

A klaszterelemzés alkalmazási lehetőségei a marketingkutatásban

Simon Judit

kandidátus, az MVM vállalati
professzora, a Budapesti
Corvinus Egyetem egyetemi
docense

E-mail: judit.simon@uni-corvinus.hu

A tanulmány célja, hogy a gyakorlati felhasználó számára ismertesse a klaszterelemzés módszerének lépéseit, a gyakorlati adatbázisok elemzésére való felhasználhatóságának problémáit, bemutatva a szakirodalomban ismeretes, esetenként eltérő megoldási javaslatokat. A klaszterelemzés alkalmazása során felmerülő problémákról a szakirodalomban is megoszlanak a vélemények. Ezeket a tanulmány azzal a céllal mutatja be, hogy segítse a felhasználó tájékozódását. A módszertan bemutatását két olyan példa kíséri, amelyek a tudományos igényű, empirikus marketingkutatás területéről származnak, és mivel hasonlítanak a mindennapi gyakorlatban legtöbbször felmerülő problémákhoz, segítségükkel speciális alkalmazási lehetőségek is megismerhetők.

TÁRGYSZÓ:

Klaszteranalízis, klasszifikáció, regressziós fák.
Marketing.

A marketingkutatás egyik leggyakoribb célja a fogyasztók, a megkérdezettek szegmentációja, vagyis olyan csoportok képzése, amelyek belül viszonylag homogének, ugyanakkor egymástól jól megkülönböztethetők (heterogének). Az osztályozás, csoportosítás mind a statisztikai módszertanban nagyon jelentős vizsgálati terület, mind a gyakorlati alkalmazások egyik kiemelt fontosságú kérdése. Az osztályozásról a statisztikában hármass követelményt fogalmaznak meg (*Hunyadi-Vita* [2004]): 1. legyen teljes (a sokaság minden egységének legyen helye valamelyik osztályban), 2. átfedésmentes (minden egyes egység egy és csakis egy osztályban szerepeljen), valamint 3. legyen homogén (az azonos osztályba sorolt egységek jobban hasonlítsanak egymáshoz, mint azokhoz az egységekhez, amelyek más osztályokba vannak sorolva). A marketingkutatásban ugyanezeket az alapelveket szokás az osztályozásra, csoportosításra alkalmazni, a fő figyelmet a homogenitási kritériumra fordítva, mivel ezzel kapcsolatban kell a legtöbb statisztikai és értelmezési követelményt megvizsgálni. A tanulmány célja az, hogy a gyakorló marketinges felhasználó számára ismertesse a csoportképzés egy fontos eszközének, a klaszterelemzés módszerének lépéseit, a gyakorlati adatbázisok elemzésére való felhasználhatóságnak problémáit, kitérve a szakirodalomban ismeretes, esetenként eltérő megoldási javaslatokra. A módszertan bemutatását két olyan, a marketingkutatás területéről származó példa kíséri, amelyek hasonlítanak a mindennapi gyakorlatban legtöbbször felmerülő problémákhoz.

A szegmentáció a marketingkutatásban legtöbbször a magatartási típusok elkülönítésére, megkülönböztetésére szolgál, és mind a magatartáselemzésének fontos eszköze, mind a további stratégiák készítésének megalapozására használható. A csoportképzés a kvantitatív típusú, kérdőíves vizsgálatok esetében végezhető el, amikor elegendő számú és megfelelő módszerrel felvett adat áll rendelkezésre ahhoz, hogy egyes ismérvek szerint csoportokat képezzünk a mintában. A leggyakrabban használt csoportképző ismérvek a fogyasztókat jellemző szocio-demográfiai jellemzők, de lehetséges a használt termék tulajdonságaiból kiindulva is szegmentálni a fogyasztókat. A magatartáskutatás igénye az, hogy minél több jellemző egyidejű figyelembevételével lehessen csoportosítani, és a változók között megjelenhessenek az attitűd típusú, életstílust jellemző, illetve az adott termék kategóriához való viszonyulást kifejező állítások, valamint ezek fogyasztói értékelése.

A szegmentáció céljára a tudományos kutatásokban és a gyakorlati alkalmazásokban is gyakran használt, többváltozós statisztikai módszer a klaszterelemzés, ami a csoportosítás, csoportképzés módszertanának összefoglaló elnevezése. A csoportképzés során a marketingben többnyire fogyasztók csoportjait (*Wind* [1978]), illetve

hasonló termékek csoportjait (*Srivastava–Leone–Schocker* [1982]), vagy tesztpiacok kiválasztását (*Green–Frank–Robinson* [1967]), valamint hasonló stratégiájú vállalatok csoportjait (*Doyle–Saundres–Wong* [1989]) határozzák meg.

A klaszterelemzés többféle csoportosítási eljárást foglal magában (*Green–Frank–Robinson* [1967]). Azon többváltozós módszerek közé tartozik, amelyek az objektumok, a megkérdezettek közötti struktúrák feltárására szolgálnak. A klaszterelemzés célja az, hogy viszonylag heterogén objektumok (egyéni fogyasztók vagy vállalatok) ismérvei alapján viszonylag homogén részsokaságokat, csoportokat tárjon fel és képezzen. A gyakorlati felhasználás szempontjából nagyon fontos jellemzője az, hogy az ismérvek egyidejű figyelembevételével tárja fel a struktúrákat, vagyis nem szükséges előzetesen meghatározni a csoportokat, vagy azok egyes jellemzőit.

A többváltozós statisztikai eljárások alkalmazása több tudományágban – pszichológia, biológia, műszaki tudományok, szociológia, közgazdaságtan, valamint üzleti gazdaságtan – terjedt el, és az alkalmazásokkal, az elterjedt gyakorlattal kapcsolatban különböző kérdések merülnek fel. A klaszterelemzés szakmailag különösen vitatott módszer, a tudományos kutatásban való alkalmazása során több kérdésben térnek el a vélemények. Az alkalmazók részben földrajzi elhelyezkedésük, részben egyes tudományos iskolákhoz való tartozásuk szerint eltérő nézeteket vallanak. Ez a helyzet elsősorban a klaszterelemzés egyes lépései során fennálló választási lehetőségek miatt alakult ki, mivel – mint a többváltozós eljárásoknál gyakran előfordul – a kutatónak nagy a döntési szabadsága.

Csupán egy, bár igen fontos kérdést tekintve egyes szerzők (*Hair et al.* [1998]), elismerik ugyan a gyakorlati alkalmazhatóság jelentőségét, de a módszert leíró jellegűnek, az általánosíthatóságot és a statisztikai következtetés levonhatóságát nélkülözőnek minősítik. Szerintük a módszert exploratív módon lehet használni struktúrafeltárára, kevésbé alkalmas viszont arra, hogy a mintából következtetéseket lehessen levonni az alapsokaságra vonatkozóan. Az adott adatbázison nyert megoldások eltérhetnek egymástól, attól függően, hogy az egyes lépésekben milyen módszert választott a kutató. Érzékeny a módszer arra is, ha az adatbázist új változókkal bővítjük, vagy elhagyunk változókat. Problémaként említik a szerzők, hogy a módszer minden esetben feltár csoportosítási struktúrákat, akkor is, ha ezek a struktúrák nem valósak, de megoldást ebben az esetben is kapunk.

Mivel ez a módszer a hazai tudományos és gyakorlati kutatásokban is gyakran használt (*Füstös et al.* [2004], *Hajdu* [2003]), így a jelen dolgozat első részében az összes módszer részletes bemutatása helyett csupán néhány olyan eljárás rövid összefoglaló jellemzését végezzük, valamint olyan eljárásbeli sajátosságokra hívjuk fel a figyelmet, amelyek a módszer marketingkutatási felhasználása során gyakran előfordulnak.

A tanulmány második részében a módszer alkalmazási lehetőségeit mutatjuk be két példa segítségével. A cikket az összefoglaló következtetések zárják.

1. A klaszterelemzés néhány elméleti kérdése, fontosabb lépései

A gyakorlati példák bemutatása előtt, éppen a gyakorlati alkalmazást elősegítő összefoglaljuk a módszerrel kapcsolatos legfontosabb lépéseket, kétségeket és döntési pontokat.

1.1. Az elemzés előkészítése

A klaszterelemzés összetett statisztikai módszer, amit igen gondosan kell elvégezni. Ezért a módszer tényleges végrehajtása előtt egy sor előkészítő műveletre van szükség. Ezek a következők:

- az objektumok definíciója és számuk meghatározása;
- a kiugró értékek (outlierek) problémájának kezelése;
- a figyelembe vehető ismérvek (változók) számának meghatározása;
- az ismérvek súlyozási módszerének kijelölése;
- az ismérvek összehasonlíthatóságának vizsgálata.

Amennyiben a klaszterelemzést egy mintavétellel nyert adatbázisra végezzük, és az így nyert csoportokat általánosítani kívánjuk az alapsokaságra, akkor biztosítanunk kell, hogy megfelelő mennyiségű elem legyen az egyes csoportokban ahhoz, hogy az egyes alsokaságok megfelelően reprezentáltak legyenek az alapsokaságban. Mivel azonban előre nem tudjuk, hogy mely részsokaságok szerepelnek az alapsokaságban – mivel éppen e csoportok feltárása a klaszterelemzés célja –, így a legnagyobb szükség az ún. kiugró értékek feltárására és eliminálására van. Kiugró értéknek tekintjük azokat az objektumokat, amelyek a többi objektummal való összehasonlításban egy, a tulajdonságok, ismérvek teljesen más kombinációjával rendelkeznek, és így az összes többi elemtől távol helyezkednek el. Ezek az értékek a többi elemmel végzett összevonási eljárást erősen befolyásolják, és így a többi elem közötti összefüggést nehezebb felismerni, illetve ezek a vizsgálat során torzulnak.

Sem a vizsgálatba bevont elemek számára, sem a bevont változók számára nincs egyértelmű előírás. Az alkalmazónak figyelnie kell arra, hogy csak olyan ismérveket vegyen figyelembe a csoportosítási eljárásban, amelyek elméleti megfontolások alapján relevánsnak tekinthetők a vizsgálat tárgya, tartalmi kerete szempontjából. Olyan ismérveket, amelyek a vizsgált jelenség szempontjából jelentéktelennek tekinthetők, ki kell hagyni a csoportosítási eljárásból.

Azt sem lehet előre meghatározni, hogy az egyes ismérvek milyen súllyal vesznek részt a csoportalkotásban, ezért a gyakorlatban minden egyes ismérvre azonos

nagyságú súlyt feltételeznek. Itt elsősorban arra kell figyelni, hogy az egymással magas korrelációban levő ismérvek miatt az összevonási algoritmus során bizonyos aspektusokat ne hangsúlyoznak túl, mert ez megint csak torzítást okozhat. Amennyiben az ismérveknek egyenlő súlyt kívánunk biztosítani és az egymással korrelált kiindulási adataink rendelkezésre állnak, javaslatunk a következő.

1. *Előzetes exploratív faktorelemzés végzése a kiinduló adatokra.* Az exploratív faktorelemzés célja az, hogy a magas korrelációjú, nagyszámú kiinduló adatból független, kevesebb számú faktort állítson elő. Amennyiben a faktorelemzés során ezeket a faktorokat előállítottuk, alapjául szolgálhatnak egy klaszterelemzésnek. Ebben az esetben azonban arra kell figyelni, hogy a faktorértékek értelmezése problémás lehet, emellett a faktorok a kiinduló változók információtartalmának csupán egy részét tükrözik.

2. *A Mahalanobis-távolságfogalom használata.* Amennyiben az elemek közötti távolság meghatározásra a Mahalanobis-távolságot használjuk, akkor a távolságok kiszámítása során a változók közötti esetleges korrelációt kiszűrhetjük. A távolságfogalom használatát a szakirodalom eltérő mértékben ajánlja, egyes vélemények szerint a Mahalanobis-távolságfogalom használatához az adatbázisra olyan előfeltételeseknek kell teljesülniük, amelyek a klaszterelemzéshez használt adatbázisra gyakran nem igazak (*Backhaus et al.* [2003]). Vele ellentétben *Hair et al.* [1998] és *Hajdu* [2003] a korreláció kiszűrésére felhasználható módszerek közül a leginkább ajánlhatónak ezt távolságfogalmat tartják. Szerintük a használat azért nem terjedt el jobban, azért választják leggyakrabban az euklideszi távolságot, mert a számítógépes programok egy része nem tartalmazza a Mahalanobis-távolságot, az euklideszi viszont valamennyinek része. Megjegyezzük, hogy a marketingalkalmazásokban is leggyakrabban az euklideszi távolságfogalmat használják.

3. *Azon változók kizárása, amelyek egymással magas szinten korrelálnak.* Amennyiben két ismerv között magas a korreláció (0,9 vagy efölötti érték), akkor érdemes mérlegelni, hogy ne zárjuk-e ki őket a kiinduló adatbázisból. Azon változó tartalma, amelynek magas a korrelációja egy másik változóval, nagyrészt ezzel a változóval úgyis megjelenik, és így az adott változót redundánsnak tekinthetjük. Ezen magas korrelációjú változók kizárása az egyik legalkalmasabb lehetőség arra, hogy e torzító hatásokat kiszűrjük.

Az alkalmazónak arra is ügyelnie kell, hogy a kiinduló adatbázisban ne legyen olyan ismerv, amelynek konstans az értéke, vagyis olyan, amelynek minden egyes eleme ugyanaz. Ezek az értékek nivellálnák a különbségeket, ami ugyancsak torzítást okozhat az összevonás során. A konstans ismérvek nem megfelelően hatékonyak az elválasztás során, ezért el kell távolítani őket a vizsgált adatok közül (különösen azon ismérveket, ahol ez a konstans nullához közeli érték). Nehézség léphet fel akkor is, mégpedig az implicit súlyozás problémája, ha az egyes kiinduló ismérveket eltérő skálán adják meg. Ezáltal ugyanis az egyes ismérvértékek közötti

különbségek felnagyíthatódnak. Jó, ha az ismérvek túl finom (sok értékkel rendelkező) skálán vannak megfogalmazva. Ahhoz, hogy az egyes változók közötti összehasonlíthatóság a különbségképzéshez fennálljon, standardizálni kell a kiinduló adatokat.

1.2. A klaszterelemzés módszerének gyakorlatorientált bemutatása

Az előző alfejezetben bemutattuk azokat a lépéseket, amelyek az előkészítést szolgálják. Mindezen módszerek átgondolása után kezdődhet a tulajdonképpeni klaszterelemzés, melynek fő lépései a következők.

1. Az adatbázis előkészítése, a hasonlóságok, illetve távolságok meghatározása és kiszámítása. A kiinduló adatbázisban szerepel K objektum, amely általában a megkérdezett személyek vagy vállalatok bizonyos (J) számú ismérvre vonatkozó adatait tartalmazza, amelyek metrikus vagy nem metrikus skálázásúak lehetnek. A metrikus adatokra távolságot, a nem metrikus adatokra hasonlósági mérőszámot számítunk, amelyek alapján a hasonlósági, illetve a távolságmátrix $K \times K$ méretű lesz.

A kiinduló adatbázissal kapcsolatban is felmerülnek olyan problémák, amelyek megosztják a kutatókat. Szakemberek nagy csoportjának véleménye szerint (például *Backhaus et al.* [2003], *Hair et al.* [1998]) mind metrikus, tehát intervallum- és arányskálán mért, mind nem metrikus, tehát nominális skálázású adatokra is jól alkalmazható a módszer, csupán arra kell ügyelni, hogy a hasonlósági vagy a távolságfogalmat használjuk. Ennek némileg ellentmond az a felfogás, hogy lehet ugyan klaszterezni a nem metrikus adatokat is, de ezekre jobban ajánlható a többdimenziós skálázás vagy a látens struktúrák feltárási módszerei (*Kinnear–Taylor* [1973], *Saunders* [1995]). Felmerülhet olyan alkalmazás is, amelyben metrikus és nem metrikus adatok vegyesen szerepelnek a vizsgálatban, azonban ennek előfordulása nem gyakori és nem is ajánlott (*Green–Tull–Albaum* [1988]). A hazai szerzők (*Füstös et al.* [2004]) nem zárkoznak el a nem metrikus adatok használatától, sőt a különböző skálátípuson mért adatok együttes vizsgálatától sem; ebben az esetben skálatranszformációt, illetve dichotomizálást ajánlanak.

A nem metrikus adatoknál használt hasonlósági mértékeket olyan módon képezzük, hogy a bináris változókat (a nominális változók átalakítva) összehasonlítjuk oly módon, hogy az adott tulajdonság mennyire van jelen a két objektum esetében. Így a következő esetek lehetségesek:

- a – mindkét objektumnál jelen van,
- b – csak a 2. objektumnál van jelen,
- c – csak az 1. objektumnál van jelen,
- d – egyiknél sincs jelen mindkét tulajdonság.

A két objektum közötti hasonlóság összehasonlítására általánosan a következő formula használható:

$$S_{ij} = \frac{a + \delta \cdot d}{a + \delta \cdot d + \lambda(b + c)},$$

ahol:

S_{ij} – az i -edik és j -edik objektum közötti hasonlóság,
 δ, λ – a lehetséges (konstans) súlyok.

A mérőszámokat azon az egyszerű példán szemléltetjük, amelyben két mosópormárkát (Tisztító és Fehérítő fantázianévvel) hasonlítunk össze a tulajdonságaik alapján (1 – az adott tulajdonság jellemző, 0 – az adott tulajdonság nem jellemző).

1. táblázat

Mosópormárkák összehasonlítása

| Mosópor | Kiváló tisztítóhatású | Extra fehérítőhatású | Nagyszemcsés | Kék színű | Nagy kiszerelésű | Folyékony kiszerelésű | Óceánillatú |
|----------|-----------------------|----------------------|--------------|-----------|------------------|-----------------------|-------------|
| Tisztító | 1 | 0 | 1 | 0 | 1 | 1 | 0 |
| Fehérítő | 1 | 1 | 1 | 0 | 1 | 0 | 0 |

2. táblázat

Az egyes koefficiens értékei

| A koefficiens neve | Súlyok | | Definíció | A mutató értéke |
|---------------------|----------|-----------|-------------------|-----------------|
| | δ | λ | | |
| Tanimoto (Jaccard) | 0 | 1 | $a/(a+b+c)$ | 3/5 |
| Simple Matching (M) | 1 | 1 | $(a+d)/(a+b+c+d)$ | 5/7 |
| Russel and Rao (RR) | – | – | $a/(a+b+c+d)$ | 3/7 |
| Dice | 0 | 1/2 | $2a/(2a+(b+c))$ | 6/8 |
| Kulczynski | – | – | $a/(b+c)$ | 3/2 |

Forrás: Backhaus et al. [2003], és saját példa.

Az általános képletben különböző súlyok alkalmazásával számíthatók az egyes mutatók, amelyeket a 2. táblázatban foglalunk össze. Példánkban $a = 3$, $b = 1$, $c = 1$, $d = 2$, az egyes koefficiensek értékeit a különböző lehetséges súlyok választása esetére pedig a 2. táblázat tartalmazza.

Ha döntünk a hasonlósági mérőszámról a legfontosabb szempont az, hogy ugyanolyan súlyú-e számunkra egy tulajdonság előfordulásának hiánya, mint amilyen az előfordulás. Abban az esetben, ha az előfordulás ugyanolyan súlyú, mint az elő nem fordulás, akkor az M -koefficiens használata javasolt, ott a számlálóban az elő nem fordulás is szerepel. A szakirodalom általában a Tanimoto-koefficiens használatát ajánlja (*Backhaus et al.* [2003]).

Metrikus változók esetén a hasonlóság helyett a távolságfogalmat célszerű meghatározni. A leggyakrabban alkalmazott távolságfogalom a Minkowski-metrika vagy L -norma, amelynek kiszámítása:

$$d_{kl} = \left(\sum_j |x_{kj} - x_{lj}|^r \right)^{1/r},$$

ahol:

- d_{kl} – a k -adik és l -edik objektum távolsága,
- x_{kj}, x_{lj} – a j -edik változó értéke a k -adik, illetve az l -edik objektumban ($j = 1, 2, \dots, J$),
- $r \geq 1$ – a Minkowski-konstans.

Ha $r = 1$, akkor a city-block metrikát kapjuk, ha $r = 2$, akkor az euklideszi távolságfogalmat. A city-block metrikát elsősorban olyan esetekben használják, amikor különböző telephelyekre vonatkozó ismérvek alapján képzünk szegmenseket. Egyébként a leggyakrabban használt távolságfogalom az euklideszi távolság. A Minkowski-metrikánál nagyon fontos, hogy a változókat azonos skálán mérjük, ezért, ha ez nem áll fenn, akkor az adatokat először standardizálni kell.

A metrikus adatokra is lehet hasonlósági mértéket használni, ez a Q -korrelációs együttható, amely két objektum hasonlóságát az egyes objektumokat jellemző összes változó figyelembevételével számítja ki. A metrikus adatokra számítható távolságfogalom és a Q -korreláció alkalmazására *Backhaus et al.* [2003] a következő szemléletes példát említik: a távolság használatát javasolják olyan esetben, ha két termék forgalmi értékének időbeli hasonlóságát vagy különbözőségét mérjük, és a Q -korrelációs együtthatót akkor, ha a két termék forgalmának időbeli tendenciáját, alakulását hasonlítjuk össze. A Q -korrelációs mutatót azért nem tárgyaljuk részletesen, mivel ritkán fordul elő a gyakorlati alkalmazásokban.

2. A következő lépés annak eldöntése, hogy milyen összevonási (amalgamációs) eljárást válasszunk. Az összevonási algoritmusokat több szempont szerint csoporto-

síthatjuk, az egyik az összevonási eljárás során figyelembe vett változók száma szerint történik. Eszerint van monotetikus és politetikus módszer, ahol a monotetikus módszer használata azt jelenti, hogy az összevonási eljárás során egy változót veszünk figyelembe. A klaszterelemzés nagy előnyének azonban éppen azt tekintjük, hogy egyidejűleg több változót is figyelembe vesz az összevonási eljárásnál és csoportképzésnél, ezért csak a politetikus eljárásokkal foglalkozunk.

Egy másik csoportosítási lehetőség az összevonási eljárás lépései szerint tehető, eszerint a következő eljárásokat különböztetjük meg:

- gráfelméleti eljárások,
- hierarchikus eljárások,
- particionáló eljárások,
- optimalizálási eljárások.

A továbbiakban a marketingkutatói gyakorlatban való elterjedt alkalmazásuk miatt a hierarchikus és a particionáló eljárásokkal foglalkozunk részletesen, a másik két módszer leírása áttekinthető a szakirodalom alapján.

A hierarchikus eljárásoknál kiinduló állapotban minden egyes elem egy önálló klasztert képez. Ezután minden egyes, a vizsgálatba bevont elemre kiszámítjuk az összes távolságot, majd azokat vonjuk össze, amelyeknek a legkisebb a távolsága. Ezt követően a leghasonlóbb csoportokat egy újabb klaszterbe vonjuk össze, ami által a klaszterek száma eggyel csökkent. A következő lépésben megint kiszámítjuk a távolságokat az új és a megmaradt többi klaszter között, ez az előző lépéshez képest redukált távolságmátrixot fog eredményezni. Az összevonás így folytatódik tovább egészen addig, míg az összes elem egy nagy klaszterbe vonódik össze.

Az egyes hierarchikus eljárások abban különböznek egymástól, hogy milyen elv alapján számítják a távolságot az adott objektum és az új klaszter között. Az összevonási kritérium szerint megkülönböztetjük a legközelebbi szomszéd vagy egyszerű lánc (single linkage), a legtávolabbi szomszéd vagy teljes lánc (complete linkage), valamint a Ward-módszert. Néhány összevonási módszer (a centroid, a medián és a Ward-módszer) csak metrikus adatokra számítható.

Az eljárásokat három nagy csoportba sorolhatjuk aszerint, hogy 1. tértágító, 2. térszűkítő vagy konzervatív, vagy 3. térkonzerváló hatásúak.¹ A tértágító eljárás, a legtávolabbi szomszéd fő jellemzője (amelyben a két klaszter összevonása esetén a legtávolabbi pontok közötti távolságot számítják ki, és ezek közül kerül összevonásra az a két klaszter, amelyeknek a legkisebb a távolsága), hogy nagyjából hasonló nagyságú sok csoportot képez. A térszűkítő, a legközelebbi szomszéd eljárás (a távolság a két klaszter legközelebbi pontjai távolságaként kerül meghatározásra) arra törekszik,

¹ Ezek az elnevezések *Füstös et al.* [2005] munkájában megtalálhatók magyarázataikkal együtt.

hogy kevés nagy elemszámú klasztert képezzen, amelyek mellett van néhány kisebb, ami jelentősen különbözik a nagy klaszterektől. Ez az eljárás nagyon alkalmas a „kiugró” (outlier) elemek kiválasztására, elkülönítésére. A térkonzerváló eljárások közé soroljuk a centroid, a medián és a Ward-módszert. A centroid és a medián eljárásoknál a klaszterek összevonásakor a távolságot a két klaszter középpontjának, illetve a pontok mediánjának távolsága jelenti, ennek alapján választják a legkisebb távolságra levőt. A Ward-módszer esetében a klaszteren belüli pontokra kiszámítjuk az átlagot, illetve a pontok átlagtól való négyzetes eltéréseinek összegét, és azt a pontot vagy klasztert vonjuk be a nagyobb klaszterképzéshez, amely bevonásával az eltérés-négyzetösszeg növekménye a legkisebb.

További megkülönböztető ismérve az eljárásoknak a láncképzés. Az ilyen eljárások során elsősorban az egyedi elemek összekapcsolása történik, vagyis sok lépésben mindig további egyedet vonunk be az eddigi klaszterbe. Ezzel viszonylag nagy klaszterek jönnek létre, majd azt a néhányat soroljuk be kisebb klaszterekbe, amik nem kerültek bele az előző nagy csoportokba, elsősorban azért, mert értékük jelentősen különbözik a többiekétől (kiugró érték).

A módszerek abból a szempontból is különböznek, hogy az összevonási eljárás során az alkalmazott, heterogenitást mérő célfüggvény monoton nő-e vagy van-e lehetőség közben korrekcióra, csökkentésre is. Mindezen szempontokat foglalja össze a 3. táblázat.

3. táblázat

A felhasznált módszerek jellemzői

| Módszer | Jellemző tulajdonság | Monoton-e a célfüggvény javítása? | Alkalmazható hasonlósági mérték | Egyéb jellemző tulajdonság |
|---------------|----------------------|-----------------------------------|---------------------------------|---|
| Egyszerű lánc | tér-tágító | igen | bármelyik | láncképzésre alkalmas |
| Teljes lánc | tér-szűkítő | igen | bármelyik | kis csoportok képzésére alkalmas |
| Átlagos lánc | tér-konzerváló | igen | bármelyik | |
| Centroid | tér-konzerváló | nem | távolság | |
| Medián | tér-konzerváló | nem | távolság | |
| Ward | tér-konzerváló | igen | távolság | nagyjából egyforma nagyságú csoportokat képez |

Forrás: Backhaus et al. [2003].

A gazdasági alkalmazásokban igen elterjedt a Ward-eljárás. A német nyelvterületen meghatározó jelentőségű szakirodalom (*Backhaus et al. [2003]; Bergs [1981]*) vizsgálataira hivatkozva állítja, hogy a Ward-módszer a legtöbb esetben nagyon jó felosztásokat eredményez, és az egyes elemek csoportba való sorolása is többnyire ennek bizonyul.

A Ward-eljárás abban az esetben nagyon alkalmas módszer, ha:

- a kiinduló adatbázisnál a hasonlóság mérésére távolságot lehet alkalmazni;
- minden változó metrikus skálán mért;
- az adatok közül kiszűrtük a kiugrókat (ajánlott az egyszerű láncceljárás elvégzése a Ward-módszer előtt);
- a kiinduló változók korrelálatlanok;
- feltehető, hogy az egyes klaszterekbe tartozó elemszám nem nagyon különbözik;
- a klaszterek mérete nagyjából hasonló nagyságú.

A particionáló eljárások feltételeznek egy adott csoportosítást, amelyből indulva egy cserélő algoritmust felhasználva az egyes elemekre jutnak el ahhoz a megoldáshoz, amely megfelel egy kitűzött célfüggvénynek. Az egyes eljárások részben a célfüggvényben, részben a cserélő algoritmusban különböznek egymástól. A particionáló eljárás során az induló megoldás klasztereire minden egyes ismérv átlaga és az eltérésnégyzetek alapján olyan cserét hajtunk végre, amelyben a varianciák minimalizálási kritériuma (az egyes ismérvek varianciájának egyszerű összege) egyre jobban teljesül, majd az eljárás akkor fejeződik be, amikor a variancia már nem csökkenthető tovább. A módszer többnyire csak a lokális optimum elérésekor ér véget, a globális optimum eléréséhez ugyanis az összes elem összes lehetséges besorolását és az ehhez tartozó varianciakritériumot vizsgálni kellene, ami olyan nagy mennyiségű számítást tételez fel, amely gazdaságosan nem végezhető el. Amennyiben például m számú objektumot g számú csoportba sorolhatunk, ez g^m besorolási lehetőséget jelent, ami $m = 10$ és $g = 3$ esetében $3^{10} = 59049$ klaszterképzési lehetőséget ad.

A lehetőségek rendkívül nagy száma miatt az eljárás a lokális optimumok teljesülését tudja csak biztosítani. A megoldás további javítása az induló particionálás változtatásával érhető el. Amennyiben a módosított induló megoldással elvégzett eljárás végén csökken a célfüggvény értéke, vagyis alacsonyabb szórásnégyzetet értünk el, akkor javult a megoldás. Az induló particionálás kiválasztása a kutató feladata, akinek két kérdésről is döntenie kell: részben az induló megoldásban alkalmazott klaszterszámról, részben arról az alapelvről, ami szerint az elemeket beosztja a klaszterekbe. (Ez történhet véletlen számgenerátorral vagy az elemek sorrendje alapján, illetve alapjául szolgálhat egy hierarchikus módszerrel nyert klasztermegoldás is.)

Összehasonlítva a particionáló és a hierarchikus eljárást, szembevetendő különbség az, hogy ha egy adott elemet a hierarchikus eljárásban besorolunk egy klaszterbe, az végig ott marad, míg a particionáló eljárás során még áthelyeződhet máshova. Mind-ezen rugalmasságuk ellenére a particionáló eljárások kevésbé terjedtek el a gyakorlati alkalmazásokban, mint a hierarchikusak. Ez valószínűleg abból következik, hogy a particionáló eljárás nagyban függ az átcsoportosítás során alkalmazott célfüggvény-

től, valamint a kiinduló felosztástól, ami sokszor szubjektívnek tűnik. Azt is a módszer hibájaként említik, hogy többnyire csak lokális optimális értékek elérésére van lehetőség.

3. Az összevonási eljárás elvégzése után, amikor már rendelkezésre állnak az adott eljárással képzett csoportok, dönteni kell a megoldásról, vagyis a *klaszterek számáról*. Az agglomeratív eljárások ugyanis abból az állapotból indulnak ki, amikor minden egyes elem külön klasztert alkot és eljutnak abba az állapotba, amikor az összes elemet egy nagy klaszterbe foglalják össze, ez a két helyzet a szegmensképzés szempontjából egyformán szélsőségesnek tekinthető. A kutatónak kell meghoznia a döntést arról a klaszterszámról, amelyet a legjobb klasztermegoldásnak tekint. Amennyiben nincs valamilyen előzetes vagy elméleten alapuló feltevés az elérendő klaszterszámról, akkor az egyes eljárások esetében különböző statisztikai lehetőségek állnak rendelkezésre. Az egyik az összevonás, ehhez a dendrogram nyújt áttekinthető ábrázolást. A Ward-módszernél követhető az egyes lépésekhez tartozó eltérés négyzetösszeg (vagy a variancia) nagysága, amelynek a klaszterszámok függvényében való ábrázolása megmutathatja azt a töréspontot (Elbow-kritérium), ahol a heterogenitási mértékben „ugrás” következik be (a belső varianciaösszeg jelentősen lecsökken), amely a diagramon törésként jelenik meg. Amennyiben nem jelentkezik jelentős ugrás, vagyis nincs könyök, akkor a „hüvelykujj-szabályt” az 50-50 százalékos belső és külső varianciaarányt érdemes figyelembe venni.

A klasztermegoldás kiválasztását segíti az F -érték is, amelyet az összes változóra ki lehet számítani olyan módon, hogy a j -edik változó g -edik csoportban mért varianciáját viszonyítjuk a j -edik változó teljes mintában mért átlagos varianciájához.² Az F értéke csökken, minél kisebb a számlálóban található csoportra vonatkozó variancia, vagyis annál megfelelőbb az adott csoportban a vizsgált változó homogenitása. Kritériumként fogalmazható meg az elfogadható homogenitási, illetve heterogenitási határra, hogy a csoporton belüli variancia ne haladja meg a változónak a teljes mintában mért átlagos varianciáját, vagyis az F értéke ne haladja meg az 1-et. Egy klasztert akkor tekinthetünk homogénnek, ha az összes klaszterképző változóra kiszámított F -érték kisebb 1-nél.

A klaszterszám megállapítását segíti az értelmezhetőség is: a klaszterek értelmezése a klaszterképző változók t értéke alapján történik.³ A t -érték számítása: a j -edik változó g -edik csoportban mért átlagának és a j -edik változó mintabeli átlagának különbségét osztjuk a j -edik változó becslött mintabeli szórásával. A t -értékek olyan normált értékeket jelentenek, amelyeknél a nulla körüli értékek az adott tulajdonság

² Ismeretes, hogy normális eloszlású változók esetén a két variancia hányadosa elég általános feltételek mellett F -eloszlást követ. Ez az állítás alapozza meg az F -próbák kiterjedt alkalmazását. Itt ennek egy egyszerű változatát láthatjuk.

³ Az előzőkhöz hasonlóan ez esetben a hányados t - (Student-) eloszlást követ.

átlagos „képviselét”, fontosságát mutatják a klaszterben, míg a negatív és pozitív értékek a skála értelmezése szerint átlaghoz képesti alul-, illetve felülreprezentáltságot jelentenek. Az egyes klaszterek összehasonlításakor az egyes változók t -értékei alapján megállapítható, hogy a csoportok valóban lényegesen különböznek-e egymástól, illetve ha különböznek, akkor mely változókban térnek el. Az egyes klaszterek értelmezése is az egyes változók adott klaszterre jellemző t -értékei alapján történhet, annak alapján, hogy mely ismérvek jellemzők a klaszterre, melyek kevésbé. A klaszterszám megállapításához az értelmezhetőség és az egyes klaszternagyságok összehasonlítása is segítséget jelentenek.

További segítség lehet az egyes klaszterek *értelmezésében* a klaszterképzésbe be nem vont változók egyes klaszterekben megjelenő elemzése, aszerint, hogy mennyire különböznek a csoportok az egyéb változók értékei alapján (demográfiai jellemzők, vásárlási szokások). A klaszterelemzés lépéseinek áttekintéséből látszik, hogy a kutató döntései jelentősen befolyásolhatják az eredményt. Ez előnyt jelent abból a szempontból, hogy a klaszterelemzésnek széles körű alkalmazási területet biztosít, mivel a módszerek alkalmazása teszi lehetővé a csoportok létrejöttét. Ugyanakkor kétségtelenül fennáll a kutató által történő manipulálás veszélye, amely kívánatos megoldásokhoz vezethet.

A szubjektív megítélésére az irodalomban szigorú és kevésbé szigorú véleményekkel találkozhatunk. *Backhaus et al.* [2003] művét olyan forrásnak tekinthetjük, amely didaktikus ismertetésével nagyban segíti az egyes módszerek alkalmazását, ugyanakkor a veszélyekre is körültekintően felhívja a figyelmet. A klaszterelemzés esetében azt tanácsolja, hogy a kutató folyamatosan vegye figyelembe és igyekezzen megválaszolni a következő kérdéseket.

1. Melyik hasonlósági mértéket és összevonási algoritmust választotta?
2. Milyen megfontolások alapján választotta ezeket?
3. Mennyire stabilak az eredmények, ha megváltozik:
 - a hasonlósági kritérium,
 - az összevonási algoritmus, és/vagy
 - a kiválasztott klasztermegoldás (vagyis a klaszterszám)?

A kérdések megválaszolása a klaszterek megbízhatósági és érvényességi vizsgálatát tételezi fel. Ez történhet részben olyan módon, hogy elvégezzük az előző kérdésben feltett ismételt eljárást más hasonlósági, illetve összevonási algoritmus-sal, valamint az előzőtől eltérő klaszterszámot is elemzünk. Másrészt lehetőség van az adatok felosztásával történő érvényességvizsgálatra (véletlenszerűen kettéválasztott két minta klaszterátlagainak összehasonlításával), vagy egyes változók elhagyásával megismételhetjük a klaszterelemzést és összehasonlítjuk az eredményeket.

2. Példák a klaszterelemzés alkalmazására

Az eddigiekben áttekintettük a klaszterelemzés alapjait, elvégzésének főbb lépéseit. A továbbiakban két jellegzetes marketingalkalmazás példáján mutatjuk be az elmondottakat.

2.1. A hibrid vásárlói magatartás modellje

A marketing egyik leggyakrabban kutatott területe az egyéni és a szervezeti vásárló magatartása. Az egyéni vásárlói magatartásra vonatkozó számos elmélet közé tartozik a hibrid vásárlói magatartás elmélete is (*Schmalen* [2002] és *Schmalen–Simon* [1998]). Mivel ez az elmélet marketingszakmai körökön kívül nem általánosan ismert, a következőkben megadjuk rövid összefoglalását.

Az a magatartás, amikor ugyanazon fogyasztó bizonyos termékekért több pénzt hajlandó kiadni, míg más termékek esetében odafigyel az árra, a fogyasztó „hibrid” vásárlói magatartásaként foglalható össze, amikor is egy vegyes magatartási típus figyelhető meg. Végeredményben minden fogyasztó, a saját magatartási szokásait követve, egyidejűleg nyitott és kevésbé nyitott pénztárcával vásárol, ami bevásárlási magatartását kiszámíthatatlanná teszi. Általánosságban abból indulhatunk ki, hogy az ember tudatos lény, ezért szükségletei kielégítésekor nem utolsó sorban arra figyel, hogy a kiválasztott termék fontos legyen számára, és megszerzése ne járjon nagy kockázattal. Adott termék fontossá válik, ha:

- a fogyasztó számára jelentős egyéni hasznossága van, ha például fontos funkciókat tölt be, egy hobbihoz kapcsolódik, kíváncsiságot csillapít vagy érzelmeket kelt (hedonista elem);
- a fogyasztói kosár kialakításakor fontos (társadalmi komponens);
- elajándékozható valakinek, aki a fogyasztó számára fontos (szituációs elem).

Az adott vásárlásnál fellépő „szubjektív” kockázat funkcionális kockázattá is válhat: gyakran hiányzik a vásárlók áttekintése és tapasztalata az adott termékkel kapcsolatban, gyakran előfordul, hogy az adott termék technikailag nagyon bonyolultnak tűnik a vásárlónak, amelyet nem tud értelmezni. A vásárlókat jellemző két fő magatartási ismérv – a márkapreferencia és az árelfogadási hajlandóság – a hibrid vásárlói magatartás esetében különböző magatartástípusokban kapcsolódik össze.

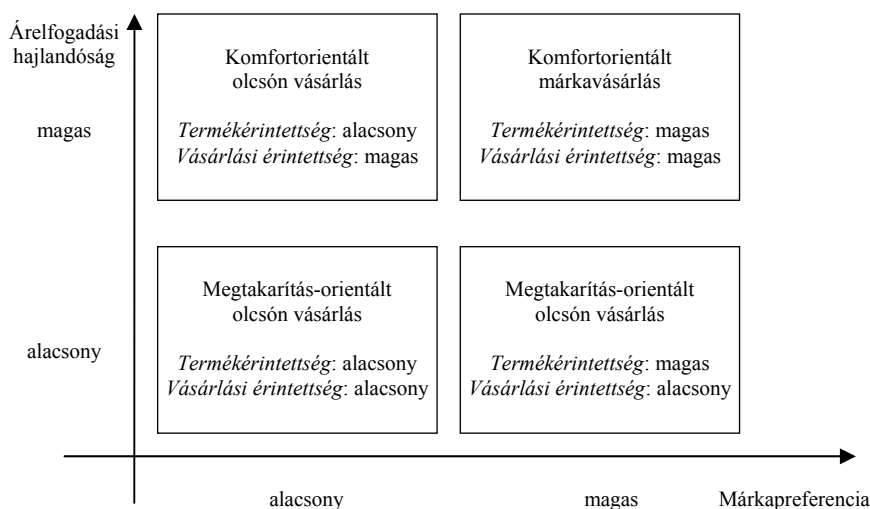
1. *A komfortorientált márkavásárló:* nagyobb árelfogadási hajlandósággal képes magasabb árat adni a márkázott termékért (a magas áron vásárló).

2. *A megtakarítás-orientált márkavásárló:* márkázott terméket preferál, de szeretné a lehető legelőnyösebb árat elérni érte (az előnyös áron való vásárló).

3. *Az olcsón vásárló:* nem márkázott, vagy kereskedelmi márkát vásárló, alacsony árelfogadási hajlandóságú, aki vagy megtakarítás-orientáltan, vagy komfortorientáltan vásárol.

A vásárlói magatartás típusai és a motivációk között teremt kapcsolatot az involvement, vagyis az érintettség az, hogy a vásárló mennyire érintett az egyes termékkel, illetve vásárlási helyzettel kapcsolatban. Az érintettségnek két fő típusa különböztethető meg: a vásárlási és a termékérintettség. A termékérintettség magas, amennyiben a fogyasztó számára valamilyen személyes vagy társadalmi befolyás miatt fontos a termék, míg a vásárlási érintettség akkor magas, ha egy adott vásárlási helyzetben a vásárló által észlelt funkcionális vagy társadalmi kockázat magas. Az egyes vásárlási típusok és az érintettség esetei összekapcsolhatók: a magas termékérintettség, vagyis amikor fontos a termék a vásárló számára általában márkázott termék vásárlásával biztosítható, amikor azonban a termék nem fontos, akkor szívesen vesz a vásárló nem márkás vagy kereskedelmi márkás terméket. Amennyiben a vásárlási helyzetet érzi kockázatosnak a vásárló, akkor ezt a kockázatot az eladótól kapható tanácsadás igénybevételével, valamint nagy termékválasztékkal és szolgáltatásokkal rendelkező bolt választásával csökkentheti. A hibrid vásárlási magatartás kategóriáit, illetve a termék- és vásárlási érintettség összefüggését mutatja az 1. ábra.

1. ábra. A hibrid vásárlói magatartás kategóriái



Konkrét példánkban az akár egyénekenként is különböző termékkategóriákkal kapcsolatos különböző magatartási típusok empirikus kutatását végezte *Schmalen* és *Lang* [1998], akik ugyanazon megkérdezésben tesztelték a sportcipő és a hifi-berendezések kategóriájával kapcsolatos vásárlói magatartást. Az alapfeltevés az, hogy ugyanaz a fogyasztó másként viselkedhet egymástól nagyon különböző termékkategóriák vásárlásakor. Kérdőívünkben mindkét termékkategória esetében olyan, az egyes magatartási típusokat leíró többtétéles skálát, állításhalmazokat fejlesztettek

ki és alkalmaztak, amelyet ötfokozatú skálán értékelték a válaszadók (a kérdőív erre vonatkozó kérdéseit a *Statisztikai Szemle* Interneten megjelenő melléklete tartalmazza [www.ksh.hu/statszemle]).

Az elemzés során azokat a válaszadókat értékelték, akik mindkét állításhalmazra érvényes válaszokat adtak, a hiányzó válaszokkal rendelkezőket kihagyták az elemzésből. Az elemzés különbözik attól az eljárástól, amellyel egy termék kategória vizsgálatakor találkozhatunk (*Schmalen–Simon* [1998]) ugyanerre a magatartástípusra. Ebben az esetben szegmentálni kell a mindkét termék kategóriával kapcsolatos attitűdök alapján, és meg kell találni a kapcsolatot a két csoportba sorolás között. A szerzők a klaszterelemzéshez szükséges korrelálatlan kiinduló változókat faktorelemzés előzetes bekapcsolásával biztosíthatják. Amennyiben ezt a módszert a két állításhalmaz esetében is alkalmazni akarják, szembekerülnek azzal a kérdéssel: vajon a két különböző állításhalmazon a faktorelemzés eredményeként egyező faktorszám lesz-e a megoldás, és a faktorsúlyok struktúrája mennyiben vezet azonos vagy hasonló eredményekhez. Csak hasonlóan értelmezett faktorok esetén lehetséges a két állításhalmazra olyan klasztermegoldásokat kapni, amelyek összehasonlíthatók egymással.

Az empirikus vizsgálatban az előfeltételezés az, hogy a sportcipő és a sztereó be rendezés vásárlására vonatkozóan eltérő magatartás jellemezheti a megkérdezettek jelentős részét. Az állítások az árelfogadási hajlandóságra, a tanácsadással kapcsolatos attitűdre, a márkapreferenciára, illetve a termék- és vásárlással kapcsolatos érintettségére vonatkoztak. Az elemzésben 216 kérdőívet elemeztek, amelyekben az állításoknál nem szerepeltek hiányzó értékek.

Mindkét állításhalmazban 3-3 állítás vonatkozott a márkapreferenciára, valamint az árérzékenységre. Az első lépésben faktorelemzést végezve a két állításhalmazra vonatkozóan, mindkét halmazra azonos struktúrájú faktorsúlymátrix adódott. A faktorelemzés jó elvégezhetőségét a statisztikai mutatók tükrözik: a kiinduló adatbázisra kiszámított KMO-érték⁴ az egyik esetben 0,7 felett van, a másik esetben hajszálnyira marad el tőle. A 0,7-es KMO-érték jó szintű az irodalmi ajánlások szerint (*Backhaus et al.* [2003]). A két faktor előállításakor az állításokban eredetileg megfogalmazott információ tartalmat kifejező teljes magyarázott varianciahányad meghaladja a 70 százalékot.

A két állításhalmazra és a belőlük számított faktorokra mint mérési modellre célszerűnek láttak a szerzők egy ellenőrző jellegű vizsgálatot végezni konfirmatorikus faktorelemzéssel, amelynek eredményét itt nem részletezzük. A konfirmatorikus faktorelemzéssel kiszámított faktorsúlyok mindkét termék kategóriára nagyon jól megfeleltethetők, és jól összhangba hozhatók az exploratív faktorelemzéssel nyert faktorsúlyokkal.

⁴ A KMO-érték a *Kaiser, Meyer és Olkin* által kifejlesztett mérőszám (measure of sampling adequacy). Ez a faktoranalízis kiinduló változóira kiszámított korrelációs mátrix alapján számítható, és egy mutatóval próbálja meg jellemezni a kiinduló változók közötti összefüggést. Az adatbázis megfelelőnek tekinthető a faktorelemzésre, ha a KMO-mutató értéke 0,6 és 1 között van, de minimumkövetelmény az, hogy legyen nagyobb, mint 0,5 (*Backhaus et al.* [2004]).

4. táblázat

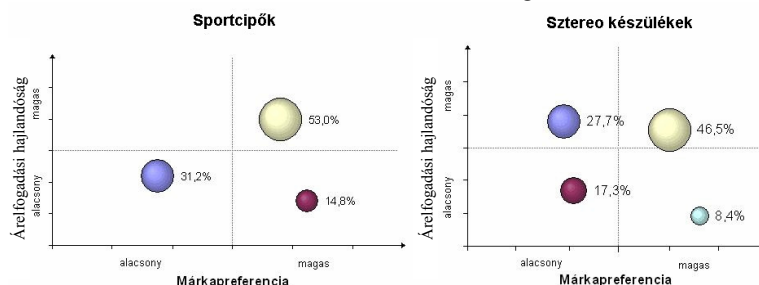
Számított faktorsúlyok hibrid vásárlói magatartás esetén

| Változók | Sportcipő | | Sztereo berendezés | |
|--|------------------|--------------------------|--------------------|--------------------------|
| | 1. faktor | 2. faktor | 1. faktor | 2. faktor |
| Kiinduló változók | márkapreferencia | árelfogadási hajlandóság | márkapreferencia | árelfogadási hajlandóság |
| Márkapreferencia-1 | 0,92271 | 0,08087 | 0,87125 | 0,07443 |
| Márkapreferencia-2 | 0,83364 | 0,10739 | 0,78204 | 0,01727 |
| Márkapreferencia-3 | 0,89432 | 0,11526 | 0,78608 | 0,16241 |
| Árelfogadás-1 | 0,22562 | 0,87551 | 0,09035 | 0,88905 |
| Árelfogadás-2 | 0,11487 | 0,89345 | 0,12852 | 0,88905 |
| Árelfogadás-3 | -0,00887 | 0,73997 | 0,04964 | 0,87462 |
| Mutatók az exploratív faktorelemzésben | (n = 209) | | (n = 209) | |
| KMO | 0,69971 | | 0,71921 | |
| A teljes variancia magyarázott aránya | 75,9% | | 73,2% | |

Forrás: Itt és a 2. ábrán, valamint az 5. táblában Schmalen–Lang [1998].

A szerzők az exploratív faktorelemzéssel kiszámított faktorértékek alapján végezték a klaszterelemzést. A Ward-eljárással elvégzett klaszterezés eredményeként a sportcipő-kategóriára a 3-klaszteres, a sztereo berendezésekre a 4-klaszteres megoldást fogadták el. A belső eltérésnégyzet nagysága a sportcipős 3-klaszteres megoldásnál 45,18 százalék a sztereo berendezéseknél a 4-klaszteres megoldásnál 33 százalék. A faktorátlagokat ábrázolva az egyes klaszterekben az ábrán látható eredményeket kapják. A körök nagysága arányos a csoportokba tartozó megkérdezettek számával. A faktorok átlagos értékeit varianciaanalízissel vizsgálva, az egyes csoportok között szignifikáns különbség tapasztalható ($p=0,0000$). Az egyes klaszterekben az F -értékek 0,5 körül ingadoznak, amely nagyfokú klaszter-homogenitásra mutat.

2. ábra. Klaszterelemzés különböző vásárlói magatartások esetén



A 2. ábrán bemutatott magatartási típusok szerint a sztereó berendezések vásárlása esetén kimutatható mind a négy viselkedéstípus, míg a sportcipők esetében a vizsgált mintában nem mutatható ki a komfortorientált olcsón történő vásárlás.

A kapott klaszterek segítségével mód nyílik az egyes klaszterek közötti váltások vizsgálatára is. Amennyiben a fogyasztónak valóban hibrid a magatartása, akkor a két termék kategóriára vonatkozóan az egyik típusú klaszterből a másik típusúba való átvándorlás lehetne megfigyelhető. Az 5. táblázatból jól látható, hogy valóban megfigyelhető az azonos magatartástípus és az átváltás is egyazon vásárló esetében. A vásárlók mintegy 47 százaléka ugyanazon magatartástípusba tartozik mindkét termék kategória esetében (lásd a diagonális elemek arányát), mintegy 53 százalékuknál viszont magatartásváltás figyelhető meg, vagyis eltérő magatartást mutatnak az egyes termék kategóriáknál.

5. táblázat

Vásárlási típusok sportcipő és sztereó készülékek vásárlása esetén

| | | Sztereó készülékek | | | | |
|-----------|---------------------------------------|---------------------------------------|--------------------------------------|---------------------------------|--------------------------------|----------------|
| | | Megtakarítás-orientált olcsó vásárlás | Megtakarítás-orientált márkavásárlás | Komfortorientált olcsó vásárlás | Komfortorientált márkavásárlás | Összesen |
| Sportcipő | Megtakarítás-orientált olcsó vásárlás | 21 | 3 | 16 | 19 | 59 (30,3%) |
| | Megtakarítás-orientált márkavásárlás | 8 | 8 | 4 | 11 | 31 (15,9%) |
| | Komfortorientált olcsó vásárlás | – | – | – | – | – |
| | Komfortorientált márkavásárlás | 4 | 6 | 33 | 62 | 105 (53,8%) |
| | Összesen | 33 (16,9%) | 17 (8,7%) | 53 (27,2%) | 92 (42,2%) | 195 |

A vásárlók viszonylag magas hányada márkavásárló, mintegy 32 százalékuk mindkét termék kategóriában komfortorientált márkavásárló. A sztereó készülékeknel találunk tanácsadást igénybe vevő, de olcsó készüléket kereső komfortorientált olcsón vásárlót, ami nem jellemző a sportcipővásárlásra. Ugyanakkor a megkérdezettek több, mint fele eltérő magatartást mutat a két termék kategória esetében. Mintegy 10 százalékuk olcsón vásárol cipőt, de a sztereó berendezésért hajlandó magasabb árat adni és márkát vásárolni, míg 17 százalék azon vásárlók aránya, akik márkás cipőt vásárolnak magasabb árelfogadással, de a sztereó készülékeknel a tanácsadással

történő vásárlás mellett olcsón vásárolnak, tehát elfogadják a kereskedelmi márkás vagy nem márkázott terméket is. A vásárlások jelentős részében (53%) megfigyelhető a vizsgált hibrid vásárlói magatartás.

2.2. A tartalomelemzés és a klaszterelemzés összekapcsolása: business-to-business alkalmazás

A szervezeti vásárlások piacán a vásárlások egyik fontos típusa a beszerzés. A szervezeti vásárlók részben vállalati, részben kormányzati, közigazgatási intézmények számára szereznek be. A kormányzati beszerzésekre a business-to-business szakirodalom jóval kevesebb figyelmet fordít, mint a vállalatiakra, így kevés azon tanulmányok száma, amelyek a kormányzati vevők beszerzési magatartását tárgyalják. Szinte alig van olyan kutatás, amely a kormányzati vagy intézményi vásárlók tendervásárlói magatartásával foglalkozna. A következőkben röviden bemutatandó kutatásban, amikor a magyar tendervásárlók beszerzési magatartásának egyes elemeit vizsgáljuk, a nem profitorientált, részben kormányzati szervezetek magatartását elemezzük, mivel a tenderkiírók többsége nem profitorientált szervezet. Vizsgálatunk célja az volt, hogy ezen a piacon tárjunk fel magatartási típusokat.

Magyarországon is, mint az európai országok többségében, kötelező megjelentetni a tenderfelhívást két országos napilapban. Amennyiben közbeszerzési eljárásról van szó, akkor a *Közbeszerzési Értesítőben* kell a felhívást közzétenni. A pályázat minősítésének kritériumai megjelennek ugyan a kiírásban, de kevés kutatás foglalkozik azzal, hogy melyek a tipikus minősítési esetek, milyen a kiíró és a pályázó közötti kapcsolat, illetve mi a pályázatok sikerességének vagy sikertelenségének a magyarázata.

Kutatásunkban (*Mandják–Simon* [2000]) tenderkiírások szövegét elemeztük abból a célból, hogy megállapítsuk a kiírások tipikus elemeit, jellemzőit, valamint a meghirdetett döntési kritériumokat. Az Observer sajtófigyelő ügynökség segítségével 1996 márciusától összegyűjtöttünk olyan tenderkiírásokat, amelyekre tartalomelemzést végeztünk. A vizsgált adatbázis az 1996. május–június folyamán a magyar napilapokban publikált tenderfelhívások (mintegy 515) szövegéből keletkezett, a szöveg elemekre bontásával képeztük a vizsgált változókat. A következő változókat határoztuk meg a felhívások alapján:

- a közlés adatai (melyik napilapban jelent meg),
- a tender típusa (a kiíró típusa szerint megkülönböztetve: önkormányzati, államigazgatási, egyéb intézményi és vállalati),
- a kiíró jellemzői (neve és az intézmény típusa),
- a kiírás tárgya (termék, szolgáltatás vagy projekt),
- a beadási határidő,

- az elvárt garanciák,
- az értékelési kritériumok,
- a szállítási időpontok,
- a kiírási szöveg hossza,
- a vevő és eladó szavak említésének száma a szövegben,
- a vásárló szakmai jellemzői.

Az egyes szövegeket a változók alapján feldolgozva, a kvalitatív típusú tartalom-elemzést kvantifikálni tudtuk, így a vizsgálatban is használhattunk kvantitatív elemzési módszereket. Az értékelési kritériumok alapján próbáltuk meg feltárni a vásárlók preferenciáit, illetve ezek alapján a tipikus vásárlói magatartási formákat. A tenderkiírásokban a legtöbb esetben említettek értékelési kritériumokat, többnyire nem csak egyet, hanem többet is. Összegyűjtve az értékelési szempontokat, mintegy 12 olyat találtunk, amelyek több mint 50 kiírásban is említésre kerültek, ezért ezeket vontuk be a vizsgálatba. A szempontok fontosságára önkényes feltételezést tettünk, amely szerint az egyes kiírásokban szereplő szempontok egymásutánisága egyben fontossági rangsort is tükröz, vagyis a kiírók által legfontosabbnak tekintett szempont az első helyen szerepel, a második legfontosabb a második helyen és így tovább. A szempontok ilyen valóságos rangsorolását pontozással érzékeltettük, amelyben az első helyen említett, legfontosabb szempont 12 pontot kapott. A második helyen említett, második legfontosabb szempont 11 pontot kapott, és így minél hátrébb volt említve egy szempont, annál kevesebb pontot kapott. Azokban a kiírásokban, ahol egy adott szempontot nem említettek, az a szempont 0 pontot kapott. A változókat tehát a pontok szerint értékelve a szempontokat metrikus változóként kezeltük. Ezután a 12 változóból faktoranalízissel képeztünk 5 faktort, ahol a teljes magyarázott változóhányad mintegy 60 százalékos volt. Az 5 faktorra mint kiinduló változóra végeztünk klaszterelemzést, a kiugró elemeket (8 elem) elimináltuk az egyszerű láncmódszerrel, majd a klasztereket Ward-módszerrel képeztük. A klaszterezés eredményeként a hétklaszteres megoldást találtuk stabilnak és ezt értelmeztük. A klaszterek értelmezése a következő:

1. klaszter: A vásárlók, akik valamilyen preferenciával rendelkeznek (14%).

Ebben a szegmensben a vásárlók az árat és az eladó referencialistáját említették mint legfontosabb tényezőt. A többi tényező jóval kevésbé fontos számukra, ugyanakkor az említett szempontok sem szerepelnek nagyon magas fontossággal, vagyis vannak preferenciáik, de azoknak nem tulajdonítanak kiemelkedő fontosságot. Ebben a szegmensben az önkormányzati vevők vannak túlsúlyban és a vásárlás tárgya legtöbbször valamilyen szolgáltatás.

2. klaszter: A hagyományos értékeket figyelembe vevők (21%).

Az ebbe a szegmensbe tartozó vevők olyan hagyományos szempontoknak tulajdonítanak kiemelkedő jelentőséget, mint a minőség és a fizetési feltételek. A vásárlás

tárgya a legtöbb esetben termék, és kissé nagyobb a vállalatok aránya, mint az önkormányzatoké vagy más állami intézményeké.

3. klaszter: A beszállító értékeire figyelők (11%).

Ebben a szegmensben a beszállítóra jellemző olyan szempontok, mint a technikai képességek, a stabilitás, a megbízhatóság és a referencialista a legfontosabbak a döntésnél. A vásárlás tárgya túlnyomórészt projekt, és a résztvevők nagyrészt állami intézmények.

4. klaszter: Az előírásokat közvetlenül alkalmazók (8%).

Ezek a vevők nem fogalmazzák meg saját döntési kritériumokat, hanem a törvényben megfogalmazott előírásokat alkalmazzák a vásárlási döntés meghozatalakor. A törvényben szabályozott szempontok közül legfontosabbnak az árat, a technikai paramétereket, a szállítási időt és a referencialistát tartják.

5. klaszter: A büntetésre figyelők (10%).

Ebben a szegmensben a büntetés, a kötbér kikötése a legfontosabb szempont, de megemlítik emellett az árat, a műszaki paramétereket és a referenciákat is. Elsősorban intézmények tartoznak ide, akik többnyire projekteket vásárolnak.

6. klaszter: A kiírás tárgyára figyelők (17%).

A vevők azokra a tényezőkre figyelnek elsősorban, amelyek az adott termék, szolgáltatás vagy projekt vételével függnek össze, mint a szállítási határidő, a referenciák, a garanciák és az ár. A szegmensre jellemző az állami és egyéb intézmények kissé magasabb aránya.

7. klaszter: Azok a vevők, akik nem közölnek értékelési szempontot (19%).

Erre a csoportban az a jellemző, hogy nem közölnek a tenderkiírásban értékelési szempontot, még a törvényben előírtakat sem.

Az egyes klaszterek értelmezése alapján találhatunk egymáshoz hasonlókat abból a szempontból, hogy a vásárlási folyamatot elsősorban tranzakcióként kezelték-e vagy inkább a kapcsolat jellege dominált a kritériumok között. E szempontok szerint a hét szegmenst két nagy típusba soroltuk be, aszerint, hogy a vásárlási döntésnél inkább a vevő és a beszállító közötti kapcsolatot vagy magát a vásárlás, a tranzakció jellemzőit tartja-e fontosnak és meghatározónak. A kapcsolatorientált típusba a mintában található vásárlók összesen 25 százaléka, a tranzakcióorientált típusba mintegy 56 százaléka tartozik, míg 19 százalékuk nem jelölt meg elbírálási szempontot.

A kutatás eredményeként kapott adatok jól interpretálhatók a business-to-business marketing területén, megfelelnek azoknak, a gyakorlatban megfigyelt tapasztalatoknak, hogy ezen a területen a tényleges cserefolyamattal kapcsolatos szempontok figyelembevétele jellemző ugyan, de egyes, vásárlásokkal kapcsolatos döntésekben (elsősorban a nagyobb értékű, projektformában megvalósuló szolgáltatások esetében, valamint a hosszabb üzleti kapcsolatok keretében megvalósuló döntéseknél) a vevő és a beszállító közötti kapcsolatoknak van döntő szerepe.

3. Következtetések

A dolgozatban a klaszterelemzés marketingkutatásban való felhasználási lehetőségeire igyekeztünk rámutatni legfőképpen azzal a céllal, hogy a szegmentáció szükségességére és módszertani lehetőségeire felhívjuk a figyelmet. A módszertan bemutatása felhasználó-orientáltan történt, ami azt jelenti, hogy a módszer statisztikai hátterét, precíz leírását és bemutatását mellőztük. A módszer tárgyalásakor hangsúlyosan szövegtünk a felhasználás feltételeiről, azzal a céllal, hogy a gyakorlati felhasználó számára segítséget nyújtsunk. Mivel a klaszterelemzés alkalmazása során több olyan problémával lehet találkozni, amelyről a (felhasználóorientált) szakirodalomban is megoszlanak a vélemények, igyekeztünk bemutatni a különböző irányzatokat olyan módon, hogy véleményt, orientációs lehetőséget is felkínáltunk a felhasználónak. A módszertan alkalmazását bemutató két példa kiválasztásában az a szempont játszott a legnagyobb szerepet, hogy két, a piackutatási gyakorlatban legtöbbször előforduló alkalmazás esetétől némileg eltérő, speciális példát kívántunk ismertetni. A kéttermékcsoportos magatartásvizsgálatban az a különleges, hogy a szegmentációt ritkán szokták összekapcsolni két termékcsoporttal kapcsolatban ugyanazon vásárlóra. A hibrid vásárlói magatartás elmélete, amely szerint adott vásárló eltérő típusú magatartással rendelkezhet más-más termékcsoportban, jól alátámasztható azzal a szegmentációval, amit a két termék kategóriára vonatkozó ismervek alapján végzünk. A klaszteranalízis alkalmas arra is, hogy a magatartási típusokat az empirikus eredményekben összehasonlítsuk.

A business-to-business kapcsolatokban igen gyakori a tenderkiírás alapján való beszerzés, a kutatás azonban mégis viszonylag kevés figyelmet szentel ilyen döntések elemzésének. A példa kiválasztásakor részben egy kevésbé kutatott területtel kívántunk foglalkozni, részben azt a módszertani megoldást akartuk bemutatni, amelyben a tartalomelemzést kvantitatív elemzéssel kapcsoltuk össze. A tenderkiírásokban előforduló döntési kritériumok alapján jól megkülönböztethető szegmenseket sikerült kimutatni, amelyek megfeleltek mind a módszertani követelményeknek, mind az alkalmazási területen használt fogalmaknak és tipológiáknak. Mindkét példa bemutatásával azt igyekeztünk érzékeltetni, hogy a többváltozós statisztikai módszertan, jelen esetben a klaszteranalízis alkalmazása milyen jelentős többlet-információkat szolgáltat a gyakorlat egyes területein végzett adatelemzéseknél, valamint azt is, hogy a módszertan szigorú követelményei összeegyeztethetők a gyakorlatban általában nem teljes mértékben teljesülő feltételekkel.

Irodalom

- BACKHAUS, K. ET AL. [2003]: *Multivariate Analysemethoden*. 10. Auflage, Springer Verlag, Berlin.
BERGS, S. [1981]: *Optimalität bei Cluster-Analysen*. Diszertáció. Münster. (Munkaanyag.)

- DOYLE, P. – SAUNDERS, J. – WONG, V. [1989]: International marketing strategies and organization: a study of US, Japanese and British competitors. *American Marketing Research*. 12. évf. 11. sz. 408-413. old.
- FÜSTÖS L ET AL. [2004]: *Alakfelismerés*. UMK. Budapest.
- GREEN, P. E. – FRANK, R. A. – ROBINSON, P. J. [1967]: Cluster analysis in test market selection. *Management Science*. 13. sz. 387–400 old.
- GREEN, P. E. – TULL, D. S. – ALBAUM, G. [1988]: *Research for marketing decisions*. Englewood Cliffs, NJ, Prentice Hall. New York.
- HAIR, J ET AL. [1998]: *Multivariate data analysis*. Prentice Hall. New York.
- HAJDU O. [2003]: *Többváltozós statisztikai számítások*. KSH. Budapest.
- HOFMEISTER TÓTH Á. [2003]: *Fogyasztói magatartás*. Aula Kiadó. Budapest.
- HOFMEISTER Á. – SIMON J. – SAJTOS L. [2003]: *A fogyasztói elégedettség*. Alinea Kiadó. Budapest.
- HUNYADI L. – VITA L. [2004]: *Statisztika közgazdászoknak*. KSH. Budapest.
- KINNEAR, T. C. – TAYLOR, J. R. [1973]: The effect of ecological concern on brand perceptions. *Journal of Marketing Research*. 10. évf. 5. sz. 191–197. old.
- MALHOTRA, N.K. [2002]: *Marketingkutatás*. KJK Kerszöv Kiadó. Budapest.
- MANDJÁK T. – SIMON J. [2000]: Hungarian tender buyers: some results of an empirical research study. In: *Woodside* (szerk.): *Getting better at sensemaking*. Jai Press Inc. Stamford, Connecticut. 385–398. old.
- NUNALLY, J.C. [1994]: *Psychometric theory*. McGraw-Hill, Inc. London.
- SAUNDERS, J. [1995]: Cluster analysis. In: *Hooley, G. J.– Hussey, M. K: Quantitative methods in marketing*. The Dryden Press. London.
- SCHMALEN, H. [2002]: *Általános üzleti gazdaságtan*. Axel Springer Budapest Kiadó. Budapest.
- SCHMALEN, H. – LANG, H. [1998]: Hybrides Kaufverhalten und das Definitionskriterium des Mehrproduktfalls. In: *Marketing*. 20. évf. 1. sz. 5–13. old.
- SCHMALEN, H. – SIMON, J. [1998]: A hibrid vásárlói magatartás és a kereskedelemre levonható következtetések. *Marketing és Menedzsment*. 31. évf. 6. sz. 67–72. old.
- SIMON J. – LANTOS Z. [2004]: A betegelégedettség mérési módszere és az empirikus vizsgálatok tapasztalatai. In: *Berács et al.: Marketingelmélet a gyakorlatban*. KJK Kerszöv Kiadó. Budapest.
- SPÁTH, H. [1977]: *Cluster-analyse Algorithmen zu Objektklassifizierung und Datenreduktion*. 2. Auflage, R. Oldenbourg-Verlag. München.
- SRIVASTAVA, R. K. – LEONE, R. P. – SHOCKER, A. D. [1981]: Market structure analysis: hierarchical clustering of products based on substitution-in-use. *Journal of Marketing*. 45. évf. 3. sz. 38–48. old.
- STEINHAUSEN, D. – LANGER, K. [1977]: *Clusteranalyse*. Walter de Gruyter Verlag. Berlin.
- WIND, Y. [1978]: Issues and advances in segmentation. *Research, Journal of Marketing Research*. 15. évf. 8. sz. 317–337 old.

Summary

In this study applications of cluster analysis to marketing problems are reviewed. Issues and problems related to the use and validation of cluster analytic methods are also discussed. Altern-

tive methods of cluster analysis are presented and evaluated in terms of theoretical background described in the literature and recent empirical works on their performance characteristics. Two empirical studies are presented conducted on scientific research problems which are relevant to the practical research as well. The approach of these studies helps to understand some special applications of the cluster analysis.

A hibrid vásárlói magatartást leíró állításhalmaz

| |
|---|
| Sportcipővásárlás: márkapreferencia |
| Sportcipővásárlásnál az ismert márkákat részesítem előnyben |
| Egyes olcsó márkák a sportcipő vásárlásakor szóba sem jöhetnek |
| Sportcipő esetében teljesen mindegy nekem, milyen márkájú |
| Sztereókészülék-vásárlás: márkapreferencia |
| Sztereókészülék-vásárlásnál az ismert márkákat részesítem előnyben |
| Egyes olcsó márkák a sztereó készülék vásárlásakor szóba sem jöhetnek |
| Sztereó készülék esetében teljesen mindegy nekem, milyen márkájú |
| Sportcipő: árelfogadási hajlandóság |
| Igen, szakszerű tanácsadásért hajlandó vagyok valamennyit fizetni is |
| Inkább fizetek valamit a tanácsadásért, csak hogy a számozásról szakszerű eligazítást kapjak |
| Sportcipővásárlásakor nem vagyok hajlandó a márkaboltban levő választékért, tanácsadásért vagy valamilyen szolgáltatásért többet fizetni |
| Sztereó készülék: árelfogadási hajlandóság |
| Igen, szakszerű tanácsadásért hajlandó vagyok valamennyit fizetni is |
| Inkább fizetek valamit a tanácsadásért, csak hogy a készülékről szakszerű eligazítást kapjak |
| Sztereókészülék-vásárlásakor nem vagyok hajlandó a márkaboltban levő választékért, tanácsadásért vagy valamilyen szolgáltatásért többet fizetni |

| |
|---|
| Sportcipővásárlás: termékérintettség |
| A sportcipők nem különösebben érdekelnek engem |
| A sportcipők mind egyformán néznek ki |
| A sportcipővásárlást inkább tehernek tartom |
| Sztereókészülék-vásárlás: termékérintettség |
| A sztereó készülékek nem különösebben érdekelnek engem |
| A sztereó készülékek mind egyformán néznek ki |
| A sztereókészülék-vásárlást inkább tehernek tartom |
| Sportcipő: vásárlási érintettség |
| A sportcipő vásárlásánál könnyen mellé lehet fogni, ami a megfelelő formát illeti |
| A sportcipő vásárlásakor könnyű a rossz választás |
| Sztereó készülék: vásárlási érintettség |
| A sztereókészülék-vásárlásánál könnyen mellé lehet fogni a hangzás minőségét illetően |
| A sztereókészülék-vásárlásakor könnyű a rossz választás |

Forrás a SCHMALEN, H. – LANG, H. [1998]: Hybrides Kaufverhalten und das Definitionskriterium des Mehrproduktfalls. In: *Marketing*. 20. évf. 1. sz. 5–13. old. található kérdőív alapján.