

A

ROPstat

statisztikai
menürendszer

Dr. Vargha András
2007

tartalomjegyzék

1. Alapok (egymintás elemzések)

- Alapstatisztikák
- Részletesebb statisztikák számítása
- Gyakorisági eloszlás, hisztogram készítése
- Közéértékekre 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

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: súsúsódéspont-elemzés
- Időszeparáció
- Időegyesítés

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; legkisebb és legnagyobb érték; legkisebb és legnagyobb standardizált érték (z-érték).

Irodalom: Vargha (2000, 3. fejezet).

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. A program egyben azt is teszteli, hogy a változó ferdesége és csúcsossága eltér-e a normális eloszlásától.

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.

Irodalom: Vargha (2000, 3. fejezet és 5.3.5. alpont).

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.

Irodalom: Vargha (2000, 3. fejezet és 17.1. alfejezet).

Középértékekre 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. 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 $\text{abs}(t) \leq 10$ 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.

Irodalom: Vargha (2000, 7. fejezet).

Normalitásvizsgálat

Ezt az elemzést legfeljebb 100 elemű minták esetén a Kolmogorov-próba, 100-nál nagyobb elemszámú minták esetén pedig a khi-négyzet-próba segítségével végzi el a program.

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.

Irodalom: Vargha (2000, 17.1.3. alpont).

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

2. Csoportok és változók összehasonlítása

Független minták egyszempontos összehasonlítása

Intervallum-skálájának beállított változókra a program az elméleti szórások, átlagok és trimmelt átlagok egyenlőségét vizsgálja (ez utóbbit csak akkor, ha a trimmelési százalék 0-nál nagyobb).

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.

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.

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.

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.

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” (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.

Ha az összehasonlítandó csoportok száma kettő és a kiválasztott függő változók száma legalább kettő, akkor a program – automatikusan – közös tömör táblázatban foglalja össze az összehasonlítások eredményeit. A feltüntetett adatok változónként: a két csoport átlaga és szórása, az átlagok standardizált különbsége (a Cohen-féle hatásmértékkel kifejezve), amit úgy kapunk, hogy az átlagok különbségét leosztjuk a két minta együttes szórásával, az elméleti szórások egyenlőségét tesztelő Levene-próba p -értéke, a kétmintás t -próba

eredménye, szabadságfoka és p -értéke, valamint a Welch-féle d -próba eredménye, szabadságfoka és p -értéke.

Ordinális skálájúnak beállított változók esetén a program az elméleti rangszórások megegyezését és a sztochasztikus homogenitást teszteli. Populációk 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 egyesítés 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.

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 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. Ennek tesztelése a khi-négyzet-

próba, illetve a Fisher-egzakt-próba segítségével történik. 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.

Ha az összehasonlítandó csoportok száma kettő és a kiválasztott függő változók száma legalább kettő, akkor a program – külön kérés nélkül – közös tömör táblázatban foglalja össze az összehasonlítások eredményeit. A feltüntetett adatok változónként: a két csoport rangátlaga, a valószínűségi fölény $A(1, 2)$ mutatója, az elméleti rangszórások egyenlőségét tesztelő Levene-próba p -értéke, a Mann-Whitney-próba p -értéke, a Fligner-Policello-próba eredménye, szabadságfoka és p -értéke, valamint a Brunner-Munzel-próba eredménye, szabadságfoka és p -értéke.

Nominális skálájúnak beállított változók esetében a program elkészíti a csoportosító változó (G) és a függő változó (X) csoportjai által meghatározott kétdimenziós gyakorisági táblázatot és χ^2 -próbával (kis elemszámok esetén Fisher-egzakt-próbával) teszteli, hogy X függ-e G -től. Szignifikáns eredmény esetén a program elkészíti a táblázat sorösszegei szerint képezett százalékos gyakoriságok táblázatát is.

Ha a feladatablakban azt állítjuk be, hogy az eredmények kilistázása csak szignifikáns különbségek esetén történjen, akkor az eredménylistán a program nem jeleníti meg azon függő változókkal kapcsolatos eredményeket, amelyek esetében semmilyen összehasonlítás nem szignifikáns.

Irodalom: Vargha (2000, 9., 10., 13. és 15. fejezet, illetve 17.3. alfejezet) és Vargha (2001, 2002, és 2005).

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 egyszempontos összetartozó mintás VA-val teszteli az elméleti átlagok megegyezését és Friedman-próbával, illetve rangszámokon végrehajtott VA-val a sztochasztikus homogenitást, mely a független mintás esettel analóg módon definiálható. A hagyományos VA mellett a program szabadságfok korrekciók (Greenhouse-Geisser és Huynh-Feldt-féle) robusztus VA-t is végrehajt. Tekintettel arra, hogy az összetartozó mintás VA szórás-homogenitási 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 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 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 ($A(k, u) - 0,5$), 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 (Johnson- és Gayen-próba), a mediánok egyenlőségének tesztelésére a Wilcoxon-próbát, a sztochasztikus egyenlőség tesztelésére pedig az előjelpróbát hajtja végre. 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 nullhipotézis 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” (trimmelni). Ezt 0-tól különböző trimmelési százalék beállításával tehetjük meg. 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.

Nominális változók esetén a változás/eltérés vizsgálatára a program a függő változók és a talált különböző értékek számától függően különböző diszkrét elemzéseket végez. 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 mind az öt értékre vonatkozóan a Cochran-féle Q-próbával teszteli, hogy az érték ugyanolyan valószínűséggel fordul-e elő a három függő változónál.

Irodalom: Vargha (2000, 8., 14. és 15. fejezet, illetve 17.4. alfejezet) és Vargha (2001).

Átlagok kétszemponos összehasonlítása

Ezzel a modullal független minták kétszemponos varianciaanalízise végezhető el. 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 változat) eredményét alapul venni.

Ha 0-tól különböző trimmelési százalékot állítunk be, a program robusztus kétszemponos trimmelt VA-t is végrehajt.

Eredménylista: átlagok, szórások, illetve a csoportonkénti megfigyelések (személyek) száma a függő változókra; a kétszemponos varianciaanalízis összefoglaló táblázata; Welch-Johansen-féle robusztus kétszemponos VA; trimmelt átlagokra vonatkozó robusztus kétszemponos VA; szignifikáns főhatás esetén a szintek páronkénti összehasonlítása.

Ha a feladatablakban azt állítjuk be, hogy az eredmények kilistázása csak szignifikáns különbségek esetén történjen, akkor az eredménylistán a program nem jeleníti meg azon függő változókkal kapcsolatos eredményeket, amelyek esetében a VA egyik hatása sem szignifikáns.

Irodalom: Vargha (2000, 16.1. és 16.2. alfejezet).

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). A mintaelemszámok eltérőek 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, valamint a két szempont interakcióját pedig egy szabadságfok-korrektív VA-val (Greenhouse-Geisser és Huynh-Feldt-féle robusztus VA).

Eredménylista: átlagok és szórások táblázata; a kétszemponos varianciaanalízis összefoglaló táblázata; robusztus kétszemponos varianciaanalízis; szignifikáns főhatás esetén a szintek páronkénti összehasonlítása.

Irodalom: Vargha (2000, 16.1. és 16.3. alfejezet).

Kétszemponos rang-varianciaanalízis

Ezzel a modullal független minták kétszemponos sztochasztikus homogenitásvizsgálata végezhető el. Egy (i, j) alpopuláció 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, 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 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.

Ha a feladatablakban azt állítjuk be, hogy az eredmények kilistázása csak szignifikáns különbségek esetén történjen, akkor az eredménylistán a program nem jeleníti meg azon függő változókkal kapcsolatos eredményeket, amelyek esetében a kétszemponos rang-VA egyik hatása sem szignifikáns.

Irodalom: Vargha (2004).

3. Változók kapcsolatának vizsgálata

Korreláció, egyszerű regresszió

Ez a modul változópárokra számítja ki az egyszerű lineáris regressziós egyenletet és a két változó közti kapcsolat szorosságának mérésére egy sor korrelációs mérőszámot, intervallumbecslésükkel együtt.

Ha 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.

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 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 konfidencia-intervallumot is. Itt a megbízhatósági szint alapértelmezése 90%, de ez az érték a feladatablakban 95%-ra, illetve 99%-ra is átállítható.

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 -hibavarianciát is, vagyis X , illetve Y regressziós becslésének átlagos négyzetes hibáját.

Eredménylista (minden elemzett változópárra): átlagok és szórások; legkisebb és legnagyobb értékek; Y egyszerű regressziós egyenlete X függvényében, illetve X egyenlete Y függvényében; Pearson-féle lineáris korrelációs együttható; Wilcox-féle r_{pb} robusztus korrelációs együttható; Kendall-féle tau és tau-b monotonitási együttható; konfidencia-intervallum az elméleti kapcsolati mutatókra.

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 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 % és a diszkordancia % értékét.

Irodalom: Vargha (2000, 11. és 12. fejezet).

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). 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.

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 véletlenül szorosabb lineáris kapcsolat a kijelölt változók csoportjában.

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.

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

Eredménylista: egyváltozós összefoglaló statisztikák; korrelációs és parciális korrelációs mátrix; Bartlett-féle függetlenségvizsgálat.

Irodalom: Vargha (2000, 11. és 12. fejezet).

Diszkrét változók kapcsolatvizsgálata

Ez a modul diszkrét (pl. nominális skálájú) változók kapcsolatának vizsgálatára alkalmas. Segítségével a kijelölt változópárok kétszemponos gyakorisági táblázata készíthető el, tesztelhető a változópár függetlenségének nullhipotézise (khi-négyzet-, 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 kilistázás, „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.

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

Eredménylista (minden kijelölt változópárnál): kétszemponos gyakorisági táblázat; sorösszegek és oszlopösszegek szerinti százalékok táblázata; a függetlenség tesztelése a khi-négyzet-próba, illetve a Fisher egzakt-próba segítségével (utóbbi csak 2x2-es táblázatokra, kisebb gyakoriságok esetén); Cramér-féle V kontingencia-együttható; esélyhányados ('odds ratio' vagy 'cross-product ratio'); Yule-féle Q és Y asszociációs együttható 2x2-es táblázatokra; monotonitási együtthatók (Kendall-féle γ és τ -b, Somers-féle D); a monotonitás tesztelése (a Kendall-féle γ segítségével); predikciós mutatók (Goodman-Kruskal-féle τ és λ); közös információ-hányad; Cohen-féle kappa.

Ha a feladatablakban azt állítjuk be, hogy az eredmények kilistázása csak szignifikáns kapcsolatok esetén történjen meg, akkor az eredménylistán a program nem jeleníti meg azon változópárokkal kapcsolatos eredményeket, amelyek esetében a khi-négyzet-próba nem szignifikáns.

Irodalom: Vargha (2000, 17. fejezet).

Itemanalízis

A program a kijelölt változók, mint teszttételek együttese által definiált ún. additív skála hagyományos megbízhatósági (reliabilitási) vizsgálatát végzi. 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 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, a definiált skála átlagát és szórását, külön kérésre az itemekre vonatkozó alapstatisztikákat, 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 korrigált item-totál korrelációnak is nevezni.

Irodalom: Nagybányai (2006).

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. 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.

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 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 korrelációs és determinációs együtthatót (R és R^2), a regressziós becslés standard hibáját, valamint a lépésenkénti elemzés harmadik blokkjában közölt információt.

Az elemzésben a regressziós becslés külön változóként létrehozható és beilleszthető az adattáblázatba.

Irodalom: Székelyi és Barna (2003, 5. fejezet).

4. Mintázattalátoló elemzések

Adattalátoló

Ebben a modulban a program alapstatisztikákat számol és tájékoztat a hiányzó adatok mintázattalátoló. 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 menüablakban kérjük a kijelölt változók korrelációs mátrixának elkészítését, akkor a program külön kérésre páronként összehasonlítja az egyazon változópárokhoz tartozó korrelációs együtthatókat.

Hiányzó adatok pótlása

Ennek a modulnak a segítségével hiányzó adatok pótolhatók a következő három módszer valamelyikével:

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.

Irodalom: Bergman, Magnusson és El-Khoury (2002, 107-111. o.).

Reziduálanalízis

Ebben a modulban az adatállomány extrém, kilógó személyei azonosíthatók a hozzájuk leghasonlóbb más esetek (legközelebbi szomszédok) megkeresése segítségével. 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 hasonlóság mértéke lehet a szokásos euklideszi távolság, 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.

Irodalom: Bergman, Magnusson és El-Khoury (2002, 109-110. o.).

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.

1. A program minden esetpárnak meghatározza a távolságát a megadott távolságmérték felhasználásával. 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. 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 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 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 (EES = error sum of square = négyzetes összhiba). Az összheterogenitást úgy határozzuk meg, hogy minden klaszter esetén kiszámítjuk a klaszterbe tartozó személyek klasztercentroidtól való távolságainak összegét, majd ezeket összeadjuk. Röviden: ennél a módszernél minden lépésben azt a két klasztert vonjuk össze, amely esetében az ESS növekedés (ESSnö.) a legkisebb.

- Béta flexibilis módszer: ez egy viszonylag bonyolult, de néha igen sikeres klasszifikációt eredményező összevonási módszer (a részletekről lásd Bergman és mtsai, 67. o.).

4. A program minden lépésben feltünteti az összevonás lépését, az adott lépésben létező klaszterek számát (KL#), a megmagyarázott ESS%-ot, az ESS növekedést (ESSnö.), valamint az összevonásra kerülő két klaszter kódját és a kód után zárójelben a klaszter elemszámát.

A megmagyarázott ESS% (EV) 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:

$$EV = 100 \cdot (ESS_{max} - ESS) / ESS_{max}.$$

Ezt a mutatót az adott osztályozáshoz tartozó megmagyarázott variancia arányának is szokták nevezni. EV értéke függ az elemzésbe bevont változók számától (több változó esetén általában kisebb) és az összelemszámtól (nagyobb minta esetén általában kisebb). Sikeres osztályozás esetén EV értéke eléri a 0,60-at.

A klaszterkód 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 klasztererek főbb jellemzőit (klaszternagyság, változónként átlag, szórás, minimum és maximum), továbbá a klaszterbeli személyek átlagos távolsága (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 EV 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 fölfelé vagy lefele az átlagostól.

Irodalom: Bergman és mtsai (2002, 4. fejezet és 113-115. o.); Székelyi és Barna (2003, 3. fejezet).

Relokáció: nemhierarchikus k-központú klaszteranalízis

Ebben a modulban relokációs elemzés, vagyis k-központú (K-means) klaszteranalízis végezhető. 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 klaszterbesorolását az elemzés kezdetekor. Ha nem adunk meg induló 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 ugyanazokat az eredményeket listázza ki, mint a nemhierarchikus elemzés relokációs elemzése után.

Ebben a modulban lehetőség van arra is, hogy a program tesztelje, vajon a kapott klaszterstruktúra jobb-e, mint egy véletlenszerűen összeállított adatállományból létrehozott klaszterstruktúra. Ez oly módon történik, hogy a program egymás után, egymástól függetlenül többször (a független ismétlések száma beállítható) a változóoszlopok adatait az oszlopokon belül véletlenszerűen összekeveri, majd az ily módon randomizált adatállományon végez K-központú klaszteranalízist és számítja ki a kapott klaszterstruktúrához tartozó megmagyarázott ESS% (EV) értéket. Ezután a program egymintás t -próbával teszteli, hogy a véletlenszerű besorolásokkal nyert EV-értékek szignifikánsan kisebbek-e, mint az eredeti adatokon végrehajtott klaszterelemzés EV-értéke.

Irodalom: Bergman és mtsai (2002, 4. fejezet és 115-117. o.); Székelyi és Barna (2003, 3. fejezet).

Konfigurációanalízis (CFA)

Ezzel a modullal diszkrét változók tipikus és atipikus értékkombinációinak 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.

Irodalom: Bergman és mtsai (2002, 5. fejezet és 117-119. o.).

ExaCon

Ezen elemzés célja kétszemponos gyakorisági táblázatok celláinak elemzése egzakt próbákkal. Ebben a modulban a függetlenség esetén elvártnál szignifikánsan nagyobb és kisebb cellagyakoriságok azonosíthatók.

Irodalom: Bergman és mtsai (2002, 5. fejezet és 125-127. o.).

Centroid

Ennek a modulnak a segítségével összehasonlítható több különböző klaszterváltozó átlagmintázata (centroidja) vagy egyetlen klaszterváltozó átlagmintázata több különböző csoportban (valamely csoportosító változó által definiálva).

Irodalom: Bergman és mtsai (2002, 5. fejezet és 125-126. o.).

DensePoint: sűsűsödéspont-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 szomszédsága milyen mértékben esik bele. Ezzel a más elemzésekben létrehozott klaszterek altípusai (prototípusai) azonosíthatók.

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.

Időegyesítés

Ezen modul segítségével az időszeparációval létrehozott új fájlban végzett elemzések új változói (pl. klaszterváltozók) visszakerülnek egy eredetivel megegyező rekordstruktúrájú fájlba.

Hivatkozások

- Bergman, L. R., Magnusson, D. és El-Khoury, B. M. (2002). *Studying individual development in an interindividual context. A Person-oriented approach. Mahwah, New Jersey, London: Lawrence-Erlbaum Associates.*
- Nagybányai Nagy Olivér (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 Mária és Barna Ildikó (2003). *Túlélőkészlet az SPSS-hez.* Budapest: Typotex Kiadó.
- Vargha András (2000). *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal.* Budapest: Pólya Kiadó.
- Vargha András (2001). Kísérleti helyzetek és csoportok összehasonlítása új statisztikai módszerekkel. In: Pléh Cs., László J. és Oláh A. (szerk.), *Tanulás, kezdeményezés, alkotás.* Budapest: ELTE Eötvös Kiadó, 371-386.
- Vargha András (2002). Független minták egyszempontos összehasonlítása új rangsorolások eljárások segítségével. *Statisztikai Szemle, 80 (4), 328-353.*
- Vargha András (2004). A kétszempontos sztochasztikus összehasonlítás modellje. *Statisztikai Szemle, 82 (1) 67-82.*
- Vargha András (2005). Sokaságok összehasonlítása új módszerekkel. *Statisztikai Szemle, 83 (5), 429-448.*
- Vargha András (2007). *Matematikai statisztika pszichológiai, nyelvészeti és biológiai alkalmazásokkal (2. kiadás).* Budapest: Pólya Kiadó. (megjelenés alatt)