

Az életstílus-alapú szegmentáció módszertani kihívásai¹

**Hornyák Miklós – Kehl Dániel –
Kruzslicz Ferenc – Szűcs Krisztián**
Pécsi Tudományegyetem

A fogyasztói szegmentációkutatás számos szempontból kihívást jelent a kutatóknak: módszertani és tartalmi vonatkozások tekintetében is felmerülhetnek olyan kérdések, amelyek megoldása elengedhetetlen ahhoz, hogy releváns, megfelelően értelmezhető csoportokat kapjunk eredményül. Jelen tanulmányban összegyűjtöttünk olyan problémákat, megoldandó feladatokat, amelyekkel szembesültünk az elemzések során és bemutatjuk az ezekre adott válaszainkat, illetve vizsgáljuk a kidolgozott megoldások eredményességét, elfogadhatóságát is.

Kulcsszavak: életstílus, yea-saying, statisztikai korrekció, dimenzióredukció, klaszterezés

BEVEZETÉS

Ebben a tanulmányban az életstílus-alapú szegmensek elemzésére koncentrálunk, amelynek kiindulópontját az Életstílus Inspiráció modell (Töröcsik 2011) szolgáltatta. Ennek az elméleti konstrukciónak az alapja egy kétdimenziós rendszer, amelyet az „élettempó” és az „értéorientáció” alakít. Ez a két látens változó határozza meg a vizsgált alapsokaságon belül azokat a szegmenseket, amelyek egymástól jelentős mértékben eltérnek, a csoporttagok viszont – életstílusukat tekintve – homogének. Az elemzések során így kettős cél fogalmazódott meg:

1. egyrészt rekonstruálni szeretnénk volna az elméleti rendszert (amelyet elsőként 2004-ben, aztán 2005-2008 között több alkalommal is megtettünk) a dimenziók mérésére kifejlesztett attitűdállítások segítségével,
2. másrészt azonosítani kellett a rendszerben megjelenő csoportokat, azok helyzetét és méretét.

A TÁMOP-4.2.3-12/1/KONV-2012-0016 projekt keretében nyílt lehetőségünk arra, hogy a 15-24 éves magyarországi fiatalok fogyasztói magatartását, életstílus-csoportjait egy komplex vizsgálat során megismerjük. A kutatás kiemelt eleme a 2.000 fős, országos mintán elvégzett (CAPI) adatfelvétel, illetve az abból előállított adatállomány volt.

Terjedelmi korlátok miatt nem foglalkozunk részleteiben az adatfelvétel során, illetve az azt megelőző potenciális hibaforrásokkal, azok kiküszöbölésének lehetőségével, pedig a téma vizsgálatához széles irodalmi bázis áll rendelkezésre, akár magyar nyelven is (Kish 1989, Hunyadi 2001). Röviden összefoglalva a kritikus pontokat azt láthatjuk, hogy az alábbi három tipikus döntési pont jellemző a kutatások tervezése során:

- az adatfelvételi technika helyes megválasztása: kezdetben a postai és telefonos adatfelvétel hatékonyságát mérték össze a kutatók (Yammarino et al. 1991, Kaldenberg et al. 1994), majd a 2000-es évek elején

megjelentek az online megkérdezések tapasztalataival kapcsolatos beszámolók (például Porter és Whitcomb 2003, Shih és Fan 2008, Galesic és Bosnjak 2009),

- a kutatási eszköz megfelelő felépítése, vagyis a kérdőív struktúrájának és hosszúságának befolyása a válaszadási hajlandóságra: ezzel a témakörrel is jelentős számú tanulmány foglalkozik, mind az adatfelvételi technikával (Pintér és Kátay 2010), mind a megkérdezés témájával (Vicente és Reis 2010, Lynn 2014) összefüggésben,
- és végül, de nem utolsó sorban fontos megismerni az alapsokaságot, a potenciális válaszadókat abból a szempontból is, hogy milyen hajlandóságot mutatnak a kérdőívek megválaszolását illetően: számos tanulmány született az 1980-as években (Kaldenberg et al. 1994), amely az életkor és a válaszadási hajlandóság közötti összefüggéseket vizsgálta és az akkori tapasztalatok azt mutatták, hogy mind a személyes, mind a telefonos megkérdezések esetében az idősebb korosztály tagjai alacsonyabb mértékű együttműködésről tettek tanúbizonyságot.

Vagyis érzékelhető, hogy a jól megtervezett kutatás ellenére a célcsoport széles körben történő megismerése jelentős korlátokba ütközhet, mind az adatfelvételi módot, mind a válaszadási hajlandóságot, vagy éppen a konformizmust tekintve.

A vizsgálat során az előzetesen tesztelt 44 attitűdállítást öt-, illetve hatfokozatú Likert-típusú skálán kellett értékelniük a válaszadóknak, amelyből az értékmérés a Schwartz által használt 21 értékváltozó (Keller 2008) használatával végeztük el.

A tisztított, labelezett adatállomány vizsgálatok derült ki, hogy az attitűdállításokra adott válaszok eloszlása jelentősen eltér a normálistól, több olyan változót is találtunk, amely eredeti állapotában nem volt alkalmas arra, hogy a csoportképzésben érdemben részt vegyen. Azt találtuk, hogy az adatállomány mérési problémákat vethet fel, ezért ezt a témakört jártuk körbe,

illetve egy egyszerű utólagos statisztikai korrekciót végeztünk el, amit a következő fejezetben mutatunk be.

A MÉRÉS PROBLÉMÁI

A mérés pontosságának biztosítása a természettudományok esetén sem egyszerű feladat, hatványozottan igaz ez azonban a társadalomtudományokban, ahol a mérés tárgya jellemzően egy egyén adott tulajdonsága. A mérési hibák lehetséges okainak és a kezelési módszerének megértéséhez röviden tekintünk át azokat. A kérdőíves kutatásokkal foglalkozó kézikönyvek gyakran négy nagy csoportra bontják a potenciális hibákat (Grover-Vriens 2006):

- *Sampling error*: minden mintavétel hibát okoz, amit úgy kell értenünk, hogy pontatlanabb információink lesznek a sokaságról, mintha az egész sokaságot megkérdeztük volna, ez a fajta hiba nem védhető ki.
- *Coverage error*: ez a hiba akkor következik be, ha a mintába potenciálisan bekerülő egyének nem fedik le a teljes sokaságot, ezt a típusú hibát a mintavételi terv gondos elkészítésével meg tudjuk előzni.
- *Nonresponse error*: a nem válaszolási hiba elsősorban akkor fordul elő, ha a potenciális válaszadók egy részét nem vagyunk képesek elérni, illetve azok megtagadják a válaszadást. Komolyabb problémát akkor okoz, ha mértéke a 10%-ot jelentősen meghaladja, illetve ha a nem válaszolás szisztematikus. Erre utaló jeleket az adatokban nem tapasztaltunk, illetve a mintanagysághoz képest elenyésző a nem válaszolás mértéke.
- *Measurement error*: a mérési hiba a válaszadó adott változóhoz tartozó tényleges, valódi értéke és a kérdőív segítségével kapott érték közötti különbséget jelöli. Természetesen ez a leggyakrabban előforduló, és egyben a legnehezebben felismerhető és kiküszöbölhető hibaforrás. Szisztematikus mérési hibát okozhat például maga

a válaszadó, illetve a megkérdezés módszere (pl. személyes interjú, más eredmények születnek, mint telefonos megkereséssel, stb.) is, bár ezeket az okokat sokszor nehéz szétválasztani. Az ilyen jellegű hibákat a külföldi szakirodalomban sok esetben common method variance-nek nevezik, ami a 60-as évek óta komoly kutatási terület. Nagyon széleskörű összefoglalót ad a területről Podsakoff és szerzőtársainak munkája (Podsakoff et al. 2003).

A mérési hibák, illetve a válaszadói torzítások igen gyakoriak, ennek ellenére ezeket sok esetben a kutatók nem veszik figyelembe. A hibák sokféle torzulást okozhatnak, a mostani vizsgálat szempontjából azonban a változók közötti (korrelációs) kapcsolatokra gyakorolt hatásra koncentrálnunk. Különböző esetekben a torzítások miatt első- és másodfajú hibát is elkövethetünk, azaz némely esetben olyan változók között is mutathatunk ki korrelációt, melyek a valóságban függetlenek, illetve fordítva: a valóságban létező kapcsolatot fedhet el a torzítás. Az ilyen szisztematikus hibák akár azt is okozhatják, hogy a két változó közötti tényleges és a mért korreláció eltérő előjelű. A hiba mértéke sok tényezőtől függ, de Podsakoff és társai tanulmánya alapján az attitűd állítások esetén a legnagyobb az ilyen jellegű torzítás (Podsakoff et al. 2003, 880). A leggyakrabban említett és a legtöbb kutatást indukáló válaszadói torzítások az alábbiak:

- *Társadalmilag elfogadott válaszadás* (socially desirable responding): az a tendencia, mely szerint a válaszadók az érvényes társadalmi normák alapján jobb színben tüntetik fel magukat, mint a valóság. Tipikus példák az adócsalással, alkoholfogyasztással, rasszizmussal kapcsolatos kérdések.
- *Beleegyező válaszadás* (acquiescent responding): az általános definíció szerint ez a torzítás azt a tendenciát jelenti, hogy a válaszadók egyetértenek az állításokkal, függetlenül azok tartalmától

(Winkler et al. 1982). A jelenség egyéb elnevezéseken is ismert: egyetértési tendencia (agreement tendency), igenlés (yea-saying), illetve pozitívizmus (positivity). Megjegyzendő, hogy a beleegyező válaszadásnak ismert, de kevésbé gyakran előforduló ellentéte a nay-saying, vagy disacquiescent responding.

- *Extrém válaszadás* (extreme responding): azt a válaszadási mintázatot jelenti, melyben a válaszadó a kérdés tartalmától függetlenül az extrém (legalacsonyabb, vagy legmagasabb) értékeket választja. Az ilyen típusú válaszadásnak sok oka lehet, kutatások szerint kisgyermek és a téma iránt erős érzelmet mutatók könnyen esnek ennek a torzításnak áldozatául.
- *Középre húzó válaszadás* (midpoint responding): ez a válaszadási mintázat tulajdonképp az extrém válaszadás ellentéte, a középső lehetőség választása, a szóban forgó kérdés tartalmától függetlenül. Ennek a torzításnak az oka általában az óvatosság, bizonytalanság, vagy az ismeretek hiánya.

Természetesen a bemutatottakon kívül is vannak egyéb, potenciális veszélyforrások, a fentiek kezelése azonban elengedhetetlenül fontos a megfelelő következtetések levonása érdekében. A következőkben a rendelkezésünkre álló adatokon keresztül vizsgáljuk meg az egyes torzító tényezők előfordulását, illetve próbáljuk azok káros következményeit kiszűrni.

Ahogy azt már említettük, a kérdőív sok egyéb változó mellett az élettempóra és az értékorientációra utaló méréseket is tartalmazott. Az élettempó mérésére a kérdőívben a v18 jelű kérdéscsoport szerepelt, amely „Mennyire értesz egyet a következő állításokkal?” formában, a v19: „Mennyire tartod fontosnak az alábbi tulajdonságokat?” kérdéscsoport pedig az értékorientációt vizsgálta. Ezek vizsgálata során azt állapítottuk meg, hogy a fenti válaszadói torzítások közül a beleegyező

válaszadás (acquiescent responding) erőteljesen megfigyelhető. Az első ábrán két, különböző változóra vonatkozó, tipikusnak mondható hisztogramot mutatunk be.

Az egyes kérdésekre adott válaszok magas átlaga mellett problémát okoz az is, hogy a változók közötti korrelációk is torzítottak lesznek, amikből képzett korrelációs mátrix a későbbi elemzési módszer (dimenziócsökkentés) alapja, ezért mindenképp bővebben foglalkoznunk kell a problémával. A következőkben a beleegyező válaszadást járjuk körül részletesebben, a saját adataink tükrében, illetve megkíséreljük a statisztikai korrekciót, bemutatjuk annak tulajdonságait.

Egyes vélemények szerint a beleegyező válaszadás személyes jellemző, amely egy régebbi tanulmány (Couch, Keniston 1960) szerint az emocionális, impulzív személyekre jellemző. A jelenséget más szempontok szerint is vizsgálták, például a különböző kultúrák közötti eltéréseket (Bachman, O'Malley 1984), esetünkben azonban erről nincs szó. Összefüggés található ezen kívül a képzettséggel, a jövedelemmel, illetve a korrall és a nemmel is.

A következőkben tekintünk át a beleegyező válaszadás mérési lehetőségeit. Steenkamp és Baumgartner (1995) konkrét példát is hoz a mérésre. Egy kérdőívben belül olyan kérdéspárt vizsgáltak, ahol az egyik

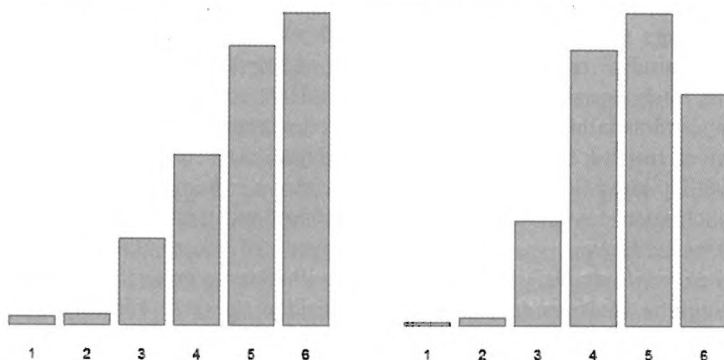
negatív, a másik pedig pozitív módon került megfogalmazásra. Az állítások a következőképp szóltak:

- Szeretek új dolgokat és változást meg tapasztalni a napi rutinban.
- A rutinszerű életet preferálom a változásokkal telivel szemben.

A fenti kérdésekre adott válasz alapján következtethetünk egy adott egyén, illetve a mintában szereplők beleegyező válaszadásának mértékére. Amennyiben a kérdésekről tudjuk, hogy egydimenziós módon egy adott jellemzőt mérnek, és azonos mennyiségű negatív és pozitív módon megfogalmazott kérdést teszünk fel, az segíthet a torzítás kiküszöbölésében.

Egy másik lehetséges megoldás, ami nem feltétlenül igényli negatív módon megfogalmazott kérdések feltevését, arra támaszkodik, hogy sok, minél heterogénebb kérdés esetén vizsgálja a válaszadók értékeit, pontosabban az egyetértésük mértékét. A kérdések heterogenitása azért fontos, hogy elkülöníthető legyen a beleegyező válaszadás és az adott téma iránti különös érzékenység. Azzal a feltételezéssel élhetünk, hogy ha a nagyon különböző kérdésekre egyaránt erős beleegyezést mutat egy adott válaszadó, akkor az nem csak a valós véleményét, hanem egyben a beleegyező magatartását is méri.

I. ábra: Tipikus válaszadási gyakoriságok két kiemelt változóra



Forrás: Saját szerkesztés

A BELEEGYZŐ VÁLASZADÁS HATÁSA A KORRELÁCIÓRA ÉS A JAVASOLT STATISZTIKAI KORREKCIÓ

A páronkénti korrelációs együtthatókat ugyancsak érinti a beleegyző válaszadás, méghozzá általában növeli a korrelációt, hisz a tényleges együttmozgáson kívül a beleegyző válaszadás hatása is belekerül a kapcsolat szorossági mérőszámába. Ennek belátásához tegyük fel, hogy a két vizsgálni kívánt változó X és Y , de a beleegyző válaszadás következtében egy additív torzító hatás figyelhető meg, ami a kérdésektől független, azaz kizárólag a válaszadó személyéhez köthető. Jelöljük ezt Z -vel, a ténylegesen megfigyelt, torzítással terhelt változókat pedig $\tilde{X} = X + Z$ és $\tilde{Y} = Y + Z$ módon. Meg kell vizsgálnunk a megfigyelt $\rho(\tilde{X}, \tilde{Y})$ és a valójában vizsgálni kívánt $\rho(X, Y)$ kapcsolatát.

Ekkor

$$\rho(\tilde{X}, \tilde{Y}) = \rho(X + Z, Y + Z) = \frac{\text{Cov}(X + Z, Y + Z)}{\sqrt{\text{Var}(X + Z)\text{Var}(Y + Z)}} \quad (1)$$

Feltételezve, hogy a beleegyző válaszadásra való fogékonyság mértéke és a tényleges vélemény korrelálatlanok, a két összehasonlítandó korreláció az alábbi módon írható:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \quad \rho(\tilde{X}, \tilde{Y}) = \frac{\text{Cov}(X, Y) + \text{Var}(Z)}{\sqrt{(\text{Var}(X) + \text{Var}(Z))(\text{Var}(Y) + \text{Var}(Z))}} \quad (2)$$

A két formula alapján néhány megállapítást azonnal tehetünk:

- amennyiben X és Y a valóságban korrelálatlan, úgy a beleegyző magatartás bármilyen kis jelenléte esetén is pozitív (megfigyelt) korrelációs együtthatót kapunk;
- amennyiben a megfigyeléseink beleegyző válaszadás szempontjából homogének, azaz $\text{Var}(Z) = 0$, úgy ez semmiféle torzulást nem okoz a korrelációban (az átlagokban természetesen igen);
- amennyiben $\text{Var}(Z)$ nagyságrendekkel meghaladja a többi paraméter értékét, akkor a korreláció minden határon túl

megközelíti a +1-es értéket, ez azonban inkább matematikai tulajdonság, a valóságban nem jellemző, hogy a beleegyző válaszadás szempontjából jóval heterogénebbek a minta résztvevői, mint a vizsgált kérdés szempontjából;

- a két korrelációs együttható közötti nagyságrendi reláció a paraméterek nagy száma miatt nem állapítható meg egyértelműen.

Mivel egyértelműen nem mondhatjuk ki, hogy a korreláció növekszik a beleegyző válaszolás heterogén jelenlétével, megvizsgáltuk, hogy mik ennek a feltételei. Ennek érdekében a megfigyelhető korreláció parciális deriváltjai közül a $\text{Var}(Z)$ szerinti tekintettük, hiszen ez mutatja meg, hogy a beleegyző válaszadás növekedése esetén mikor növekszik, illetve mikor csökken a mért korreláció a ténylegesen vizsgálni kívánt értékhez képest. A parciális derivált-ról belátható, hogy:

$$\frac{\partial \rho(\tilde{X}, \tilde{Y})}{\partial \text{Var}(Z)} = \frac{-\text{Cov}(X, Y)(\text{Var}(X) - \text{Var}(Y) + 2\text{Var}(Z)) - \text{Var}(Z)(\text{Var}(X) - \text{Var}(Y)) - 2\text{Var}(Z)\text{Var}(Y)}{2[(\text{Var}(X) + \text{Var}(Z))(\text{Var}(Y) + \text{Var}(Z))]^2} \quad (3)$$

alakban írható fel. Mivel tudjuk, hogy a $\text{Var}(Z) = 0$ esetben a két korreláció megegyezik, valamint (3) alapján az alábbiakat mondhatjuk el:

- Amennyiben az elméleti kovariancia/korreláció negatív, a megfigyelt korreláció minden esetben magasabb értéket vesz fel, mint az elméleti, hiszen (3)-ban minden tag pozitív, azaz az eredeti függvény növekvő.
- Amennyiben az elméleti korreláció pozitív, a csökkenés/növekedés (3) számlálójának előjelétől függ, hiszen a nevező nemnegatív. A számlálóból felírhatjuk a különböző paraméterek melletti feltételeket, amire az adódik, hogy a

$$\text{Var}(Z) > \frac{\text{Cov}(X, Y)(\text{Var}(X) + \text{Var}(Y)) - 2\text{Var}(X)\text{Var}(Y)}{\text{Var}(X) - \text{Var}(Y)} \quad (4)$$

reláció teljesülése esetén a torzítás növeli az elméleti korrelációt, ellenkező esetben csökkenti.

- A fentiek azt jelentik, hogy amennyiben a valós, elméleti korreláció nagyon magas pozitív értéket vesz fel, úgy alacsony varianciájú Z változó (azaz minden válaszadó hasonlóan érintett a beleegyező válaszadásban) csökkentheti a megfigyelt korrelációt, méghozzá viszonylag kis mértékben, egyre nagyobb variancia esetén ez a hatás eltűnik. Ennek magyarázata az, hogy ha a két nem megmérhető változó között erős a korreláció, a hozzájuk adott alacsony „zaj” csökkenti a korrelációt, míg ha ez a „zaj” egyre erősebb, akkor a korreláció újra erősödni fog (immár megközelítve az egyet). Mindenesetre fontosnak tartjuk megemlíteni azt, hogy a fenti modellben a valós korrelációt csökkentheti is egy közös tényező bevezetése, amit nem minden korrekciós módszer vesz figyelembe (Lindell, Whitney 2001).

A fentiekben azt láttuk tehát, hogy a gyakorlatban releváns szituációk esetén a beleegyező válaszadás a megfigyelt korrelációs együtthatókat növeli, különösen igaz ez a negatív korrelációk esetére. A továbbiakban arra teszünk kísérletet, hogy a megfigyelhető, már torzítással terhelt adatokat statisztikai korrekciónak vessük alá.

A későbbiekben bemutatandó eredeti (megfigyelt, vélhetően torzított) korrelációs együtthatók jól mutatják, hogy az egymással ellentétesnek tűnő kérdések esetén sem igen látunk gyengénél erősebb negatív irányú kapcsolatot. A pozitív korrelációk minden esetben sokkal nagyobb abszolút értéket vesznek fel, mint a negatívok. A már bemutatott hisztogramokon kívül ez a jelenség is a beleegyező válaszadás jelenlétére utal.

A következőkben a detektált beleegyező válaszadás általunk javasolt statisztikai módszerekkel történő korrekcióját tekintjük át. A fentiek, valamint az adatok vizsgálata alapján azt a következtetést vontuk le, hogy az egyes válaszadások értékében nem csak

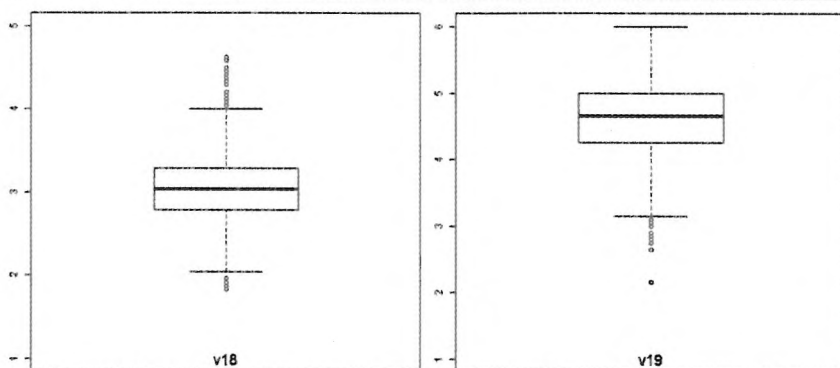
a tényleges egyetértés, hanem a pozitív válaszadási hajlandóság is szerepet játszik. Mivel a beleegyező magatartást sok tényező befolyásolhatja, úgy döntöttünk, hogy az egyének szintjén korrigálunk, illetve a korrekciót a két kérdéscsoport (v18: Mennyire értesz egyet a következő állításokkal?, v19: Mennyire tartod fontosnak az alábbi tulajdonságokat?) esetén külön végezzük el. Ehhez vezessük be a következő jelöléseket:

- \bar{x}_{ij} az i . válaszadó j . kérdésre adott tényleges (megfigyelt) válasza a kérdéscsoporton belül ($i=1,2,\dots,n$ $j=1,2,\dots,J$)
- $\bar{x}_i = \frac{\sum_j x_{ij}}{J}$ az i . válaszadó heterogén kérdésekre adott válaszainak átlaga

Esetünkben z_i gyakorlatilag a beleegyező válaszadás (becsült) mérőszámának tekinthető az adott kérdéscsoportra. Abban az esetben, ha ez az érték az i . válaszadó esetén magas, az azt jelenti, hogy a válaszadóra erősen jellemző a beleegyező válaszadás jelensége. A két kérdéscsoport közül a második esetben találtunk különösen magas mértékét a beleegyező válaszadásnak. A valamennyi válaszadóra és a két kérdéscsoportra meghatározott z_i mérőszámok eloszlását a második ábra segítségével mutatjuk be.

A boxplotok alapján a válaszadók igen heterogének a kérdésekre adott válaszaik átlaga alapján, ami nem lenne meglepő, ha a kérdések azonos jelentéssel bírnának. Mivel azonban heterogén, egymással akár ellentétes kérdésekre válaszoltak, az átlagok ilyen mértékű különbsége meglepő, azt mutatja, hogy egyes válaszadók a skálák különböző értékeit „szokták” használni. A v19-es kérdéscsoport esetén látható, hogy a válaszadók mintegy negyede 5 feletti átlagot „produkált” az 1-6 skálán, illetve a válaszadók fele 4,65 feletti mutatóval rendelkezik. A 3 alatti átlagúak pedig már kiugró értékűnek számítanak, ami újabb bizonyítékát adja annak, hogy a beleegyező válaszadás torzításával terhelték az adatok. Ezen a kérdéscsoporton belül egyébként az átlagok varianciája, $\text{Var}(Z)$, azaz a korrelációk torzítása is jóval nagyobb.

2. ábra: Boxplot a válaszadások átlagai alapján



Forrás: Saját szerkesztés

1. táblázat: Korrelációs együtthatók a v19 kérdéscsoport alapján az eredeti (főátló felett) és a korrigált (főátló alatt) értékekre

	v19_2	v19_8	v19_9	v19_13
Gazdagság, pénz	1	-0,098	-0,037	0,350
Más emberek meghallgatása	-0,283	1	0,398	0,078
Alázatos, szerény, visszafogott	-0,182	0,274	1	0,084
Sikeresség, elismertség	0,170	-0,278	-0,222	1

Forrás: Saját szerkesztés

A jelenség megállapítása után a legfontosabb feladat a hatás csökkentése. Feltevésünk szerint a beleegyező válaszadás additív módon kapcsolódik a kérdésekről alkotott valódi véleményhez és független attól, a korrelációk közötti viszonyt is ezek alapján a feltevések alapján vizsgáltuk. Ennek megfelelően a tisztítást az alábbi módon végeztük el:

$$x_{ij} = \tilde{x}_{ij} - z_i \quad (5)$$

ahol a már bevezetett jelölések mellett x_{ij} a beleegyező válaszadói torzítástól tisztított, korrigált érték az i . válaszadó és j . változó esetében. A fenti eljárást valamennyi, a későbbiekben alkalmazandó változó esetén elvégeztük.

Az eredeti, valamint (5) segítségével tisztított értékek közötti korrelációs együtthatók az alábbi táblázatban találhatóak meg néhány – a v19 csoportból származó, későbbiekben használandó – változó esetére. Jól láthatóan negatív korrelációk is megjelentek az ellen-

tés jellemzőket mérő változók esetén, ami alapján arra a következtetésre jutottunk, hogy az alkalmazott statisztikai korrekció legalább részben sikeresnek tekinthető.

A két korrelációs mátrix összehasonlításában elsődlegesen azt tartjuk fontosnak, hogy a következő lépésben, a dimenziócsökkentésben mennyire értelmezhető főkomponensek alakulnak ki a mátrixok alapján. Arra, hogy két sokaságból származó korrelációs/kovariancia mátrix azonosnak tekinthető-e, a Box-M statisztika (Box 1949) adhat választ. Az eljárás kifejezetten érzékeny a normalitás nem teljesülésére, tájékoztató jelleggel azonban elvégeztük a tesztet, annak ellenére, hogy ez a feltétel a vizsgált adatok esetén nyilvánvalóan nem teljesül. A bemutatott mátrixok esetén a próbafüggvény értéke 180 feletti, a p-érték gyakorlatilag 0, azaz a két mátrix minden szignifikancia szinten különbözik tekinthető. A tesztet elvégeztük a másik kérdéscsoport esetén is, és ott is szignifikáns különbségeket találtunk.

Azt gondoljuk, hogy a normalitás nem teljesülése ellenére is jelentősen különböző korrelációs mátrixot kapunk mindkét kérdéscsoport esetén, főként amennyiben figyelembe vesszük a nagy mintaelemszámot is. A korrekció jelentősége azonban elsősorban nem ebben van, hanem a dimenziócsökkentés utáni mesterséges változók értelmezhetőségének jelentős javulásában.

A következő fejezet célja tehát kettős:

- elsőként a klaszterezhetőség egy mérőszámát mutatjuk be, majd a mérőszám segítségével vizsgáljuk a fentiekben bemutatott korrekció és a dimenziószám csökkentés hatását a klaszterezhetőségre,
- a mérőszám alakulásán túl bemutatjuk a dimenziócsökkentés eredményének értelmezhetőségét a korrigált és az eredeti adatok tekintetében.

DIMENZIÓCSÖKKENTÉS ÉS KLASZTEREZHETŐSÉG

A klaszterezhetőség mérése az adatok véletlen elrendezésének vizsgálatát jelenti, melynek mérőszáma a Hopkins-index (Hopkins 1954). Számításához normál eloszlású, véletlenszerűen generált adathal-

mazon végzett legközelebbi szomszédító való távolság számítási értékének össze hasonlítás szükséges a klaszterezhetősége vizsgálandó adathalmaz adataira számított hasonló mértékkel. Számítása az alább képlet alapján történik:

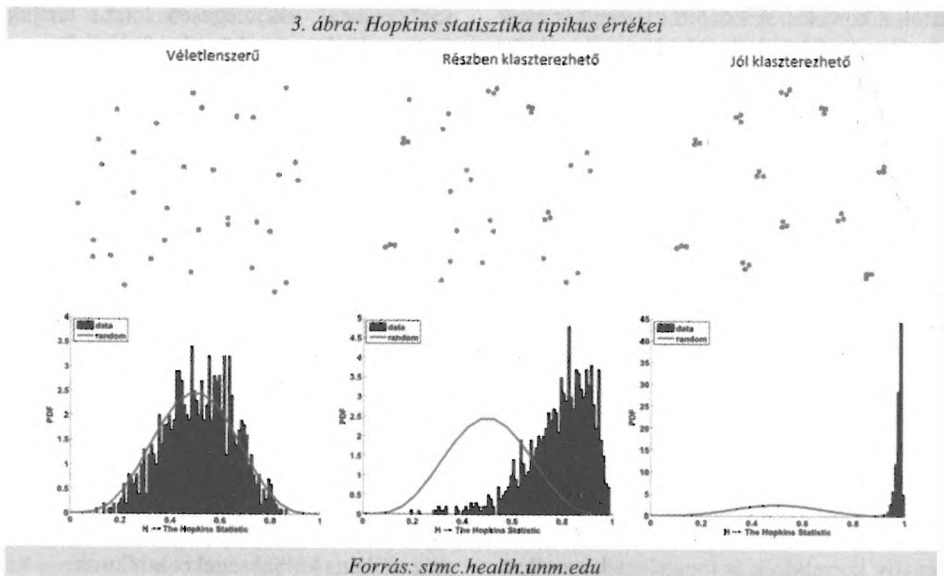
$$H_{ind} = \frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n q_i + \sum_{i=1}^n w_i} \quad (6)$$

ahol q_i a véletlenszerűen generált pontok közti távolságmérték, w_i pedig a valós adathalmaz pontjai közti távolságmérték.

A 3. ábrán a Hopkins statisztika tipikus értékeit láthatjuk különböző adathalmazokra. Véletlenszerűen elhelyezkedő pontra 0,5 alatti, míg klaszterezhetőség szempontjából a részben vagy a jól klaszterezhető esetekben az index értéke közelít az 1-hez.

Amennyiben az index értéke nagyobb, mint 0,5 az adathalmazban klaszterek azonosíthatók, de a mérőszám a klaszterek számáról nem ad információt.

A klaszterezhetőség szempontjából fontos, hogy az eredeti adatok korrekciója nem változtatott az attribútumok számán. Mivel az attitűdállítások által alkotott 44 dimenziós térben a válaszok közötti távolságok nagyjából egyformák lesznek (különösen 1500 feletti adatpont esetén),



így a „dimenzionalitás átkaként” ismert jelenség miatt távolság alapú klaszterezési eljárások az eredeti adathalmazra nem alkalmazhatóak. Ezt igazolja az a tény is, hogy az eredeti adathalmazra számított Hopkins mutató értéke 0,590 (lásd az összefoglaló jellegű 3. táblázatban), ami tulajdonképpen a sokdimenziós térben a pontok véletlenszerű elhelyezkedésére utal.

Ahhoz, hogy a válaszok közötti távolság értelmet kapjon, csökkenteni kell az attribútumok számát. Ezt két módon tettük meg. Először is olyan dimenziócsökkentési módszereket alkalmaztunk az eredeti, illetve a korrigált adatokra, amelyek az adathalmazokban található legtöbb információt megőrzik a két dimenziós térbe való leképezés során. Második lépésben pedig ugyanezen módszereket alkalmaztuk, de már olyan, kevesebb attribútumot tartalmazó adathalmazokra, ahol a rendelkezésre álló változókat szakmai szempontból előzetesen szelektáltuk (a szelekció szempontjait a későbbiekben bemutatjuk).

A dimenziócsökkentő módszerek közül a főkomponens analízis (PCA), az ISOMAP algoritmus és a Sammon-leképezés hatékonyságát hasonlítottuk össze aszerint, hogy ezek alkalmazása mennyire javítja a klaszterezhetőséget. A PCA módszer a látens változókat az eredeti attribútumok olyan lineáris kombinációjaként állítja elő, melyek csökkenő mértékben tartalmazzák az eredeti adatok változékonyságát. A Sammon-leképezés a többdimenziós skálázás egy speciális esete, ahol a pontok közti távolságmegőrzés hibáját az eredeti távolságokkal normalizáljuk, ezért ez nem lesz lineáris leképezés. A Sammon algoritmus kisebb távolság esetén jobban megőrzi az eredeti távolság viszonyokat, mellyel az adathalmaz belső struktúrája jobban megőrződik (Sammon 1969, Tan et al. 2011). Ez éppen megfelelt a vizsgálati feltételeknek is, hiszen az egymáshoz jobban hasonlító válaszadókat a transzformált térben is közelebb helyezte el egymáshoz, amit még a Manhattan távolság megválasztásával is elő lehetett segíteni. Az ISOMAP algoritmus

vetítéskor már nemcsak a pontok közötti szomszédságot próbálja megőrizni, hanem a szomszédsági gráfban található legrövidebb útvonalakat is, így értelemszerűen ez sem lesz lineáris leképezés.

Bár a korrigált adatokra alkalmazott klaszterezhetőségi index értéke lett a legjobb (0,727), az attribútum kiválasztás után mégis a PCA módszer használata mellett döntöttünk, mert annak linearitása miatt egyedül ott lehetett a látens változókhöz a gyors/lassú, és modern/tradicionális fogalmakat igazolhatóan hozzárendelni.

Miután az összes attitűdállításon végzett vizsgálataink azt mutatták, hogy a korrigált adatok mindenféle módszer esetén klaszterezhetőségi szempontból jobban viselkednek, mint az eredetiek, az attribútum kiválasztást ezen a halmazon végeztük el. A kiválasztás során megfelelő attribútumok megkeresésére az alábbi ket-tős kritériumrendszerrel dolgoztuk ki:

1. a statisztikai elvárásoknak megfelelőeknek, így például elfogadható Cronbach- α értéket (Cronbach 1951) mutassanak, vagy megfelelő szinten járuljanak hozzá a főkomponensek kialakításához (a kommunalitás-értékek vizsgálatakor), mindemellett pedig
2. mindkét dimenzióhoz (élettempó, értékorientáció) kapcsolódva, megfelelő számú változót kapjunk, amelyek mind a pozitív, mind a negatív értelmezési tartományt képviselni tudják, amelyek az élettempó esetében a „gyors” (+) és lassú (-), az értékorientáció esetében pedig a modern (+) és a tradicionális (-) végpontokat jelentik.

Ezt követően, az átlagos válaszadói jellemzőtől tisztított adatok alapján az élettempó és értékorientáció látens változók meghatározása volt a következő cél. A fentiekben bemutatott modellnek megfelelően a transzformált értékekre vonatkozóan főkomponens analízist végeztünk az élettempót és értékorientációt mérő változókra egyszerre.

A megmaradt változók két főkomponenst alkottak, melyeket élettempónak és

értékorientációnak neveztünk el az elméleti konstrukciónak megfelelően. Az egyes változók kódjait is megadjuk, amire a későbbiekben még hivatkozunk.

- Élettempó (Mennyire értesz egyet a következő állításokkal?, 1-5):
 - v18_2: Szeretek mindenhol ott lenni, pörögni (-)
 - v18_3: Nagy gondot fordítok arra, hogy nézek ki (-)
 - v18_10: Kimaradok a dolgokból (+)
 - v18_15: Inkább csak szemlélem a világot (+)
 - v18_18: Másokhoz képest lassan élek (+)
- Értékorientáció (Mennyire tartod fontosnak az alábbi tulajdonságokat?, 1-6):
 - v19_2: Gazdagság, pénz (+)
 - v19_8: Más emberek meghallgatása, akikkel nem értek egyet (-)
 - v19_9: Alázatos, szerény, visszafogott (-)
 - v19_13: Sikeresség, elismertség (+)

Az egy feletti sajátértékkel rendelkező főkomponensek meghagyásával a magyarázó erő mintegy 50%-ot ért el, illetve a kommunalitások 0,4-0,5 közötti értékeket vettek fel. A szokásosan használt KMO teszt 0,764-es értéke elfogadhatónak mondható (Hajdu, 2003).

Annak érdekében, hogy a korrekció értelmezhetőségre gyakorolt hatását is vizsgálni tudjuk, a rotált komponens mátrixot meghatároztuk az eredeti és a korrigált

adatok alapján is, a rotált komponens mátrixokból elhagyva az alacsony abszolút értékű értékeket a 2. táblázatot kapjuk.

A rotált komponens mátrix alapján megállapítható, hogy a módszer jól szétválasztotta az értékorientációt és az élettempót mérő változókat a korrigált adatok alapján. A mátrixban található együtthatók előjelei megfelelnek előzetes várakozásainknak. Az eredeti adatok esetén a főkomponens analízis paraméterei is rosszabbak, illetve az eredmény marketing szakmai szemmel sem értelmezhető.

A főkomponensek segítségével kapott score-okat elmentettük, amit a továbbiakban a csoportok kialakításánál fogunk felhasználni. A két kapott változó korrelálatlan, a tartalmukat megállapítottuk, ezért azokat könnyedén fel tudjuk használni a csoportok jellemzésére. A kapott score-ok -1 -gyel való szorzására volt szükség a későbbi jobb értelmezhetőség miatt, ami csak az irányt változtatja meg (épp ellentétesre).

A 3. táblázatban gyűjtöttük össze a fejezetben bemutatott módszerek eredményeinek klaszterezési tendencia értékeit. Megfigyelhető, hogy a két főkomponens kiválasztása után a klaszterezhetőségi indexek visszaesnek az attribútum kiválasztási állapothoz képest, de ez a korrigált adatok esetében sem esik vissza az eredeti, sokdimenziós érték alá. Vagyis az elvégzett transzformációk során a dimenziót úgy sikerült minimálisra lecsökkenteni, hogy ezzel az adatok klaszterezhetősége nem romlott.

2. táblázat: Rotált komponens mátrixok az eredeti és korrigált adatokra

Adatok	Eredeti		Korrigált	
	1	2	1	2
v18_2		0,627	-0,712	
v18_3		0,714	-0,625	
v18_10	0,667		0,646	
v18_15	0,707		0,610	
v18_18	0,702		0,718	
v19_2		0,538		0,676
v19_8	0,399			-0,722
v19_9	0,540			-0,559
v19_13		0,690		0,602

Forrás: Saját számítás

3. táblázat: Hopkins index értékei az egyes adatcsoportokra

Adatok	ISOMAP ($k=7$, $ndim=2$) - Manhattan	Sammon (del, dup , $ndim=2$) - Manhattan	PCA ($ndim=2$) - Euklideszi	Eredeti adatok	Főkompo- nensek
v1819	0,559	0,645	0,640	0,590	0,563
v1819_korr	0,587	0,727	0,657	0,676	0,586
v1819_subset	0,598	-	-	0,559	0,815
v1819_korr_subset	0,597	-	-	0,651	0,588

Forrás: Saját számítás

Összességében tehát megállapíthatjuk, hogy a korrekció egyrészt javította az eredmények értelmezhetőségét, másrészt a módosított adatok esetén a klaszterezhetőségi mérőszámok is jobbak.

A következő fejezetben a klaszterezés néhány gyakran használt módszerét, valamint a klaszterezés minőségének mérőszámait mutatjuk be röviden, illetve alkalmazzuk ezeket az adatainkra.

KLASZTEREZÉS ÉS ANNAK MINŐSÉGE

Klaszterezésnek hívjuk az adatelemek halmazán végzett elemzési műveletet, melynél nem ismertek előre az egyes elemekhez tartozó osztályok tulajdonságai. Az osztályba sorolás során az egy klaszterbe tartozó elemek hasonlósága nagy, míg a különböző klaszterekbe tartozó elemek között kicsi a hasonlóság. A hasonlóságot/különbséget az elemek előre kiválasztott tulajdonságai alapján határozhatjuk meg. A klaszterezés minőségét az elemek különbségének mértéke alapján lehet megállapítani. (Tan et al. 2011)

Particionáló klaszterezés – k-közép (k-means)

Az alább ismertetendő algoritmus esetében bemeneti paraméterként a létrehozandó partíciók számát várja el, ezt k -val jelöli. Előfeltételnek tekintjük, hogy minden csoport elemszáma legalább 1 legyen és minden elem csak egy csoportba tartozzon.

Távolság alapú klaszterezési algoritmusok (Distance-based Clustering Algorithm) esetében a hasonlóság mértékének meghatározása alapján történő klaszterezés, azaz

az elemek közelségének meghatározása történik, valamely hasonlósági függvény segítségével. Az algoritmus elve szerint indulásnál n elemszámon létrehozza a $k < n$ partícióból álló halmazokat (klasztereket), majd ezen kiinduló klaszterek tagjait áthelyezi egyik partícióból a másikba, ezzel csökkentve a klaszterelemek közti távolságot. Az áthelyezést iteratív áthelyezési technika segítségével oldja meg, melynél objektív particionáló függvény (hasonlósági függvény, esetünkben k -átlag) optimumának keresésével történik a klaszterek alakítása. Az adatok így egymást nem átfedő részhalmazokban jelennek meg, vagyis elem csak egy részhalmaz tagja lehet.

A k -means (k -átlag) függvény esetében az egy csoportba tartozó elemek átlaga reprezentálja a klasztert, elem áthelyezésével a súlypontot újraszámolja; az elemek és a súlypont(ok) közti különbségek összegét minimalizálja. Előnye, hogy jó skálázhatóságának következtében nagy adatbázisokon is használható. Hátrányai közé tartozik, hogy csak abban az esetben használható, ha meghatározható a klaszter átlaga, továbbá nem konvex alakú klaszter felismerésre nem alkalmas és érzékeny a szélsőséges értékekre is.

Esetünkben a k -átlag paramétereit az alábbiakban állítottuk be:

- kialakítandó klaszterek száma=8
- futás száma = 100 – véletlen kiindulás pontokból történő klaszterkialakítás
- mértéktípus: Numericalmeasure: ManhattanDistance – távolságmérték típusa
- iterációk száma = 100 – klaszterek kialakításához használt iterációs szám

4. táblázat: Klaszterezés performancia értékei az egyéni átlagokkal korrigált (subset) adatokra és a főkomponensekre

Mérték	K-means	
	Egyéni átlaggal korrigált subset	Főkomponensek
SSE	0,173	0,148
Average Silhouette Coefficient	0,309	0,326

Forrás: Saját számítás

A főkomponensek használatával és az előző fejezetben bemutatott modell adatain (subset) létrehozott klasztereket a 4. táblázatban összefoglalt mértékek mentén hasonlítottuk össze. Felügyelet nélküli belső mértékeket használtunk, melyekkel a klaszterezés minőségét külső információk felhasználása nélkül határozhatjuk meg.

Az egyes mértékek jelentése:

- *Négyzetes távolság hibája (Sum of Squared Error – SSE):*

Négyzetes távolság hibája esetében az alacsonyabb érték jelzi a klaszterezés minőségének magasabb fokát. Értéke a klasztert alkotó pontoknak a klaszter középpontjától való távolságuk négyzetösszegének a számítása alapján képződik:

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist^2(m_i, x) \quad (7)$$

ahol K a klaszterek száma, x a C_i klaszter egy pontja és m_i a C_i klaszter középpontja.

- *Sziluett együttható átlaga (Average Silhouette Coefficient):*

Az együttható számítható mind a k-means, mind a sűrűség alapú, mind a hierarchikus klaszterezések eredményeire, így jó lehetőséget nyújt az különböző klaszterezési eljárások által alkotott klaszterek összehasonlítására.

Képlete:

$$s_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (8)$$

ahol a_i az i -edik elem átlagos távolsága a klasztere többi elemétől, b_i az i -edik elemre és minden azt nem tartalmazó klaszterre számított elem átlagos távol-

ságának legkisebb értéke az aktuális klaszter elemeitől.

Az együttható értéke -1 és $+1$ között helyezkedhet el. A negatív érték jelentése, hogy a klasztert alkotó pontoktól vett átlagos távolság nagyobb, mint másik klasztert alkotó pontoktól vett legkisebb átlagos távolság. Ez a klaszterek kohéziójának alacsony voltára utal, így értéke a pozitív tartományban kívánatos. (Rouseeuw 1987)

A klaszterezések összehasonlításához átlagos sziluett együtthatót számítottunk az összes pont sziluett együttható értékének átlagával.

Az SSE érték esetében a főkomponens adatok klaszterezése esetében mutat jobb eredményt (0,148), a sziluett együttható szintén a főkomponensekből képzett klasztereink magasabb kohézióját jelzi (0,326). Ezek alapján kijelenthető, hogy a főkomponensekkel történő klaszterezés eredményeképpen kialakított csoportok minősége magasabb, mint az egyéni átlagokkal korrigált adatokon kiválasztott kérdéseken (subset) végzettéke.

A fentiekben bemutatuk az elemzés során felmerülő módszertani problémákat, illetve az általunk ezekre adott válaszokat. A végső klasztermegoldás alapján megvizsgáltuk a csoportok demográfiai hátterét, és ezek is megerősítették a csoportok tulajdonságainak értelmezhetőségét. A Z-generáció életstílus alapú szegmenseinek részletesebb bemutatása a <http://www.zgeneracio.hu/> tanulmányok oldalon található meg.

ÖSSZEFOGLALÁS

Tanulmányunkban az életstílus-alapú szegmentáció néhány módszertani problémáját jártuk körül. Azt tapasztaltuk, hogy – ahogyan ez a kérdőíves felmérések esetén gyakori – az adataink mérési hibákkal terheltek. A szegmentáció során felmerült problémák közül a beleegyező válaszadás okozta a legnagyobb fejtörést, ami leginkább az értékorientációt mérő kérdések esetén volt tetten érhető. Ezért egy egyszerű modell alapján megvizsgáltuk a torzítás elméleti hatásait, majd az ennek tükrében kidolgozott korrekciót alkalmaztuk sikerrel az adatokon. A tanulmány második felében egyrészt bemutattuk azokat a dimenzió-csökkentő eljárásokat, amelyeknek alá vetettük az adatokat annak érdekében, hogy a kiinduló modell dimenziót rekonstruálni tudjuk, másrészt foglalkoztunk a csoportképzés menetével, döntési pontjaival. Mindkét fázisban az eredeti, illetve a korrigált adatokat is vizsgáltuk, arra voltunk kíváncsiak, hogy az objektív mérőszámok szerint melyik eljárás tűnik jobbnak, illetve a korrekción átesett adatok minősége mennyiben segíti a célok teljesülését. Az objektív mutatók mellett az értelmezhetőséget is vizsgáltuk, a két elgondolást együttesen figyelembe véve hoztuk meg döntéseinket. Eredményül jól értelmezhető, a demográfiai változók alapján is hiteles szegmensek alakultak ki a Z-generáció életstílusát mérő kérdőív alapján.

JEGYZET

- 1 A Szerzők köszönetet mondanak a lektornak a tanulmány korábbi verziójához fűzött kritikái észrevételeiért és értékes megjegyzéséiért.

HIVATKOZÁSOK

Bachman, J. G., O'Malley, P. (1984), „Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles”, *Public Opinion Quarterly*, 48 2, 491-509

Box, G. E. P. (1949), „A general distribution theory for a class of likelihood criteria”. *Biometrika*, 36 3/4, 317-346

Couch, A., Keniston, K. (1960), „Yeasayers and naysayers: Agreeing response set as a personality

variable”, *Journal of Abnormal and Social Psychology*, 60 2, 151-172

Cronbach, L. J. (1951), „Coefficient alpha and the internal structure of tests”, *Psychometrika*, 16 3, 297-334

Galesic, M., Bosnjak, M. (2009), „Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey”, *The Public Opinion Quarterly*, 73 2, 349-360

Hajdu, O.: (2003), *Többváltozós statisztikai számítások*, Budapest, KSH

Grover, R., Vriens, M. (2006), *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*, Thousand Oaks, Sage Publications

Hopkins, B. (1954), „A new method for determining the type of distribution of plant individuals”, *Annals of Botany*, 18, 213-227

Hunyadi, L. (2001), *A mintavétel alapjai*, Budapest, Számalk Kiadó

Kaldenberg, D.O. – Koenig, H.F. – Becker, B.W. (1994): „Mail Survey Response Rate Patterns in a Population of the Elderly: Does Response Deteriorate With Age?” *The Public Opinion Quarterly*, 58 1, 68-76

Keller, T. (2008), *Értékrend és társadalmi pozíció*, Századvég, Új Folyam 47, 151-187

Kish, L. (1989), *Kutatások statisztikai tervezése*, Budapest, Statisztikai Kiadó Vállalat

Lindell, M.K, Whitney, D.J. (2001), „Accounting for common method variance in cross-sectional research designs”, *J. Appl. Psychol*, 86 1, 114-121

Lynn, P. (2014), „Longer Interviews May Not Affect Subsequent Survey Participation Propensity”. *The Public Opinion Quarterly*, published online July 22, 2014

Pintér, R., Kátay, B. (2010), „A hibrid adatfelvétel módszertani kihívásai”, *Statisztikai Szemle*, 88 7-8, 723-738

Podsakoff, P. M., MacKenzie, S. B., Lee, J-Y., Podsakoff, N. P. (2003), „Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies”, *Journal of Applied Psychology*, 88 5, 879-903

Porter, S.R. and Whitcomb, M.E. (2003), „The Impact of Contact Type on Web Survey Response Rates”, *The Public Opinion Quarterly*, 67 4, 579-588

Rousseeuw, P. (1987), „Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”, *Journal of Computational and Applied Mathematics*, 20 1, 53-65

Sammon, J.W. (1969), „A non-linear mapping for data structure analysis”, *IEEE Transactions on Computers*, 18 5, 401-409

Shih, T., Fan, X. (2008), „Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis”, *Field Methods*, August 2008 20, 249-271

Steenkamp, J.-B. E.M., Baumgartner, H. (1995), „Development and cross-cultural validation of

a short form of CSI as a measure of optimum stimulation level”, *International Journal of Research in Marketing*, 12 2, 97-104

Tan, P-N, Steinbach, M., Kumar V. (2011), *Bevezetés az adatbányászathoz*, Budapest, Panem

Töröcsik, M. (2011), *Fogyasztói magatartás – Insight, trendek, vásárlók*, Budapest, Akadémiai Kiadó

Vicente, P. and Reis, E. (2010), „Using Questionnaire Design to Fight Nonresponse Bias in Web Surveys”, *Social Science Computer Review*, May 2010 28, 251-267

Winkler, J. D., Kanouse, D. E., Ware, J.E. Jr. (1982), „Controlling for acquiescence response set in scale development”, *Journal of Applied Psychology*, 67 5, 555-561

Yammarino, F.J., Skinner, S.J. And Childers, T.C. (1991), „Understanding Mail Survey Response Behavior A Meta-Analysis”, *The Public Opinion Quarterly*, 55 4, 613-639

Hornyák Miklós, tudományos munkatárs

hornyakm@ktk.pte.hu
Pécsi Tudományegyetem
Közgazdaságtudományi Kar
Gazdaság-módszertani Intézet

Kehl Dániel, tanársegéd

kehd@ktk.pte.hu
Pécsi Tudományegyetem
Közgazdaságtudományi Kar
Gazdaság-módszertani Intézet

Kruszlicz Ferenc, egyetemi docens

kruszlic@ktk.pte.hu
Pécsi Tudományegyetem
Közgazdaságtudományi Kar
Gazdaság-módszertani Intézet

Szűcs Krisztián, egyetemi adjunktus

szucsk@ktk.pte.hu
Pécsi Tudományegyetem
Közgazdaságtudományi Kar
Gazdálkodástudományi Intézet

Methodological challenges in lifestyle-based segmentation

Research on market segmentation poses challenges in various aspects: in order to obtain realistic and consistent findings, it is important not to solely rely on methodological considerations but to also incorporate hands-on experience and practical insight. In this paper, we discuss the challenges related to this subject by evaluating the results and examine the relevance of our solutions.

Miklós Hornyák – Dániel Kehl – Ferenc Kruszlicz – Krisztián Szűcs