037/0033-2909.83.5.762>
- BERGMAN, L. R. – MAGNUSSON, D. – EL-KHOURI, B. M. [2003]: *Studying individual development in an interindividual context. A person-oriented approach*. Lawrence-Erlbaum Associates. Mahwa, London.
- BERGMAN, L. R. – NURMI, J.-E. – VON EYE, A. A. [2012]: I-states-as-objects-analysis (ISOA): Extensions of an approach to studying short-term developmental processes by analyzing typical patterns. *International Journal of Behavioral Development*. Vol. 36. No. 3. pp. 237–246. <http://dx.doi.org/10.1177/0165025412440947>
- COHEN, J. [1992]: A power primer. *Psychological Bulletin*. Vol. 112. No. 1. pp. 155–159. <http://dx.doi.org/10.1037/0033-2909.112.1.155>
- NAGYBÁNYAI N. O. [2006]: A pszichológiai tesztek reliabilitása. In: Rózsa S. – Nagybányai Nagy O. – Oláh A. (szerk.): *A pszichológiai mérés alapjai. Elmélet, módszer és gyakorlati alkalmazás*. Eötvös Loránd Tudományegyetem. Budapest.
- SZÉKELYI M. – BARNA I. [2003]: *Túlélőkészlet az SPSS-hez*. TypoTex Kiadó. Budapest.
- TAKÁCS SZ. [2013]: Többdimenziós skálázás. *Psychologia Hungarica Caroliensis*. 1. évf. 1. sz. 140–151. old. <http://dx.doi.org/10.12663/PSYHUNG.1.2013.1.7>
- TAKÁCS SZ. – MAKRAI B. – VARGHA A. [2015]: Klasszifikációs módszerek mutatói. *Psychologia Hungarica Caroliensis*. 3. évf. 1. sz. 67–88. old.
- VARGHA A. [2001]: Kísérleti helyzetek és csoportok összehasonlítása új statisztikai módszerekkel. In: Pléh Cs. – László J. – Oláh A. (szerk.): *Tanulás, kezdeményezés, alkotás*. ELTE Eötvös Kiadó. Budapest. 371–386. old.

⁷ Lásd közvetlenül https://www.youtube.com/results?search_query=ropstat&sm=12

- VARGHA A. [2002]: Független minták egyszempontos összehasonlítása új rangsorolási 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. [2005]: Sokaságok összehasonlítása új módszerekkel. *Statisztikai Szemle*, 83. évf. 5. sz. 429–448. old.
- VARGHA A. [2007]: *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. Pólya Kiadó. Budapest.
- VARGHA A. [2008]: Új statisztikai módszerekkel új lehetőségek: a ROPstat a pszichológiai kutatószolgálatában. *Pszichológia*. 28. évf. 1. sz. pp. 81–103. <http://dx.doi.org/10.1556/Pszi.28.2008.1.5>
- VARGHA, A. – BERGMAN, L. R. [2012]: A method to maximize the information of a continuous variable in relation to a dichotomous grouping variable: Cutpoint Analysis. *Hungarian Statistical Review*. Vol. 90. Special Number 16. pp. 101–122. http://www.ksh.hu/statszemle_archive/2012/2012_K16/2012_K16_001.pdf
- VARGHA, A. – BERGMAN, L. R. – TAKÁCS, S. [2016]: Performing cluster analysis within a person-oriented context: Some methods for evaluating the quality of cluster solutions. *Journal for Person-Oriented Research*. Vol. 2. Nos. 1–2. pp. 78–86. <http://dx.doi.org/10.17505/jpor.2016.08>
- VARGHA A. – BORBÉLY A. [2016]: Modern mintázathatózó módszerek alkalmazása a kétnyelvűség kutatásában. In: *Kissné Viszket M. – Puskás-Vajda Zs. – Rácz J. – Tóth V. (szerk.): A pszichológiai tanácsadás perspektívái. Tisztelgő kötet Ritoók Magda 80. születésnapjára*. L'Harmattan. Budapest. 173–186. old.
- VARGHA, A. – TORMA, B. – BERGMAN, L. R. [2015]: ROPstat: A general statistical package useful for conducting person-oriented analyses. *Journal for Person-Oriented Research*. Vol. 1. Nos. 1–2. pp. 87–98. <http://dx.doi.org/10.17505/jpor.2015.09>
- VON EYE, A. – MAIR, P. – MUN, E.-Y. [2010]: *Advances in Configural Frequency Analysis*. Guilford Press. New York.

Summary

ROPstat is a general, mostly univariate statistical program package that offers specialties in three different domains: 1. robust techniques, 2. ordinal analyses, and 3. pattern-oriented methods. Many of these are not available in other common statistical software packages. In the present paper, the main features and menu structure of ROPstat are outlined, introducing numerous statistical methods available in the modules. In the last section, an easy-to-use feature of ROPstat is presented by an example.