

FOGALMAK ÉS MÓDSZEREK

FÜSTÖS LÁSZLÓ—MESZÉNA GYÖRGY—SIMONNÉ MOSOLYGÓ NÓRA

A sokdimenziós skálázás egyes újabb módszerei, III.

Bevezetés

Az első cikkben egy általános bevezetés után a MINISSA és az MRSCAL eljárásokat tárgyaltuk, egyes kapcsolódó meggondolásokkal együtt. Megjelent a Sigma 1982. XV. évf. 3. számában. A második részben — Sigma 1983. XVI. évf. 3. szám — az INDSCAL eljárást és változatait, a PROFIT és PREFMAP eljárásokat tekintettük át. A most következő harmadik részben a MINIRSA, az MDPREF, a PARAMAP és az UNICON eljárásokat mutatjuk be.

Az egyes cikkek közötti időbeni távolságok áthidalása érdekében néhány mondatban emlékeztetünk a sokdimenziós skálázás problémakörének egyes gondolataira.

A megfigyelési egységek rangsorolása sok tényezőnek együttes figyelembevételével, ez a kérdés mind elméleti, mind gyakorlati oldalról az érdeklődés előterébe került. Számos gyakorlatilag is jól használható eljárás született — a kompromisszum kötés módozataiban különbözve egymástól —, melyek a klasszikus értelemben vett egy dimenziós output rangsort, „skálát” adták eredményül. Cikksorozatunk nem érinti ezeket a gondolatmeneteket.

Egy másik erőteljesen fejlődő irányzat a több dimenziós outputot adó módszerek létrehozását szorgalmazza. Feloldva az egy dimenziós eredő rangsor igen erős következményét, remélhető, hogy a csökkenő kompromisszum igénye mellett tisztábban lesznek felismerhetők a rendszerbeli kapcsolatok, belső összefüggések, mindaddig, amíg a magasabb dimenziószámok más oldalról meg nem növelik a technikai nehézségeket. A cikksorozatban szereplő eljárások közvetlenül vagy közvetve ehhez a gondolatkörhöz tartoznak.

1. A MINIRSA eljárás (*Mini-Rectangular (smallest) Space Analysis*)

Az ismertető modell — a MINISSA eljáráshoz hasonlóan — kétdimenziós adatmátrixok elemzésére szolgál, s az I. cikk elején adott definíció értelmében nem metrikus módszer.

Tekintsük az alábbi adatmátrixot:

		objektumok					
		1	2	...	j	...	m
„személyek”	1						
	2						
	.						
	.						
	i						
	.						
	n						

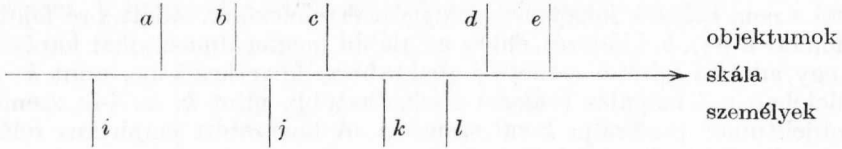
az i -ik sor j -ik helyén álló rangszám adja meg azt az információt, hogy az „ i ” indexű személy a „ j ” objektumot hogyan preferálja. Tehát minden sor az objektumoknak valamely személy szemszögéből történő preferencia sorrendjét tartalmazza.

A MINIRSA-modell r dimenziós térben helyezi el a személyeknek és objektumoknak megfelelő $(n + m)$ pontot az alábbi elv szem előtt tartásával. Jelöljük az „ i ” személyt és „ j ” objektumot reprezentáló pont távolságát d_{ij} -vel.

Megköveteljük, hogy a $d_{i1}, d_{i2}, \dots, d_{im}$ távolságok relatív nagysága feleljen meg a fenti adatmátrix i -ik sorában található preferencia rendezésnek. ($i = 1, 2, \dots, n$). A megfeleltetés jóságát egy alkalmas hibafüggvénnyel mérjük, melyet a pont-kijelölés iteratív folyamatában minimalizálunk.

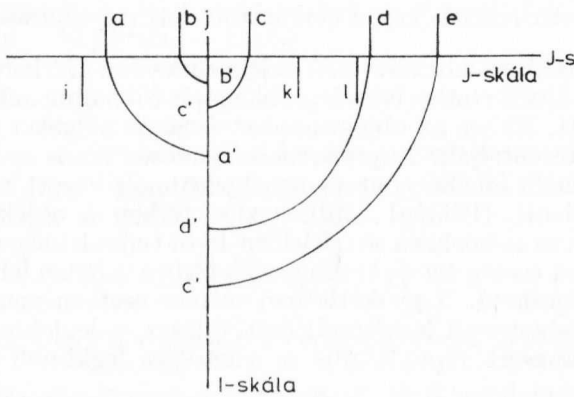
Tételezzük most fel, hogy az m számú objektum valamilyen közös tulajdonság alapján kvantitatívan is összehasonlítható, amit úgy is kifejezhetünk, hogy az objektumoknak egy közös egydimenziós skálán vannak skálaértékei.

Ugyanezen az egydimenziós skálán most az előbbieken bevezetett adatmátrixban szereplő személyeket is megkíséreljük reprezentálni egy-egy ponttal. E pontokat a személyekhez tartozó „maximum preferencia” vagy „ideális” pontoknak fogjuk nevezni. E pontok kijelölését a következő megfontolás alapján végezzük. Feltételezzük, hogy minden személy annál inkább részesíti előnyben (preferálja) az egyes objektumokat, minél közelebb helyezkednek el a skálán (az objektumokat jelentő pontok) a szóban forgó személyt reprezentáló ideális ponthoz. Így az objektumokat rögzítő skála pontok és a személyek ideális pontjai között mért távolságok monoton függvényei lesznek az adatmátrixban rögzített személyenkénti preferencia rangsornak. Az 1. ábra öt objektumot (a, b, c, d, e) és négy személyt (i, j, k, l) ábrázol egy egydimenziós skálán:



1. ábra. Objektumok és személyek az egydimenziós skálán

Ezt az egydimenziós skálát *J*-skálának (joint scale) nevezzük. A *J*-skála alapján egy személy preferencia rangsorának megfelelő *I*-skálát vetítéssel kaphatjuk meg. A vetítést a *j* személyre a *Coombs*-féle „kibontó” (unfolding) technikával a 2. ábrán látható módon végeztük el.



2. ábra. *I*-skála a „*j*” személyre

Ha a *J*-skála alapján a példában szereplő többi három személy ideális pontjához is hasonló módon elvégezzük a vetítést, a következő rangsorokat kapjuk:

személy	preferencia rangsor				
i	a	b	c	d	e
j	b	c	a	d	e
k	e	d	b	e	a
l	d	c	e	b	a

Könnyen belátható, hogy egy adott *J*-skála (illetve az objektumok egy már rögzített skála-érték rendszere esetén) nem jöhet létre minden elvileg lehetséges preferencia rangsor. Például, ha az objektumok skálaértékei olyanok — mint a fenti példában —, hogy: *b, c, d* skálaértékei az *a* és *e* értékei közé esnek, akkor nem rögzíthető a skálán olyan személy (azaz az őt reprezentáló ideális pont), amelyhez: *e, a, b, c, d* preferencia sorrend tartozik.

Vezessük be a „ \geq_e ” szimbólummal leírható, gyenge rendezést létrehozó „empirikus relációt”. Az empirikus jelző kívánja hangsúlyozni az eltérést az aritmetika „ \geq ” relációjától, valamint arra is figyelmeztet, hogy itt „dolgok”

közötti s nem számok közötti összefüggést értelmezzük. Adott i -re fejezze ki relációnkat a $p^i(j, k)$ kifejezés, ehhez az alábbi megfogalmazásokat kapcsolhatjuk: egy adott i feltétel mellett j gyakrabban következik be, mint k ; egy i szituációban a j reagálás (válasz) elfogadhatóbb, mint k ; az i -ik személy a j -ik objektumot preferálja k -val szemben. A bevezetett empirikus relációra érvényesek a következő tulajdonságok:

1. $p^i(j, j)$ (reflexivitás)
2. ha: $p^i(j, k)$ és $p^i(k, j)$ akkor:
 j ekvivalens k -val (antiszimmetria)
3. ha: $p^i(j, k)$ és $p^i(k, h)$ akkor:
 $p^i(j, h)$. (tranzitivitás)

Ha meggondolásunkban m objektum szerepel a J -skálán az őket reprezentáló pontok $m!$ sorrendben helyezkedhetnek el. Kimutatható, hogy a példa szerinti egydimenziós esetre szorítkozva, a bemutatott vetítéses eljárással csak $\binom{m}{2} + 1$

féle rangsort tudunk előállítani: (pl. 5 objektum esetén 120 helyett csak 11-et.) Az ilyen módon kezelhetetlen esetek problémáját több dimenziós terek felhasználása hidalja át. Ekkor az objektumokat és a személyeket az n dimenziós tér pontjai reprezentálják. A preferencia rendezés itt is a több dimenziós térben elhelyezkedő ideális pont és az objektumok között mért távolságok rangsorolását jelenti. (Például kétdimenziós térben 5 objektum lehetséges 120 sorrendjéből az előzőeknek megfelelően 46-ot tudunk megadni.) Belátható, hogy m objektum esetén $(m - 1)$ dimenziós térben minden lehetséges sorrend realizációja elvégezhető. A gyakorlatban számos esetben nem is cél minden elvileg lehetséges sorrendi konfiguráció előállítását, a legjobb megoldás a lehető legtöbb rangsort reprodukálja a lehetséges legkisebb dimenziószámú térben.

Jelölje az objektumokat j és k index ($j = 1, 2, \dots, m$ és $k = 1, 2, \dots, m$) a személyeket (megfigyelési egységeket) az i index ($i = 1, 2, \dots, n$). Rendelkezésünkre áll a személyeknek az objektumokra vonatkozó preferencia rangsora. [E megállapítások rögzíthetők a $p^i(j, k)$ szimbólummal.] A preferencia mértékét a sokdimenziós térben az i -edik személy ideális pontjának a j -edik objektumot reprezentáló ponttól mért d_{ij} távolsága adja meg.

Legyenek a személyek ideális pontjainak koordinátái az X , az objektumok konfigurációja pedig az Y mátrixban. Ekkor:

$$d_{ij} = \left\{ \sum_t |x_{it} - y_{jt}|^p \right\}^{1/p} \quad (t = 1, 2, \dots, Z),$$

ahol: x_{it} az i -ik ideális pont (személy) t -ik koordinátája

y_{jt} a j -ik objektum t -ik koordinátája

d_{ij} az „ i ” és „ j ” pontok közötti távolság.

Ha $p = 2$ akkor a megszokott euklideszi távolsággal dolgozunk.

Az eljárás indításakor egy alkalmas kezdő konfigurációt veszünk fel s erre kiszámítjuk a d_{ij} távolságokat. Ha rendelkezünk megfelelő információkkal, ez az indulórendszer már tartalmazhatja a valóság egyes markáns tulajdonságait, ellenkező esetben magával a programmal is lehet teljes általánosságban egy induló pont konfigurációt létrehozni.

Ezután — később részletezendő módon — a preferenciákra támaszkodva: $\hat{d}_{ij} \leq \hat{d}_{ik}$ ha: $p^i(j, k)$ kritérium szem előtt tartásával meghatározzuk a d_{ij}

becsléseket. E becslések jóságát az alábbi S függvénnyel mérjük:

$$S = \sqrt{\frac{\frac{1}{n} \sum_i \sum_j (d_{ij} - \hat{d}_{ij})^2}{\sum_j (d_{ij} - \bar{d}_i)^2}}, \text{ ahol: } \bar{d}_i = \frac{1}{m} \sum_j d_{ij}.$$

S értéke a 0 és 1 közé esik. Az iterációs eljárás során a pontkonfigurációt úgy igyekszünk változtatni, hogy S értéke csökkenjen. A konfiguráció változtatásakor a gradiens módszert használjuk fel.

Az elmondottak alapján az eljárást pontokba szedve a következőképpen foglalhatjuk össze:

1. Ha a felhasználó nem biztosít kezdeti konfigurációt, akkor a program maga generál egy induló pont-rendszert. (Először az objektumok helyét jelölve ki, majd a személyek ideális pontjait helyezi el az általuk legjobban preferált két pont – objektum – közé.)
2. Az origót az összes pont centroidjába helyezzük. (Normalizálva a rendszert: a koordináták négyzetösszege $(n + m)$ r -rel lesz egyenlő.)
3. Kiszámítjuk a d_{ij} távolságokat.
4. Az I. cikkben ismertetett, *Kruskal*-féle monoton regressziós eljárással előállítjuk a \hat{d}_{ij} értékeket. (Ha az aktuálisan előálló \hat{d}_{ij} értékek nem tesznek eleget a fentebb említett:

$$\hat{d}_{ij} \leq \hat{d}_{ik} \text{ ha } p^i(j, k)$$

kritériumnak, akkor a megfelelő d_{ij} értékek átlagával tesszük őket egyenlővé.)

5. Az illeszkedés ellenőrzéséhez kiszámítjuk az S mérőszám értékét.
6. Ha az S eléri a megkívánt szintet, az eljárás befejeződik.
7. Ellenkező esetben a gradiens módszer alkalmazásával módosítjuk a pontkoordinátákat, s ezzel a teljes pont konfigurációt.
8. Az iterációt a 2. ponttól folytatjuk.
9. A végső megoldást rotáljuk úgy, hogy a rendszer tengelyei egybe essenek a konfiguráció főkomponenseivel. Az így rögzíthető tengelyek a rendszert hordozó szerepük, annak „kifeszítéséhez történő hozzájárulásuk” mértékében rendezettek lesznek.

Végül meg kell jegyezni, hogy a gradiens módszer nem biztosítja a globális, csak a lokális minimum elérését. Ezért ilyen helyzetben a program automatikusan egy teljesen eltérő induló konfigurációból is végig viszi az iterációt és a végső megoldást a minimális célfüggvény (S) érték alapján választjuk ki.

2. Az MDPREF eljárás (*Multi Dimensional Preference Scaling*)

Induljunk ki változatlanul az 1. pont elején rögzített adatmátrixból. Ez n személy m objektumra vonatkozó preferencia rendszerét tartalmazta. Az MDPREF-modell a MINIRSA eljáráshoz hasonlóan, mind a személyeket, mind az objektumokat reprezentáló, összesen $(n + m)$ pontot szimultán egy r dimenziós térbe illeszti. A két eljárás elvileg az illesztés kritériumában tér el egymástól. A jelen esetben azt követeljük meg, hogy az objektumok vektorainak

egy-egy személy vektorára vonatkozó merőleges vetületei relatív nagyságukat tekintve feleljenek meg az illető személy preferencia-sorrendjének.

A gondolatmenet némileg hasonlít a PROFIT eljárásnál (l. a II. sz. cikk, Szigma XVI. évf. 3. sz.) látottakhoz, csak amíg ott az objektumok a priori terébe illesztettük a személyek preferenciáinak legjobban megfelelő vektort, az MDPREF-eljárás az objektumok és személyek terét egyidejűleg határozza meg.

Az eljárás kissé részletesebb leírásához fejezzük ki az induló adatmátrixban levő preferenciákat az eddigiektől eltérő technikai megoldással. Tartalmazza az i -ik személy preferenciáit a $P_i = \{p_{i,jk}\}$ mátrix, ahol

$$p_{i,jk} = \begin{cases} +1, & \text{ha: az } i \text{ személy a } j \text{ objektumot preferálja } k\text{-val szemben} \\ -1, & \text{ha: az előzővel ellenkező eset áll fenn} \\ 0, & \text{ha: az } i \text{ személy egyik objektumot sem részesíti a kettő közül} \\ & \text{előnyben.} \end{cases}$$

Legyen továbbá $x_j = [x_{j1}, x_{j2}, \dots, x_{jz}]$ a j objektum $y_i = [y_{i1}, y_{i2}, \dots, y_{iz}]$ az i személy z elemű vektora.

E jelölésekkel az i személy részéről a j objektum iránt megnyilvánuló preferencia becslése (egy konkrét pont konfiguráció pl. az induló rendszer felhasználásával) a vetület és skaláris szorzat összefüggése értelmében

$$s_{ij} = y_i' x_j.$$

Általánosabban az m objektum koordinátáit az r dimenziós térben tartalmazza az: $\mathbf{X}_{(m \times r)} = \{x_{jt}\}$; az n személy koordinátáit pedig az: $\mathbf{Y}_{(n \times r)} = \{y_{it}\}$ mátrix.

A becsült preferencia értékeket a következőképpen kapjuk:

$$\mathbf{S} = \{s_{ij}\} = \mathbf{Y}\mathbf{X}'.$$

Problémánk tehát úgy fogalmazható, meg kell határozni olyan \mathbf{Y} és \mathbf{X} mátrixokat (olyan pontokat az adott r dimenziós térben), hogy a belőlük számított preferencia értékek (\mathbf{S}) minél jobban közelítsék a megfigyelt (mért) preferencia adatokat. CARROL és CHANG (1964) adott két eljárást a feladat megoldására. Az egyik egy iteratív módszer, a másik pedig az Eckart—Young dekompozíciós eljárást használja. Az MDPREF-program az utóbbi eljárással dolgozik. Az Eckart—Young-eljárás az \mathbf{SS}' vagy az $\mathbf{S}'\mathbf{S}$ mátrix sajátértékét és a sajátvektorát számítja. Carroll és Chang Monte-Carlo-módszerrel végzett elemzéssel arra a megállapításra jutott, hogy az Eckart—Young-eljárás ugyanolyan jól dolgozik, mint a feladat iteratív megoldása.

3. A PARAMAP-eljárás (Parametric Mapping)

A három cikkben ismertetett skálázó módszerek egy része a többdimenziós output alapgondolatát módosítja, fejleszti tovább, hogy az elgondolásban rejlő kompromisszum mértéke és tulajdonságai minél jobban követhetők legyenek. A többi eljárás a skálázás alapproblémájához kapcsolódó más kérdés-felvetésekre keresi a választ.

A PARAMAP-eljárás az első csoportba tartozik, kereteiben szorosan kapcsolódik a MINISSA-modellhez, de amíg ott az input és output jellemzők

között monoton, illetve az MRSCAL esetében alapvetően lineáris kapcsolatot tételeztünk fel, itt nemlineáris, esetleg nem monoton kapcsolatok feltételezésére kerül sor.

A PARAMAP-eljárás alap gondolata szerint az m megfigyelt változó kifejezhető r számú ($r < m$) látens változó nemlineáris függvényével. Így az n objektum az input m dimenziós teréből átvihető egy redukált r dimenziós térbe. Az eljárás a mondottak szerint a nemlineáris faktorelemzés egy megvalósításának is tekinthető. A program kidolgozói SHEPPARD és CARROLL (1966), illetve CARROLL és CHANG (1973).

Az eljárás inputja lehet:

1. n objektum m változóra vonatkozó mérési eredményeit tartalmazó, $(n \times m)$ típusú adatmátrix, vagy

2. egy $(n \times m)$ típusú szimmetrikus különbözőségi (hasonlósági) mátrix.

A második változat esetén, ha a hasonlósági mátrix kovarianciákat tartalmaz, áttérhetünk az objektumok közötti „távolságokra”, a következőképpen:

$$d_{ij}^2 = c_{ii} + c_{jj} - 2c_{ij}.$$

Ha a korrelációs együtthatókat ismerjük, akkor: $r_{ii} = r_{jj} = 1$, s így

$$d_{ij}^2 = 2(1 - r_{ij}).$$

Egyéb különbözőségi mérték alkalmazása esetén:

$$d_{ij}^2 = \sum_{k=1}^n (\delta_{ik} - \delta_{jk})^2.$$

A következőkben összefoglaljuk az eljárás gondolatmenetét. Legyen $\mathbf{Y}_{(n \times m)} = \{y_{ik}\}$ az előzőekben említett 1-es típusú input adatmátrix.

A megfigyelt változók tehát: y_1, y_2, \dots, y_m ;

A látens változók: x_1, x_2, \dots, x_r ; $r < m$.

Az alapvető feltételezés a következő:

$$y_k = f_k(x_1, x_2, \dots, x_r)$$

vagy az i -edik objektumra: $y_{ik} = f_k(x_{i1}, x_{i2}, \dots, x_{ir})$.

A PARAMAP-eljárás az r dimenziós output térben keresi azt a pont konfigurációt, melyek közötti D_{ij} távolságok a Carroll-féle κ „folytonossági mérték” szerint jól illeszkednek az input d_{ij} „megfigyelt” távolságaihoz. κ általános alakja:

$$\kappa = \frac{\sum_{i \neq j}^n \frac{d_{ij}^{2a}}{D_{ij}^{2b}}}{\left[\sum_{i \neq j}^n D_{ij}^{2c} \right]^{-b/c}},$$

ahol:

$$d_{ij}^2 = \sum_{k=1}^m (y_{ik} - y_{jk})^2$$

$$D_{ij}^2 = l^2 \sum_{k'=1}^r (x_{ik'} - x_{jk'})^2$$

l : normalizáló konstans.

KRUSKAL és CARROLL (1968) az a , b , c paraméterek következő praktikus megválasztási lehetőségeit javasolta:

$$\begin{array}{lll} a = 1 & b = 1 & c = -1 \\ a = 0,5 & b = 1 & c = -1 \\ a = 0,5 & b = 0,5 & c = -1 \\ a = 1 & b = 2 & c = -1 \end{array}$$

α értelmezéséből következik, hogy egyenletesebb függvénykapcsolatokhoz kisebb α értékek tartoznak.

A α mutató további ekvivalens alakjai is szerepelnek az irodalomban. Ezekből az alakokból megállapítható, hogy a modellben a nagy távolságok kis súlyt, a kis távolságok nagy súlyt kapnak. Az eljárás α -t minimalizálja és eredményül az objektumoknak a redukált térbe illesztett konfigurációját adja.

Az eljárás programja egy önkényesen felvett $(n \times r)$ típusú \mathbf{X} mátrixból indul ki. Ez lehet valamely más skalázó eljárás outputja is. Ezután meghatározza a hozzá tartozó α értéket, majd a gradiens módszerrel iteratív úton lépésről lépésre változtatja az \mathbf{X} mátrixot, míg ki nem alakul egy stacionárius konfiguráció az adott dimenziószámú output térben. A program az iteráció eredményeképpen kapott \mathbf{X} mátrixot normalizálja és a fő tengelyekben rotálja.

4. UNICON-eljárás

(Unidimensional Conjoint Measurement for Multifaceted Design)

Az UNICON-eljárás maximum öt független változónak egy függő változóra vonatkozó összetett, közös hatását határozza meg additív, szubtraktív vagy multiplikatív, illetve ezen műveletek kombinációjával felállított modell segítségével.

A társadalomtudományokban gyakran vizsgálják egy független változó-halmaz együttes hatását egy függő változóra. Például amikor feltételezzük, hogy az életkor és a nem multiplikatívan hat a keresetre, vagy hogy a társadalmi státus és az értékrendszer együttesen befolyásolja az emberek önértékelését, akkor úgynevezett egyesített mérést hajtunk végre. A független változókat, amelyek gyakran nominális vagy ordinális tulajdonságok, egyesítjük valamilyen formában, hogy így előállíthassuk a függő változó értékeit.

A legszélesebb körben ismert és gyakran alkalmazott formája az egyesített mérésnek az N -utas szórás-elemzés (N -way analysis of variance — ANOVA). Az ANOVA-modellben a független változók hatását additív módon összesítjük.

Az ANOVA-modellben a függő változóról feltételezzük, hogy legalább intervallum mérési szintű változó, a független változók pedig nominális mérési szintű változók. KRUSKAL (1964) mutatta meg, hogy ha a függő változó ordinális szintű, egy nem-metrikus MDS-eljárást használhatunk a függő változó értékeinek újraskálázására úgy, hogy a független változók hatása jó közelítéssel additív legyen. Ezt az eljárást Kruskal MONANOVA-nak (monoton ANOVA) nevezte el.

Az UNICON-eljárás továbbfejlesztése a MONANOVA-rendszernek. Az UNICON-eljárásban lehetőség van az additív modell mellett, amint azt már említettük, a szubtraktív és multiplikatív, illetve ezek kombinációjából álló modellek alkalmazására is.

Legyen $Q = \{p_{jkl}\}$ a függő változó,

$$A = \{\alpha_j\} \quad (j = 1, 2, \dots, m_\alpha)$$

$$B = \{\beta_k\} \quad (k = 1, 2, \dots, m_\beta)$$

$$C = \{\gamma_l\} \quad (l = 1, 2, \dots, m_\gamma)$$

pedig jelöljék a független változókat (kategóriákkal együtt).

Megfigyelési adataink ekkor egy háromutas táblázatba rendezhetők. A táblázat általános eleme legyen $\{z_{jkl}\}$ az A változó j -edik a B változó k -edik és a C változó l -edik kategóriájába (a táblázat $\{jkl\}$ -háromdimenziós cellájába) eső megfigyelési egységeknek a függő változóra vonatkozó átlagos értéke.

Az általános modellben a függő változó értéke a független változók (ismertlen) f függvényével fejezhető ki:

$$Q = z_{jkl} = f(A_j, B_k, C_l).$$

Az additív modell:

$$f(A_j, B_k, C_l) = \alpha_j + \beta_k + \gamma_l.$$

A szubtraktív modell:

$$f(A_j, B_k, C_l) = \alpha_j - \beta_k - \gamma_l.$$

A multiplikatív modell:

$$f(A_j, B_k, C_l) = \alpha_j \beta_k \gamma_l.$$

Egy lehetséges további változat például:

$$f(A_j, B_k, C_l) = (\alpha_j - \beta_k) \gamma_l.$$

Az UNICON-eljárás a független változók lehetséges kategóriáihoz numerikus skálaértékeket rendel:

$$a_j = g_A(\alpha_j)$$

$$b_k = g_B(\beta_k)$$

$$c_l = g_C(\gamma_l),$$

úgy, hogy a felhasználásukkal reprodukált független változó értékek (\hat{z}_{jkl}): (például: $\hat{z}_{jkl} = a_j + b_k + c_l$) maximálisan illeszkedjenek az eredeti értékekhez, a z_{jkl} -hez. Az illeszkedés jószágát az alábbi *Stress* (S_2) függvénnyel mérjük; amely három független változó esetén a következő alakban írható fel:

$$S_2 = \sqrt{\frac{\sum_j \sum_k \sum_l (z_{jkl} - \hat{z}_{jkl(h)})^2 e_{jkl(h)}}{\sum_j \sum_k \sum_l n_{jkl} (z_{jkl} - \bar{z})^2}},$$

ahol:

$z_{jkl} = f(a_j, b_k, c_l)$ az aktuálisan specifikált modell szerint,

$\hat{z}_{jkl(h)}$ = monoton regressziós becslése z_{jkl} -nek a h -ik iterációban,

$e_{jkl(h)} \begin{cases} = 1, & \text{ha } j, k, l, \text{ a } h\text{-ik iterációban sorrendezett} \\ = 0, & \text{ha } j, k, l \text{ hiányzó adat a } h\text{-ik iterációban,} \end{cases}$

$n_{jkl} = \sum_h e_{jkl(h)}$

$$\bar{z} = \left(\sum_j \sum_k \sum_l n_{jkl} z_{jkl} \right) \setminus \left(\sum_j \sum_k \sum_l n_{jkl} \right).$$

A $z_{jkl(h)}$ értéke nem meghatározott, ha $e_{jkl(h)} = 0$, de ez akkor nincs hatással S_2 -re.

Az UNICON-eljárásban S_2 -t gradiens módszerrel minimalizáljuk. A függő változó becült értékeit a Kruskal-féle monoton regressziós eljárással számítjuk. Leírását l. az I. cikkben. (Sigma XV. 3. szám.)

5. Összefoglalás

A három cikkben tárgyalt tucatnyi sokdimenziós skálázó eljárás, eljárás-változat hazai alkalmazói gyakorlatunkban még szinte teljesen ismeretlennek tekinthető. Az alap eljárások esetében ezért nagyobb gondot fordítottunk a kérdésfeltevések gondosabb megfogalmazására, illetve példa készítésére is, más esetekben csak röviden közöltünk egy-egy eljárást. Remélhető, hogy a speciális programcsomagok terjedésével az ismertetett eljárások is egyre szélesebb körben nyernek majd polgárjogot.

A skálázó eljárások több dimenziós output terében az egyes tengelyekhez kapcsolódó eredeti változók a lényegkiemelés problémájának egyfajta megközelítését szolgáltatják, alternatívát adva így a faktoranalitikus vizsgálatokhoz. A megfigyelési egységek térbeli csoportosulásai ugyanakkor a cluster képződés egyfajta realizációját jelenthetik. Ezek a megfontolások táplálják azokat a nézeteket, hogy a skálázó módszerek elmélete keretül szolgálhat a sokváltozós statisztikai eljárások rendszerbe foglalásához.

(Beérkezett: 1985. november 4-én.)

IRODALOM

- CARROLL, J. D.—CHANG, J. J.: *Models and Algorithms for Multidimensional Scaling*. Bell Laboratories, Mimeo. 1973.
- GOWER, J. C.: Some characterisations of matrix multidimensional scaling methods. *Journal of the Royal Statistical Society*, Vol. 49. 1980.
- KRANTZ, D. A.—LUCE, R. D.—TVERSKY, A.: *Foundations of measurement: additive and polynomial representations*. Academic Press, New York 1971.
- KRUSKAL, J. B.: Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, No. 29. 1964.
- KRUSKAL, J. B.: Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the royal statistical society*, No. 27, 1964.
- KRUSKAL J. B.—CARROLL, J. D.: *Geometric models and badness-of-fit functions (from Multivariate Analysis.)* Academic Press, New York 1966.
- SHEPARD R. N.—CARROLL, J. D.: Parametric Representation of Nonlinear Data Structures. in: P. R. Krishnaiah (ed.) *Multivariate Analysis*. Academic Press New York 1966. pp. 561—592.