

A

ROPstat

statisztikai
menürendszer

Vargha András

2016

Tartalomjegyzék

Bevezetés: A ROPstat általános jellemzői

- Főbb vonások
- A ROPstat installálása
- Adatimport és adatexport a ROPstatban

1. Alapok (egymintás elemzések)

- Alapstatisztikák
- Részletesebb statisztikák számítása
- Gyakorisági eloszlás, hisztogram készítése
- Középértékekre vonatkozó hipotézisek vizsgálata
- Normalitásvizsgálat

2. Csoportok és változók összehasonlítása

- Független minták egyszempontos összehasonlítása
- Összetartozó minták egyszempontos összehasonlítása
- Átlagok kétszempontos összehasonlítása
- Kétszempontos vegyes varianciaanalízis
- Kétszempontos rang-varianciaanalízis
- Inter-rater reliabilitás kvantitatív skálákkal

3. Változók kapcsolatának vizsgálata

- Korreláció, egyszerű regresszió
- Korrelációs és parciális korrelációs mátrixok készítése
- Diszkrét változók kapcsolatvizsgálata
- Itemanalízis
- Többszörös lineáris regresszió

4. Mintázatfeltáró elemzések

- Adatleírás

- Hiányzó adatok pótlása
- Reziduálanalízis
- Hierarchikus klaszteranalízis
- Relokáció: nemhierarchikus K-központú klaszteranalízis
- Konfigurációanalízis (CFA)
- ExaCon
- Centroid
- DensePoint: sűsűsödéspont-elemzés
- Időszeparáció
- Időegyesítés

Bevezetés: A ROPstat általános jellemzői

Főbb vonások

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 jelen cikk szerzője az elmúlt évtizedekben mindhárom területen számos újdonsággal gazdagította a statisztikai adatfeldolgozás eszköztárát (pl. *Vargha* [1988], [1996], [2002], [2004], [2005], [2011], *Vargha–Delaney* [1998], *Vargha–Torma–Bergman* [2015], *Vargha–Bergman–Takács* [2016]). A nem statisztikus felhasználók azonban nem tudják felhasználni a legjobb, sőt legegyszerűbb módszereket sem, ha nem áll rendelkezésükre megfelelő statisztikai szoftver, mégpedig oly módon, hogy az új módszerek a régi, hagyományosan használtakkal együtt, közös programcsomagban legyenek elérhetők. Egy másik motivációja új statisztikai programcsomag létrehozásának az egyetemi oktatás volt, ahol hiánycikk volt a magyar nyelven kommunikáló, az egyváltozós statisztikát teljesen lefedő, széles spektrumú statisztikai programcsomag a pszichológia és a társadalomtudományok szakjain.

Első próbálkozásként a PSZICHOSTAT statisztikai programcsomagot alkottuk meg Spektrum és Commodore-64 számítógépekre a 80-as évek közepén (*Vargha–Izsó*, 1989). Ahogy fejlődtek az asztali számítógépek, úgy nőtt meg az igény a fejlettebb szoftverekre is. Így készült el a 90-es évek közepére a Microsoft DOS operációs rendszere alatt futtatható MiniStat (*Vargha* [1999], [2000], *Vargha–Czigler* [1999]). Tíz évvel később, a Windows rendszerhez való igazodásként született meg a MiniStat Windows változata, a ROPstat, mely amellet, hogy elődje minden kedvező tulajdonságát megőrizte, számos új funkcióval és statisztikai elemzéstípussal bővítette palettáját.

A ROPstat jelenlegi változata két (magyar és angol) nyelven kommunikáló szoftver, melyet a jelen cikk szerzője tervezett és írta különböző statisztikai moduljainak a programjait (Delphi programozási nyelven). Bánsági Péter matematikus mérnöknek köszönhető a programok Windows rendszerhez igazítása és a teljes képernyő-kezelés. Lars Bergman szakmai konzultánsként segített a mintázatfeltáró modulok többségének a kidolgozásában.

A ROPstat teljesen menüvezérelt és használata nem igényel programozói ismeretet. A programcsomag használatának egyik fontos feltétele, hogy a felhasználó ismerje változóinak

mérési szintjét (pszichometriai skálatípusát), valamint legyen tudatában annak, hogy milyen szakmai kérdésre szeretne adatai felhasználásával választ kapni. A ROPstat moduljainak közös jellemzője, hogy a statisztikák kiszámíthatók külön, egy ún. feltételes csoportosító változó meghatározott szintjeire is (pl. külön férfiakra és nőkre, vagy csak azokra a személyekre, akiknek 120 és 150 közötti az IQ-ja).

A ROPstatot az SPSS-sel összevetve azt mondhatjuk, hogy az SPSS teljesebb körben tartalmazza a többváltozós statisztikai eljárásokat. Ezek köréből a ROPstatban csak a többszörös lineáris regressziót, a kétszemponos varianciaanalízist és a személyek különböző klasszifikációs elemzéseit (pl. a személyek klaszteranalízisét) találjuk meg. Ugyanakkor a klasszifikációs elemzéseket is tartalmazó mintázatfeltáró elemzésekből a ROPstat gazdagabb választékot kínál a felhasználó számára, több olyan elemzéssel (pl. konfigurációelemzés, sűrűsödéspont-elemzés, centroidok összehasonlítása stb.), melyek az SPSS-ben nem elérhetők. A ROPstat gazdagabb a robusztus és a nemparaméteres módszerekben is, melyekre részletekbe menően majd később fogunk kitérni, viszont nem találhatók meg benne olyan speciális elemzések, mint az idősoelemzések (Time Series Analyses), túlélőgörbék készítése (Survival Analysis), ROC-görbe elemzés (ROC Curve Analysis) vagy minőség-ellenőrzés (Quality Control). A ROPstatban egyébként az SPSS-hez hasonlóan egy adathalmazt mintának és nem sokaságnak (teljes populációnak) tekintünk. Végül a ROPstat és az SPSS kapcsolatáról megjegyezzük, hogy az esetleges programozói hibák kiszűrése céljából a ROPstat minden olyan eljárását, amely az SPSS-ben is megtalálható (ilyen pl. az összes hagyományos egyváltozós statisztikai elemzés, a többszörös regresszió-elemzés vagy a kétszemponos varianciaanalízis), a programfejlesztés során konkrét példák segítségével az SPSS-sel összevetettük, hogy ugyanarra a numerikus eredményre vezetnek-e, s ahol kellett, javítottunk.

Hasonlókat mondhatunk a ROPstat és az Excel összehasonlítása terén is. Mindamellet az Excel használata kitüntetett a ROPstatban, ugyanis ez utóbbiban bármely statisztikai elemzés teljes eredménylistája egy gombnyomásra átküldhető az Excelbe, aminek két fontos előnye van. Az egyik az, hogy az Excelben megjelenített táblázatok ugyanilyen táblázatos formában átmásolhatók Word dokumentumokba, ami műhelymunkák, szakdolgozatok, publikálandó tanulmányok esetében igen hasznos. A másik jó tulajdonsága ennek a lehetőségnek az, hogy az Excelben a megjelenített ROPstat-táblázatokból igen könnyen lehet tetszetős ábrákat, grafikonokat szerkeszteni. Maga az adatátvitel (adat export-import) egyébként igen egyszerű a ROPstat, valamint az SPSS és az Excel között, melynek technikáját az alábbiakban majd kissé részletezzük is.

A ROPstat készítésekor elsősorban a pszichológiai oktatás és kutatás szempontjait vettük figyelembe, de ezt a szoftvert jól használhatják társadalomtudósok (nyelvészek, szociológusok, pedagógusok, közgazdászok), de más szakterületek képviselői (például orvosok, biológusok stb.) is.

A ROPstat installálása

A ROPstat installálásának első lépéseként a szoftver demó verzióját kell letölteni a ROPstat saját honlapjáról (www.ropstat.com). Ennek 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 1000 különböző változó és esetenként 100000 személy statisztikai elemzésére. A normál és profi verziók elérhetőségéről minden információ megtalálható a szoftver honlapján.

Adatimport és adatexport a ROPstatban

A ROPstat saját típusú és kiterjesztésű – msw – adatfájlokkal dolgozik, ilyeneket hoz létre és olvas be. Egy msw egyébként egy 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 előtt¹. 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 (pl. férfi és nő) alkalmazása esetén a „Változók deklarációi” ablak „Típus” oszlopában. Ugyanebben az ablakban szerkeszthetők a változók rövid (maximum 8 karaktert

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.

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 adatfájl SPSS-be vagy Excelbe szeretnénk átkonvertálni, akkor a Fájl menü 'Mentés SPSS formátumban', illetve 'Mentés tabulált szövegfájl formátumban' menüpontját kell használnunk.

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

Alapstatisztikák

Ez a modul változónként alapstatisztikákat számít. A statisztikák kiszámíthatók külön, egy feltételes csoportosító változó meghatározott szintjeire (pl. külön férfiakra és nőkre, vagy csak azokra a személyekre, akiknek 120 és 150 közötti az IQ-ja). 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).

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 eloszlásától.

Még részletesebb lesz az eredménylista, ha trimmelést (nyesést) 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.

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. Ezen változótípus esetén lehetőség van minden kiválasztott X változóból speciális új változók képzésére és elmentésére:

(i) 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 $X \neq c$.

(ii) 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.

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

- Csoport. def. változótípus beállítása esetén a program kilistázza az illető változó különböző csoportjait, azok gyakoriságát, továbbá külön kérésre a khi-négyzet-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 (vö. Vargha [2007] 17.1. alfejezet).

Közéértékekre vonatkozó hipotézisek vizsgálata

Ez a modul minden kijelölt változó esetében végrehajtja az egymintás t -próbát, annak két robusztus változatát (Johnson- és Gayen-próba), külön kérésre a trimmelt t -próbát, valamint a Wilcoxon-próbát 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: $P(X > A) = P(X < A)$. Eredménylista: átlag, szórás, minimum és maximum; egymintás t -próba; Johnson- és Gayen-próba (csak 500-at meg nem haladó mintanagyság és $|t| \leq 10$ esetén); trimmelt t -próba; Wilcoxon-próba; előjelpróba. 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.

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). Megjegyezzük, hogy a normalitás tesztelésére a legtöbb esetben érzékenyebb és hatékonyabb eljárás a ferdeségi és a csúcsossági együttható segítségével végzett elemzés a 'Részletesebb statisztikák számítása' modulban.

Ebben a menüpontban lehetőség van lineáris transzformációval minden kiválasztott változó két speciális standardizáltjának képzésére és elmentésére:

- (i) Egyszerű standardizálás: ebben az esetben az új változó átlaga 0, szórása 1 lesz.
- (ii) 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.

2. Csoportok és változók összehasonlítása

Független minták egyszempontos összehasonlítása

Intervallum-skálájúnak beállított változókra a program az elméleti szórások és átlagok egyenlőségét vizsgálja. A populációk (sokaságok) szórásainak összehasonlítására az O'Brien és a Levene-próba (ezeknek is a robusztus, Welch-féle változata) áll rendelkezésre, 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 csoport esetén pedig az egyszempontos varianciaanalízis (VA) független mintás eljárása. Ennek az elemzésnek az eredményét – különösen számottevően eltérő elemszámok esetén – csak akkor tekinthetjük érvényesnek, ha a szórások nem különböznek túlságosan nagy mértékben 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 csoport esetén pedig a Welch-, a James-, és a Brown-Forsythe-féle robusztus VA. Ezek közül 20-nál kisebb minták esetén a Welch-, nagyobb minták esetén pedig a James-próba tűnik a legmegfelelőbbnek. A Brown-Forsythe-féle eljárás csak legfeljebb mérsékelten különböző szórások esetén megbízható (vö. *Vargha* [2007] 13.2. alfejezet).

Ha kettőnél több csoport összehasonlítása esetén 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 módszer, 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 (outlier) adatok is vannak, célszerű a minták szélsőséges elemeit trimmeléssel

„levágni” (trimmelni) a mintákról. Ezt 0-tól különböző trimmelési százalék beállításával tehetjük meg. Ilyenkor a program kettőnél több csoport összehasonlítása esetén trimmelt VA-t, két csoport összehasonlítása esetén 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 közti korrelációt a program csoportonként kiszámítja és összehasonlítja.

Ordinális skálájúnak beállított változók esetén a program az elméleti rangszórások megegyezését és a sztochasztikus homogenitást teszteli. Populációk (sokaságok) összehasonlítása esetén 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 köznap 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 összpulációbeli 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. Szignifikánsan 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 minták esetén pedig a normális eloszlással való közelítés segítségével állapítja meg.

Ha kettőnél több minta összehasonlítása esetén 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 fölény mutatónak, valamint a $P(\text{Csop1} > \text{Csop2})$, $P(\text{Csop1} < \text{Csop2})$ valószínűségeknek a pontbecslését, illetve $A(1, 2)$ -re 95%-os intervallumbecslést is készít.

Ha a feladatablakban kérjük az eloszlások részletes összehasonlítását, akkor a program osztópontelemzést is végrehajt, vagyis a függő változó értékskálájának maximum 10 pontjában (ún. binarizáló osztópontokban) 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 ki. Az osztópontelemzéssel sok esetben érzékenyebben lehet például két eloszlást összehasonlítani, mint az erre a célra hagyományosan alkalmazott Kolmogorov-Szmirnov-próbával (vö. *Vargha* [2005], [2008] és *Vargha–Bergman* [2012]), melyet a ROPstat ebben az esetben szintén elvégez.

A ROPstat különleges vonása, hogy ha az összehasonlítandó csoportok száma 2 és 10 közé esik és a kiválasztott függő változók száma legalább kettő, akkor intervallum és ordinális típusú változók esetén a program az eredménylista végén közös tömör táblázatban foglalja össze az összehasonlítások eredményeit (alapstatisztikák, hatásmértékek, próbastatisztikák, p -értékek).

Nominális skálájúnak beállított változók esetében 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 esetén Fisher-egzakt-próbával) teszteli, hogy X függ-e G -től (vö. *Vargha* [2007] 17. fejezet).

Összetartozó minták egyszempontos összehasonlítása

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

Intervallum skálájú változókra a program kettőnél több függő változó esetén az egyszempontos összetartozó mintás varianciaanalízis (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. Tekintettel arra, hogy az összetartozó mintás VA szóráshomogenitási feltétele igen 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 kelletténél ritkábban szignifikáns), a Huynh-Feldt-féle pedig általában enyhén liberális (a kelletténél gyakrabban szignifikáns).

A program a Friedman-próba G mennyiségének szignifikanciáját a χ^2 -eloszlással való közelítéssel 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 $(A(k, u) - 0,5)^2$, s ezen eltérések szignifikanciáját is megvizsgálja.

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

² $A(k, u)$ 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.

Ha a normalitás erősen sérül amiatt, hogy az adatmintában a többi közül nagyon kilógó extrém (outlier) adatok is vannak, célszerű a változók szélsőséges elemeit trimmeléssel „lenyesni” (trimmelni). Ilyenkor a program kettőnél több változó összehasonlítása esetén trimmelt VA-t, két változó összehasonlítása esetén 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). E sztochasztikus összehasonlítások elméleti értelmezése ugyanúgy történik, mint a független mintás esetben.

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 változó esetén 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 a Cochran-féle Q-próbával mind az öt értékre vonatkozóan teszteli, hogy az érték ugyanolyan valószínűséggel fordul-e elő a három függő változónál.

Átlagok kétszemponos összehasonlítása

Ezzel a modullal független minták kétszemponos varianciaanalízise (VA) végezhető el a súlyozatlan átlagok módszerével (vö. *Vargha* [2007] 16.1. és 16.2. alfejezet). A mintaelemszámok eltérőek is lehetnek a csoportosító faktorok különböző kombinációiban. A csoportosító faktoroknak keresztezetteknek kell lenniük (és nem beágyazottnak). A program a VA szóráshomogenitási feltételének fennállását a Levene-próba robusztus Welch-féle változatával teszteli. Ha ez a próba szignifikáns, 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 jelenléte esetén 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 számít ki alternatív érvényes VA-t, ha ez a feltétel sérül.

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 eltérőek is lehetnek.

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

Kétszemponos rang-varianciaanalízis

Ezzel a modullal független minták kétszemponos sztochasztikus homogenitás-vizsgálata (vö. *Vargha* [2004]) végezhető el. Egy (i, j) alpopuláció (részsokaság) P_{ij} sztochasztikus dominancia értéke 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:

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

A sztochasztikus dominancia érték az X függő változó nagyság szintjét méri a két szempont (jelölje őket mondjuk A és B) szintjeinek minden kombinációjára. Az A szempont sztochasztikus homogenitása (az A szempont szintjeinek számát a -val jelölve) 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 dominancia értékek 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, pedig a pszichológiában és a társadalomtudományokban szép számmal fordulnak elő ordinális skálájú változók, melyek kétszemponos függésvizsgálata sokszor fontos kutatási kérdés.

Inter-rater reliabilitás vizsgálat kvantitatív skálákkal

Ezzel a modullal kvantitatív skálák inter-rater reliabilitása (a bírálói egyetértés mértéke) vizsgálható az ICC intraklaszter (intraclass) korrelációs együttható segítségével (vö. *Bartko* [1976] vagy *Landers* [2015]). Az ICC segítségével azt lehet tesztelni, hogy az azonos személyek különböző bírálóktól kapott minősítései között nagyobb hasonlóság van-e, mint a különböző személyek értékátlagai között. A bírálói egyetértés akkor tekinthető elfogadhatónak, ha ICC nagysága eléri a 0,70-es szintet. Sokak (pl. *Bartko* [1976]) szerint ICC helyett annak korrigált értéke, ICC(U) a megbízhatóbb.

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 vizsgált személyek azonosító kódját is (esetazonosító változó). A bírálók azonosítóját nem kötelező megadni, de a hiteles tájékoztatás szempontjából célszerű akár numerikus, akár szöveges formában feltüntetni.

3. Változók kapcsolatának 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. Ha 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 közti kapcsolatot vizsgálja.

Korreláció, egyszerű regresszió

Ez a modul változópárookra számítja ki az egyszerű lineáris regressziós egyenletet és a két változó közti 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 Wilcoxon-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 kiszámításra kerül. Előbbi csak folytonos változók esetén javasolt mutató. Folytonos esetben tau és tau-b megegyezik. A Kendall-féle tau, illetve tau-b mutató kiszámítása esetén a program mindig kiszámítja az adott mintában a pozitív kapcsolati arányt (konkordancia %), illetve a negatív kapcsolati arányt (diszkordancia %). A program az összes kiszámított kapcsolati mutató elméleti értékeire megad (beállítható szintű) konfidencia-intervallumot is.

A program a fentiekén 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.

E modul feladatablakában lehetőség van egy rövid korrelációs output kérésére is. Ezt kijelölve a program minden kijelölt X változó esetén egymás alatt listázza ki X -nek a kijelölt Y változókkal való legfontosabb korrelációit, a hagyományos Pearson-féle r , a Wilcoxon-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 % és a diszkordancia % értékét.

Korrelációs és parciális korrelációs mátrixok készítése

Ezzel a modullal a kijelölt változópárok korrelációját és parciális korrelációját számíthatjuk ki három különböző korrelációs együttható típusra (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 változó esetén.

A korrelációkat (parciális korrelációkat) a program alapesetben az X és az Y változók között számítja ki (tehát X -eken, illetve Y -okon belül nem). Viszont ha csak X vagy csak Y változók vannak kijelölve, a program az e csoporton belüli összes párra vizsgálja a változók kapcsolatát, valamint a Pearson-korreláció típus kiválasztása esetén elvégzi a Bartlett-féle függetlenségvizsgálatot. Ez azt teszteli, hogy van-e a véletlennél szorosabb lineáris kapcsolat a kijelölt változók csoportjában legalább két változó között.

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

együtthatókat is. A program külön kérésre nagyság szerinti sorba is rendezi a kiszámított korrelációkat.

Diszkrét változók kapcsolatvizsgálata

Ez a modul diszkrét (pl. 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-egzakt-próbával) és mérhető a változók különféle típusú kapcsolatának szorossága is. Ezen mutatók, valamint különféle speciális gyakorisági táblázatok (várt gyakoriságok, sor- és oszlopösszegek szerinti százalékok, összelemszám szerinti százalékok) kilistázása opcionálisan beállítható. „No” beállítás esetén sosincs kilistázás, „Yes” beállítás esetén mindig van, „If signif.” beállítás esetén pedig a program csak akkor listázza ki a szóban forgó táblázatot, illetve mutatókat, ha a függetlenség nullhipotézise elutasítható legalább 10%-os szinten.

A modulban az alábbi kapcsolati mutatók számíthatók ki (pontos definíciójukkal és értelmezésükkel kapcsolatban lásd Vargha [2007] 17.2.4. alpont).

- Bármilyen típusú diszkrét változók esetén a Cramér-féle V , mely 2×2 -es táblázatok esetén a ϕ kontingencia-együtthatóra egyszerűsödik.
- Ordinális skálájú változók esetén értelmes a monoton asszociációs mérőszámok (Kendall-féle tau és tau-b, Goodman-Kruskal-féle gamma, Somers-féle egyirányú monotonitási együtthatók) lekérése.
- Kétértékű, illetve két csoportra beállított változók (2×2 -es táblázatok) esetén az alfa esélyhányados.
- Tetszőleges diszkrét változók esetén a Goodman és Kruskal által javasolt tau és lambda predikciós mutatók.
- Közös információ-hányad.
- Ha a gyakorisági táblázat sorainak és oszlopainak a száma megegyezik, akkor a program opcionálisan Cohen-féle bírálói (interrater) kappa reliabilitást is számol.

Ebben a modulban lehetőség van arra is, hogy a program az elemzéseket egy beírt (bemásolt) gyakorisági táblázat alapján végezze el. Ha a táblázatnak J sora és K oszlopa van, akkor ehhez a következőket kell megtenni.

- Nyissunk meg egy új adatfájlt a ROPstatban J sorral (esettel) és $K+1$ oszloppal (változóval). Az első változót szöveges változóként definiáljuk és ennek oszlopába írjuk be az ehhez tartozó csoportneveket (a sorváltozó címkéit).

- A többi (K számú) változó legyen numerikus (ez az alapértelmezés), neveik legyenek az oszlopváltozó címkéi (az oszlopnevek).
- Végül a modul feladatablakában jelöljük be az „Adatmátrix gyakorisági táblázat” opciót, s indítsuk el az elemzést.

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 Cronbach-alfa azt méri, hogy a kijelölt változók mennyire húznak egy irányba, mennyire mérik tartalmilag ugyanazt. Hogy mit, az itt most nem érdekes, azt validitásvizsgálattal lehet kideríteni.

A skála tartalmazhat pozitív és negatív tételeket (itemeket) is, amelyeket külön ablakban kell kijelölni. Negatív tételek kijelölése esetén a tételek átfordításhoz meg kell adni a skálák közös elvi minimumát és maximumát. Ha ez nem azonos a különböző negatív tételek esetén, az elemzések bármely érték beállítása esetén 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 item-totál (IT), az item-maradék (IM) korrelációt, továbbá a Cronbach-alfa értékét abban az esetben, ha az illető itemet elhagynánk a skálából. IT az item és a teljes skála közti 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 (pl. az SPSS-ben) korrigált item-totál korrelációnak is nevezni. Az SPSS-ben egyébként a Scale statisztikai menüpont Reliability Analysis elemzésével lehet ugyanezeket a számításokat elvégezni.

Többszörös lineáris regresszió

A modulban több kvantitatív független változók (X -változók) segítségével elkészíthetjük egy kvantitatív függő változó (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, vagyis 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 közti 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, ami az SPSS-sel szemben itt alapértelmezés, 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 e lépéshez tartozó regressziós becslés hibavariációjáról (Res) és standard hibájáról (SH).
2. A második blokkban a program tájékoztat arról, 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 bekerült 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 korrelációt (R) és a determinációs együtthatót (R^2), a regressziós becslés hibavariációját és standard hibáját, valamint a lépésenkénti elemzés harmadik blokkjában közölt információt. Az R^2 többszörös korreláció négyzet (determinációs együttható) egyben az a varianciaarány, melyet az X független változók – lineáris predikcióval – magyaráznak az Y változó varianciájából. Mivel ez az adott konkrét mintára van optimalizálva, ez az elméleti R^2 értéknek kisebb-nagyobb mértékben (minél kisebb a minta, illetve minél több a független változó, annál inkább) torzított becslése. Az elméleti értékre jobb becslést ad a program által

szintén mindig kiszámított korrigált R^2 érték. R^2 egyébként a következőt jelenti: a többszörös lineáris regressziós becslés Y -ra vonatkozó hibavariációjára ilyen arányban kisebb, mint ha Y -t minden személy esetén az Y -átlaggal becsülnénk.

4. Mintázatfeltáró elemzések

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 elhagyása esetén mennyivel nő a minden változóra érvényes értékkel rendelkező 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 össze is hasonlítja az ugyanazon változópárokhoz tartozó korrelációs együtthatókat.

Hiányzó adatok pótlása

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

1. Adathelyettesítés az átlaggal. Ebben az esetben 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. Ebben az esetben 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 a megfelelő változóértékével helyettesíti. A hasonlóság mértéke lehet a szokásos euklideszi távolság, vagy az értékmintázatok Pearson-féle korrelációja. A feladatablakban kérhető a változók standardizálása, mely feltétlenül indokolt abban az esetben, 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 variancia-arányt is.

Reziduálanalízis

Ebben a modulban az adatállomány extrém, kilógó személyei (többdimenziós outlier adatsorok) 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 értékek esetén 0, a megmaradók esetében 1. Ezt a program – kérésre – hozzáilleszti az adattáblázathoz, s ennek felhasználásával a kilógó esetek más elemzésekből kihagyhatók. A programban kérhető sűrűsödési változók létrehozása is, melyek esetenként megadja – különböző sugarú szomszédsági köröket alkalmazva – az eset szomszédsági sűrűségét (sok közeli szomszéd esetén nagy, kevés szomszéd esetén kis értékkel).

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

Hierarchikus klaszteranalízis

Ez a modul hierarchikus klaszteranalízist végez a személyeken (eseteken) a kijelölt változók felhasználásával. A klaszteranalízis típusfeltáró többváltozós statisztikai eljárás, 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 egy 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), vagy az értékmintázatok Pearson-féle korrelációja (két személy esetén a kijelölt változók értékeire az összetartozó értékpárok közti Pearson-korreláció). A feladatablakban kérhető a távolságok kiszámítása esetén a változók standardizálása, mely feltétlenül indokolt abban az esetben, amikor 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). A lehetőségek az alábbiak:

- Átlagos távolság módszere: ekkor az a két klaszter kerül összevonásra, amely esetében a két klaszterbe tartozó személyek átlagos távolsága a legkisebb.
- Minimális távolság (legközelebbi szomszéd) módszere: ekkor az a két klaszter kerül összevonásra, amely esetében a legkisebb a két klaszter egymáshoz legközelebbi lévő két személyének a távolsága.
- Maximális távolság (legtávolabbi szomszéd) módszere: ekkor az a két klaszter kerül összevonásra, amely esetében a legkisebb a két klaszter egymástól legtávolabbi lévő két személyének a távolsága.
- Centroid módszer: ekkor az a két klaszter kerül összevonásra, amely esetében a két klaszter centroidjának (átlagvektorának) a távolsága a legkisebb.
- Medián módszer: ekkor az a két klaszter kerül összevonásra, amely esetében a két klaszterbe tartozó személyek távolságainak a mediánja a legkisebb.
- Ward-féle módszer (csak euklideszi távolsággal): ekkor az a két klaszter kerül összevonásra, amely esetében a két klaszter összevonásával a legkisebb mértékben nő a klaszterek összheterogenitása (ESS = error sum of square = négyzetes összhiba). Az összheterogenitást úgy határozzuk meg, hogy minden klaszter esetén kiszámítjuk a klaszterbe tartozó személyek klasztercentroidtól való távolságainak négyzetösszegét, majd ezeket összeadjuk. Röviden: ennél a módszernél minden lépésben azt a két klasztert vonjuk össze, amely esetében az ESS növekedés (ESSnövé.) a legkisebb.

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

4. A program minden lépésben feltünteti az összevonás lépését, az adott lépésben létező klaszterek számát (KL#), az ESS növekedést (ESSnövé.), a megmagyarázott varianciaarányt (EESS%), egy sor más klaszter adekvációs mutatót (PB, XBmod, SC, HCátlag, HCterjedelem; 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:

$$\text{EESS\%} = 100 \cdot (\text{ESSmax} - \text{ESS}) / \text{ESSmax}.$$

EESS% értéke függ az elemzésbe bevont változók számától (több változó esetén általában kisebb) és az összelemszámtól (nagyobb minta esetén általában kisebb). Sikeres osztályozás esetén EESS% 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 program listázza ki a klaszterek főbb jellemzőit (klaszternagyság, változónként átlag, szórás, minimum és maximum), továbbá a klaszterbeli személyek átlagos távolsága (HC=homogenitási együttható).

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

Relokáció: nemhierarchikus k-központú klaszteranalízis

Ebben a modulban relokációs elemzés, vagyis K-központú (K-means) klaszteranalízis végezhető (vö. *Bergman–Magnusson–El-Khoury* [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 vagy 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 szétszítja (besorolja) a többi kód által meghatározott klaszterekbe, minden személyt a hozzá legközelebbi nem 0 kódú klaszterbe téve be.

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, elvégző azon egy K-központú klaszteranalízist és kiszámítja a kapott megoldáshoz tartozó különböző klaszter adekvációs mutatók (QC-k) 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 vagy 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]).

Konfigurációanalízis (CFA)

Ezzel a modullal diszkrét változók tipikus és atipikus érték kombinációinak azonosítása végezhető el. E modul választása esetén a program a véletlennél szignifikánsan gyakrabban és ritkábbal 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]).

A CFA tulajdonképpen egy sűrűsödéspont-elemzés kvalitatív változókkal, hiszen a szignifikáns értékszekvenciák az input diszkrét változók többdimenziós terének azon pontjai (a típusok), ahol a vártnál többen tömörülnek. Azok a pontok pedig, ahol pedig a függetlenség esetén vártnál szignifikánsan kevesebben vannak (az antitípusok), azok e tér fekete lyukaiként értelmezhetők.

ExaCon

Ezen elemzés célja kétszemponos gyakorisági táblázatok celláinak elemzése egzakt próbákkal. Ebben a modulban a függetlenség esetén elvártnál szignifikánsan nagyobb és kisebb cellagyakoriságok azonosíthatók (vö. *Bergman–Magnusson–El-Khour*i [2003] 5. fejezet és 125-127. old.). Ezzel az elemzéssel meg lehet vizsgálni például, hogy különböző klaszter kód változók által képviselt klaszterezések milyen mértékben fedik át egymást. A hasonlóság egyik mértéke a RAND és a korrigált RAND (ARAND) együttható, melyeket a program mindig kiszámít. Az ExaCon elemzés emellett arra is alkalmas, hogy megvizsgáljuk egy kapott kedvezőnek tűnő klasztermegoldás kapcsolatát az adatállomány egyéb diszkrét változóival (pl. nemmel, iskolai végzettséggel, életkori szinttel stb.).

Centroid

Ennek a modulnak a segítségével összehasonlítható két különböző klaszterváltozó átlagmintázata, ún. centroidja (vö. *Bergman–Magnusson–El-Khour*i [2003] 5. fejezet és 125-126. old.). A program ennek érdekében kiszámítja a két klaszterváltozó által definiált klaszterek centroidjai közti páronkénti távolságokat (ASED mértékkel) és sorba rendezi ezeket, amiből megállapítható, hogy a két megoldás mely klasztereinek középpontjai esnek egybe.

A modul elemzése arra is alkalmas, hogy megnézzük: egyetlen klaszterváltozó átlagmintázata egy csoportosító változó által definiált több különböző csoportban (pl. férfiaknál és nőknél) mennyire ugyanaz, illetve eltérő.

DensePoint: sűrűsödéspont-elemzés

Ezzel a modullal olyan személyek azonosíthatók, akiknek sok szomszédjuk van. A szomszédságot egy beállítható küszöbérték (szomszédság küszöbe) segítségével definiáljuk. Bármely személy (eset) 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 (két személy esetén a kijelölt változók értékeire az összetartozó értékpárok átlagos négyzetes eltérése), vagy az

értékmintázatok Pearson-féle korrelációja (két személy esetén a kijelölt változók értékeire az összetartozó értékpárok közti Pearson-korreláció). A feladatablakban kérhető a távolságok kiszámítása esetén a változók standardizálása, mely feltétlenül indokolt abban az esetben, amikor az elemzésre kiválasztott változók különböző dimenziójúak. 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 küszöbe 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 szomszédtsága milyen mértékben esik bele. Ezzel a más elemzésekben létrehozott klaszterek altípusai (prototípusai) azonosíthatók.

A sűrűsödéspont-elemzéssel kapcsolatban lásd még *Vargha* [2008].

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 változók listájában olyan sorrendben helyezkednek el, hogy időpontonként ugyanolyan sorrendben egymás mellett vannak a megfelelő változók (pl. 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 változók listájában olyan sorrendben helyezkednek el, hogy egymás mellett vannak egy-egy változó különböző időpontokhoz tartozó mérései (pl. 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.

Ha mondjuk 5 attitűdváltozót mérünk 3 alkalommal (3 egymást követő hónapban) és ez a 15 változó szerepel – esetlegesen más változók társaságában – minden személy adatsorában, akkor ezzel az időszeparációs elemzéssel létrehozhatunk egy olyan új adattáblázatot, melyben a különböző időpontok 5-5 adata külön sorokban, külön esetekként szerepel. Az új adatállomány így 3-szor annyi sorból (esetből) áll. Az 5 változóval különböző elemzések végezhetők (pl. klaszteranalízis), mely az adatállomány mérési időponttól független struktúráját deríti fel. Ugyanez a fájl átmenthető természetesen SPSS-be és ott is lehet releváns elemzéseket (pl. faktoranalízist) végezni. A különböző időpontokhoz tartozó adatok ismételt egyesítése az időegyesítés modul segítségével végezhető el (lásd alább).

Időegyesítés

E modul segítségével az időszeparációval létrehozott új fájlban végzett elemzések új változói (pl. klaszterváltozók) visszakerülnek egy eredetivel megegyező rekordstruktúrájú fájlba. Ez alkalmat teremt arra, hogy pl. megvizsgáljuk egy klaszterstruktúra időbeli stabilitását, illetve változását, valamint ugyanezt az egyes személyek vonatkozásában. Érdekes lehet azonosítani a minden időpontban ugyanabba a klaszterbe tartozó személyeket és megvizsgálni jellemzőiket, vagy azokat, akik a legnagyobbat változtak a vizsgált időszakban. Ezzel és az előző modullal könnyen végrehajtható az ISOA komplex klasszifikációs elemzés (vö. Bergman–Nurmi–von Eye [2012]) is.

Irodalom

- BARTKO, J. J. [1976]: On Various Intraclass Correlation Reliability Coefficients. *Psychological Bulletin*, Vol. 83. No. 5, pp. 762–765.
- BERGMAN, L. R. – MAGNUSSON, D. – EL-KHOURI, B. M. [2003]: *Studying individual development in an interindividual context. A Person-oriented approach*. Mahwa, New Jersey, London: Lawrence-Erlbaum Associates.
- 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.
- COHEN, J. [1992]: A power primer. *Psychological Bulletin*, Vol. 112. No. 1, pp. 155–159.
- LANDERS, R. N. [2015]: Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower* 2:e143518.81744. DOI: 10.15200/winn.143518.81744
- NAGYBÁNYAI, N. O. [2006]: A pszichológiai tesztek reliabilitása. In: Rózsa Sándor, Nagybányai Nagy Olivér és Oláh Attila (Szerk.): *A pszichológiai mérés alapjai. Elmélet, módszer és gyakorlati alkalmazás*. ELTE, CD tankönyv, Budapest, 6. fejezet.
- SZÉKELYI M. – BARNÁ I. [2003]: *Túlélőkészlet az SPSS-hez*. Budapest: Typotex Kiadó.
- TAKÁCS, SZ. – MAKRAI, B. – VARGHA, A. [2015]: Klasszifikációs módszerek mutatói. *Psychologia Hungarica Caroliensis*, 3. évf. 1. sz. pp. 67–88.
- VARGHA A. [1988]: Jogos-e többszemponos varianciaanalízist alkalmazni dichotóm függő változók esetén? *Pszichológia*, 8. évf. 3. sz. pp. 409-446.
- VARGHA A. [1996]: Az egymintás t-próba érvényessége és javíthatósága. *Magyar Pszichológiai Szemle*, 52. évf. pp. 317-345.
- VARGHA A. [1999]: *MiniStat 3.1 verzió. Felhasználói kézikönyv*. Pólya Kiadó, Budapest.
- VARGHA A. [2000]: *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal*. PÓLYA KIADÓ. BUDAPEST.
- VARGHA A. [2001]: Kísérleti helyzetek és csoportok összehasonlítása új statisztikai módszerekkel. In: Pléh Cs. – László J. – Oláh A. (szerk.), *Tanulás, kezdeményezés, alkotás*. Budapest: ELTE Eötvös Kiadó, 371–386.
- VARGHA A. [2002]: Független minták egyszemponos összehasonlítása új rangsorolások eljárások segítségével. *Statisztikai Szemle*, 80. évf. 4. sz. pp. 328–353.
- VARGHA A. [2004]: A kétszemponos sztochasztikus összehasonlítás modellje. *Statisztikai Szemle*, 82. évf. 1. sz. pp. 67–82.

- VARGHA A. [2005]: Sokaságok összehasonlítása új módszerekkel. *Statisztikai Szemle*, 83. évf. 5. sz. pp. 429–448.
- VARGHA A. [2007]: *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal (2. kiadás)*. Pólya Kiadó. Budapest.
- VARGHA A. [2008]: Új statisztikai módszerekkel új lehetőségek: a ROPstat a pszichológiai kutatások szolgálatában. *Pszichológia*. 28. évf. 1. sz. pp. 81–103. <http://dx.doi.org/10.1556/Pszi.28.2008.1.5>
- VARGHA A. [2011]: A parciális korrelációs együttható értelmezési problémái a többdimenziós normalitás feltételének sérülése esetén. *Statisztikai Szemle*, 89. évf. 3. sz. pp. 275-293.
- 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. No 1–2. pp. 78–86. http://www.personresearch.ouradmin.se/articles/volume2_1_2/filer/5.pdf
- VARGHA, A. – BORBÉLY, A. [2016]: Modern mintázatfeltáró módszerek alkalmazása a kétnyelvűség kutatásában. In: Kissné Viszket M. – Puskás-Vajda Zs. – Rácz J. – Tóth V. (szerk.) [2016]. *A pszichológiai tanácsadás perspektívái. Tisztelgő kötet Ritoók Magda 80. születésnapjára*. L'Harmattan, Budapest, pp. 173-186.
- 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*, Vol. 23, No 2, pp. 170-192.
- VARGHA A. – IZSÓ L. [1989]: *PSZICHOSTAT statisztikai programcsomag felhasználói füzet (Commodore-64 változat)*. Munkaügyi Kutatóintézet, Budapest.
- 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. No. 1–2. pp. 87–98.
- von Eye, A. – Mair, P. – Mun, E.-Y. [2010]: *Advances in Configural Frequency Analysis*. New York: Guilford Press.