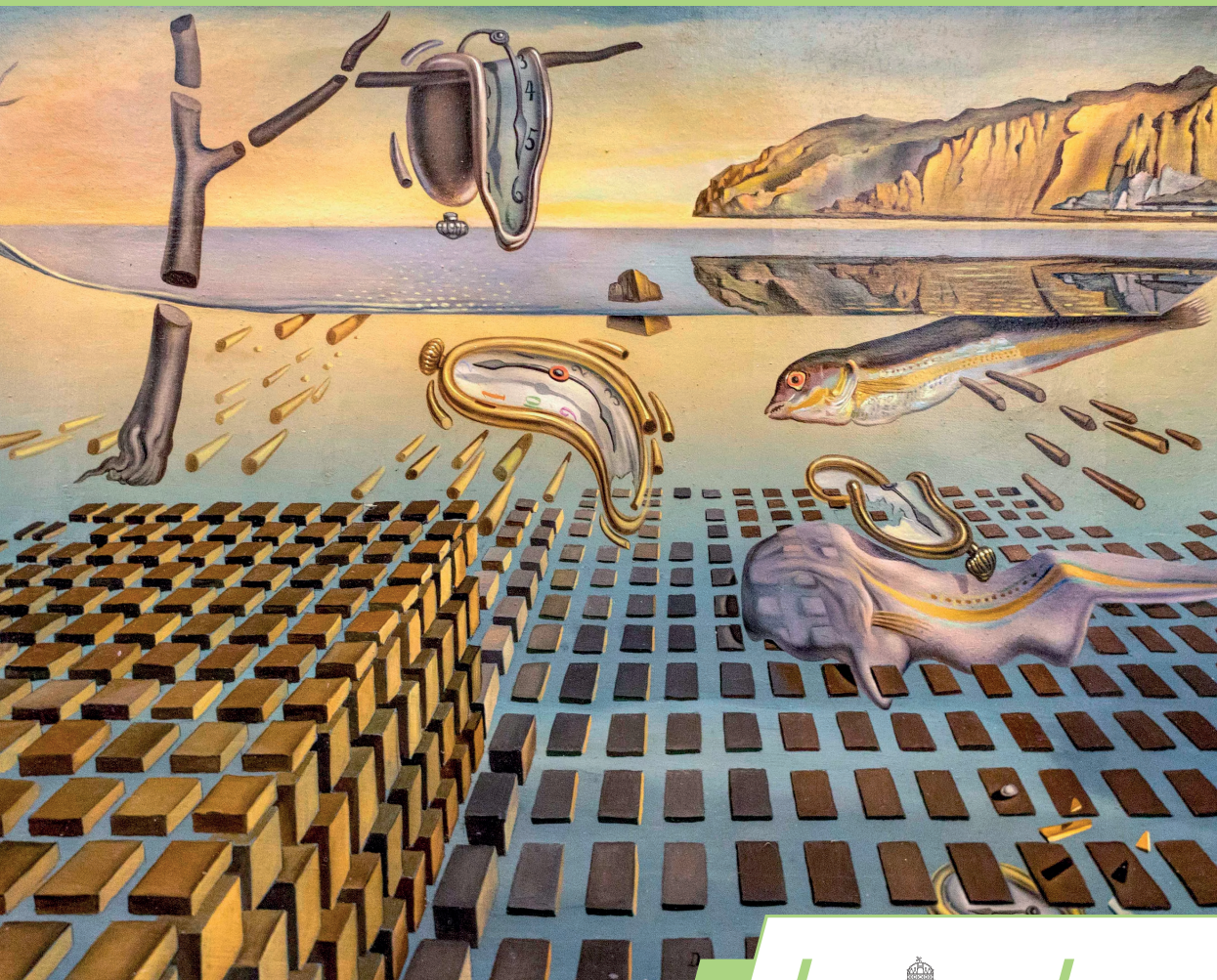


KVANTITATÍV ELEMZÉSI MÓDSZEREK

SPSS HASZNÁLATA A KUTATÁSI GYAKORLATBAN

Simon Judit, Berezvai Zombor, Kemény Ildikó,
Kun Zsuzsanna, Pusztai Tamás



KVANTITATÍV ELEMZÉSI MÓDSZEREK

SPSS használata a kutatási gyakorlatban



Budapest
2024

Kvantitatív elemzési módszerek

SPSS használata a kutatási gyakorlatban

Szerzők:

© SIMON JUDIT, BEREZVAI ZOMBOR, KEMÉNY ILDIKÓ,
KUN ZSUZSANNA, PUSZTAI TAMÁS

Olvasószerkesztő:

SZÉKELY KRISZTINA

ISBN (Print) 978-963-503-953-1

ISBN (e-Book) 978-963-503-954-8

DOI: 10.14267/978-963-503-954-8

A kiadvány megjelenését az Innovációs és Technológiai Minisztérium
Nemzeti Kutatási Fejlesztési és Innovációs Alapból nyújtott
támogatásával a Mecenatúra 2021 pályázati program finanszírozásában
megvalósuló MEC_K 141188 számú projekt tette lehetővé.



A borító *Salvador Dalí: Az emlékezet állandósága* című festményének
felhasználásával készült. (A WikiArt jóvoltából)

Kiadó: Budapesti Corvinus Egyetem
Nyomdai kivitelezés: CC Printing Kft.

Tartalomjegyzék

BEVEZETÉS	7
A könyv felépítése	9
BEMUTATKOZÁSOK	11
A szerzők	11
A kérdőív és az adatbázis bemutatása	13
A feldolgozáshoz használt statisztikai program: az SPSS	14
01 A MEGKÉRDEZÉSES VIZSGÁLAT ELŐKÉSZÍTÉSE, A KÉRDŐÍV ÖSSZEÁLLÍTÁSA	17
A marketingkutatás folyamatának áttekintése	18
A kérdőív felépítése	20
Szűrő kérdés	22
Mérhetőség, skálázás	26
A mérések skálásintje	27
02 AZ ADATBEVITEL	33
Nézetek és fájlok az SPSS-ben	33
A kérdőívből az SPSS fájlba	37
Adatbázis műveletek	39
03 LEÍRÓ STATISZTIKA	43
1. Gyakorlati példa	44
2. Gyakorlati példa	46
04 HIPOTÉZISVIZSGÁLAT	49
A vizsgálati módszer megválasztása	51
05 KERESZTTÁBLA ELEMZÉS	53
Elméleti bevezetés	53
1. Gyakorlati példa	55
2. Gyakorlati példa	60
Gyakorló feladat	63
06 VARIANCIAELEMZÉS	65
Elméleti bevezetés	65
Gyakorlati példa	68
Gyakorlati példa	76
Gyakorló feladat	82

07 KORRELÁCIÓ ELEMZÉS.....	83
Elméleti bevezetés.....	83
Gyakorlati példa.....	84
Gyakorló feladat	87
08 REGRESSZIÓELEMZÉS.....	89
Elméleti bevezetés.....	89
Gyakorlati példa.....	94
Szakirodalmi példa	106
Gyakorló feladat	108
09 LOGISZTIKUS REGRESSZIÓ.....	109
Lineáris valószínűségi modell.....	109
Logisztikus regresszió	111
Gyakorlati példa.....	115
Szakirodalmi példa	123
10 FAKTORELEMZÉS	127
Elméleti bevezetés.....	127
Gyakorlati példa.....	134
Gyakorló feladat	149
11 KLASZTERELEMZÉS.....	151
Elméleti bevezetés.....	151
Gyakorlati példa.....	159
Gyakorló feladat	171
12 A CONJOINT ALAPJAI	173
Elméleti bevezetés.....	173
A különböző conjoint módszerek	179
Conjoint value analysis (CVA).....	181
Choce-based conjoint (CBC)	182
Adaptív conjoint módszerek.....	183
További conjoint módszerek.....	184
A kísérleti design előállítása	185
A kísérleti design legfőbb jellemzői.....	185
A döntési feladatok számának meghatározása	191
A CBC kísérleti design sajátosságai.....	194
Hold-out döntési feladatok.....	195
Egy ajánlatos CVA kísérleti design létrehozása SPSS segítségével.....	196
Páros CVA kísérleti design létrehozása.....	200

A részhasznossági értékek előállítása CVA esetén	202
Az egy ajánlatos CVA módszer kísérleti designjának kódolása	202
A páros CVA módszer kísérleti designjának kódolása	204
Az elemzéshez szükséges adatbázis előállítása	205
Kísérleti design adatbázis és a válaszokat tartalmazó adatbázis összefűzése	210
A részhasznossági értékek becslése	214
Az átskálázott részhasznossági értékek előállítása	220
A dimenzió fontossági értékeinek előállítása	227
A piacmodellezés alapjai	228
A lineáris interpoláció használata	231
A részhasznossági értékek előállítása CBC esetén	232
Az adatok előkészítése	233
13 ESETTANULMÁNYOK	241
Minőségmérés a fesztiválpiacon – Az EFOTT fesztivál példája	241
Marketing mesterszakos hallgatók munkahelyi preferenciáinak elemzése adaptív conjoint módszerrel	248
A MOL Bubi használata a koronavírus-járvány idején.	255
FELHASZNÁLT IRODALOM.	268
MELLÉKLET – A KÉRDŐÍV	269

BEVEZETÉS

A társadalomtudományi, a közgazdaságtani és gazdálkodástudományi kutatásokban, ezen belül is hangsúlyosan a marketingkutatásban jelentős szerepe van az adatok kvantitatív jellegű elemzésének. A kvantitatív elemzést lehetővé tevő adatok előállításának az egyik leggyakrabban alkalmazott módja a primer megkérdezéses adatfelvétel.

A könyv célja a gyakran használt kvantitatív vizsgálati módszerek bemutatása a megkérdezéses kutatások adatainak elemzésére, a statisztikai módszerek lépéseinek gyakorlatorientált bemutatásával, valamint az egyik leggyakrabban használt statisztikai elemző szoftver, az SPSS használatának részletes, videoanyaggal segített leírásával (az IBM SPSS 29.0.1 verzió).

A könyv a megkérdezéses vizsgálatok előkészítésével, a kérdőívszerkesztés módszertani aspektusaival, az egy- és kétváltozós elemzési módszerekkel, az összefüggésvizsgálattal, a többváltozós strukturfeltáró módszerekkel, a preferenciavizsgálatok módszerével: a conjoint elemzéssel foglalkozik. A módszerek tárgyalását és a szoftver alkalmazásának lépéseit egy, didaktikai és elemzési célú adatbázis segítségével mutatjuk be. Emellett minden egyes módszernél gyakorlati példák és az olvasó számára megfogalmazott feladatok és megoldások segítik az elsajátítást és a gyakorlást.

A könyv szerzői a Budapesti Corvinus Egyetem oktatói (főállású vagy más típusú jogviszonyban), akik jelentős tapasztalattal rendelkeznek a marketingkutatás oktatásában a különböző szintű (alap- mester és doktori) programokban. Az oktatás mellett mindannyian foglalkozunk tudományos kutatással és publikációs tevékenységgel, valamint eltérő mértékű, de számottevő tapasztalatokkal rendelkezünk a gyakorlati piackutatásban is. Mindezen tevékenységünkben jelentős tapasztalatokra tettünk szert a kvantitatív elemzési módszerek használatában, amelyben túlnyomórészt az SPSS programcsomag különböző verzióit használtuk. A könyv megírásában az egyik legfontosabb motivációnk éppen annak a felhalmozódott oktatási és gyakorlati tapasztalatnak az összefoglalása volt, amivel mint Corvinusos szerzőcsapat rendelkezünk (röviden a könyv szerzői részben mutatkozunk be).

A könyvben az elmélet bemutatása tekintetében csak annyira szorítkozunk, amennyire oktatási tapasztalatunk szerint a módszerek használatához szükséges. Ugyanakkor nem csak a szoftver használatát kívánjuk bemutatni, fontos szándékunk az eredmények értelmezésének bemutatása.

A könyv fő fókusza a marketingkutatásban való felhasználhatóság, de mivel a megkérdezéses vizsgálatokat a gazdálkodástudományi, közgazdasági és társa-

dalomtudományi kutatásokban is gyakran alkalmazzák, ezért a felhasználók köre túlterjedhet a marketing- és piackutatással foglalkozók körén. Szándékunk szerint ezt a magyar nyelvű szakkönyvet mind a kutatásban, mind a felsőoktatási programokban haszonnal lehet kézikönyv jelleggel felhasználni. Könyvünknek a hozzáférhető szakkönyvek mellett az alábbi megkülönböztető jellemzői vannak:

- az elemzési módszerek gyakorlatközpontú bemutatásakor ugyanazon kérdőívet és a hozzá tartozó adatbázist elemezzük
- az egy-két és többváltozós adatelemzési módszerek közül a gyakorlatban leggyakrabban használtakat válogattuk ki
- a könyvben a mélységében és terjedelmében is legjelentősebb fejezet a conjoint módszertanát és gyakorlati alkalmazását mutatja be, ez egyedi és hiánypótló a magyar nyelven rendelkezésre álló szakkönyvekben található leírások között.

A könyvben felhasznált elméleti forrásokban olyan magyar és idegen nyelvű „alapkönyveket” jelöltünk meg elsősorban, amelyeknek a fogalmi rendszerét, megközelítési módját könyvünkben felhasználtuk. A könyv kézikönyv jellege miatt úgy döntöttünk, hogy a szakirodalmi forrásainkat a könyv végén közöljük, a szövegben nem alkalmazzuk a tudományos publikációkban szokásos hivatkozási módot, csak ott, ahol egy-egy ábra, táblázat vagy esetleg megfogalmazás közvetlen átvételéről van szó. A magyar nyelvű „alapkönyvek” közül négy könyv szemléletét tükrözi illetve ráépül, ezek a következők: Malhotra, Simon (k.m.), (2009): „Marketingkutatás”, Gyulavári, Mitev, Neulinger, Neumann-Bódi, Simon, Szűcs (2011): „A marketingkutatás alapjai”, valamint a Kemény, Simon, Berezvai, Kun (2021): „Marketingkutatás kvantitatív módszerei – segédanyag SPSS program használatához” (egyes módszereknél a könyvből átvett részeket is felhasználunk), Sajtos, Mitev (2007): „SPSS kutatási és adatelemzési kézikönyv”.

A könyv szerzőiként számos köszönettel tartozunk. Köszönjük kollegáinknak a Marketing- és Kommunikációtudományi Intézetben, hogy évtizedeken keresztül ebben a tudományos és oktatási műhelyben eszmecserével, vitákkal, fórumokkal segíthettük egymás fejlődését. Hasonlóan köszönjük a piackutatói gyakorlatban dolgozó kollegák segítségét, a Nielsen Piackutató Intézetben, a Hoffmann Research International-ban és a piackutatói konferenciákon is, ahol gyakorlati tapasztalataink nagy részét szereztük.

Köszönjük a Budapesti Corvinus Egyetemnek, hogy az SPSS programot biztosította a képzés és a kutatás számára, hasonlóan a Clementine-nak, személyesen Körmenyi Györgynek, aki segítette az SPSS különböző verzióihoz való folyamatos hozzáférésünket.

Kiemelt köszönettel tartozunk a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatalnak, a MEC_21 kódszámú „Tudományos Mecenatúra Pályázat” keretében elnyert MEC_K 141188 számú projektünk pénzügyi támogatásáért.

A legnagyobb köszönetet hallgatóinknak mondjuk, akik elsősorban a Marketing öt éves képzésben, majd a Marketing mester szakon több, mint 30 éven át segítettek oktatási módszereink fejlődését visszajelzéseikkel és saját tapasztalataik megosztásával. A könyv alapját képező első oktatási segédanyagokat hallgatóink kezdték el összeállítani, megérdemlik a név szerinti említést: Lajtós Gábor a kiinduló anyagot, majd a továbbfejlesztett verziót, Benke András, Csanda Gergely és Pusztai Tamás készítették.

Reméljük, hogy a könyv hasznos ismeretekkel tud szolgálni a gyakorlatorientált kutatási felhasználásokban, ehhez nagyon jó olvasást és gyakorlást kívánunk a szerzők:

Simon Judit, Berezvai Zombor, Kemény Ildikó, Kun Zsuzsanna, Pusztai Tamás

A nyílt hozzáférésű (Open Access) elektronikus kiadvány, a felhasznált adatbázis és a könyvben bemutatott elemzési módszerekről készült oktatóvideók elérhetőek:

<https://www.youtube.com/@SPSSagyakorlatban>

<https://unipub.lib.uni-corvinus.hu/10287/>

A könyv felépítése

A könyv elején a bemutatkoznak a szerzők, ahol megismerhetjük a szakmai hátterüket és tapasztalataikat. Ezután a könyvben ismertetett SPSS program, valamint a használt kérdőív és adatbázis bemutatása következik.

Az első nagyobb egység a kérdőíves vizsgálatokra összpontosít. Itt bemutatásra kerül a kérdőívek összeállításának folyamata, és részletesen tárgyaljuk, hogyan lesz a nyers kérdőívből elemezhető adatbázis.

A leíró statisztikákat bemutató fejezetet követően a hipotézisvizsgálatok menetét mutatjuk be, amelyet a különböző vizsgálati módszerek közötti választás logikájának ismertetésével zárunk. A klasszikus hipotézisvizsgálati módszerek közül a keresztábla-elemzés, a varianciaanalízis (egyszempontos és többszempontos ANOVA) és a regresszió (lineáris és logisztikus) módszertanát részletezzük. A struktúrafeltáró statisztikai módszerek közül a faktorelemzés és a klaszterelemzés kapott helyet a könyvben, melyek segítenek az adatok

mélyebb összefüggéseinek megértésében. A módszertani fejezetek sorát a conjoint elemzés zárja, amely hiánypótló helyet foglal el a magyar nyelvű kézikönyvek között, bemutatva a különböző conjoint módszereket, azok alkalmazását és a kísérleti design előállítását, az eredmények értelmezését. Az elemzési módszereket bemutató fejezetek végén legalább egy gyakorlati példa követhető nyomon képernyőképeken keresztül, ugyanazon adatbázisból merítve.

A könyv zárásaként három esettanulmány kerül bemutatásra, amelyek a könyvben ismertetett elemzési eszközök gyakorlati alkalmazását demonstrálják. Ezek a gyakorlati példák segítenek az olvasóknak megérteni, hogyan alkalmazhatók az elméleti ismeretek a valós világban.

A könyv végén a felhasznált irodalom jegyzéke található, amely összegyűjti a könyv készítése során használt forrásokat és további olvasmányokat ajánl az érdeklődőknek. Mellékletként a tárgyalt kérdőív szerepel.

BEMUTATKOZÁSOK

A következőkben bemutatjuk a könyv szerzőit és a könyvben szereplő fontos eszközöket, a kérdőívet és a hozzá tartozó adatbázist, amely az összes fejezetben a módszerek bemutatására szolgáló gyakorlati példa alapja, valamint a feldolgozáshoz alkalmazott statisztikai programot.

A szerzők



SIMON JUDIT a Budapesti Corvinus Egyetem egyetemi tanára, a Marketing és Média Intézet igazgatója 2008-2016 között, a Marketing PhD specializáció vezetője 2001 és 2021 között, jelenleg Professor Emerita. Kutatási, oktatási és publikációs területe az egészségügyi és gyógyszermarketing és -kutatás, a marketingkutatás, a fogyasztói magatartás és a B2B marketing. A Marketingkutatás tantárgyat több, mint 30 évig tárgyfelelősként fejlesztette és oktatta, magyar és német nyelven. Vállalati tapasztalatait a Nielsen Piackutató Kft. értékesítési

és ügyfélkapcsolati igazgatójaként, valamint az LHS Consulting egészségügyi tanácsadásra specializálódott cég alapítójaként és kutatási igazgatójaként szerezte. A Budapesti Corvinus Egyetemen a Passaui Egyetemmel közösen működtetett Német Nyelvű Gazdálkodástudományi Központ (DSG) alapítója és igazgatója volt 2016-ig. A Passaui Egyetem díszdoktora és díszpolgára. 2023-ban a Piackutatók Magyarországi Szövetsége életműdíjjal tisztelte meg.

BEREZVAI ZOMBOR a Gazdasági Versenyhivatal vezető közgazdásza és a Budapesti Corvinus Egyetem adjunktusa. Pályafutását pénzügyi menedzserként kezdte a Procter & Gamble-nél, majd 2021-ben PhD-fokozatot szerzett a Budapesti Corvinus Egyetemen. Több hazai felsőoktatási intézményben oktatott és oktat módszertani tárgyakat (kutatásmódszertan, ökonometria, piackutatás), amely módszereket munkája és kutatásai során is napi szinten használja. Kutatási érdeklődése a



mikromobilitáshoz, a kiskereskedelemhez, az egészséggazdaságtanhoz, az oktatói munka hallgatói értékeléséhez és egyéb empirikus piacelméleti kérdésekhez kötődik. 2023-ban Akadémiai Ifjúsági Díjban részesült.



KEMÉNY ILDIKÓ a Budapesti Corvinus Egyetem Digitális Marketing Tanszékének habilitált docense. PhD fokozatát 2015-ben szerezte. Az elmúlt években legfőbb kutatási fókuszja az omnichannel vásárlói magatartás megértésére és modellezésére irányul, melynek támogatására kutatótársaival megalapították a hazai Omnichannel kutatócsoportot. További kutatási területei az online fogyasztói magatartás, valamint az e-health páciens oldali elfogadása, valamint a hollandiai University of Twente vendégkutatójaként modellezési módszer-

tani kutatásokban is aktívan részt vesz.

Szakmai tapasztalai a piackutatás területén is szerteágazóak: több mint 15 éve vesz részt vállalati piackutatási projekteken. Oktatási tevékenysége zömében a marketingkutatás tanítására fókuszál.

KUN ZSUZSANNA szakmai útjának jellemzője az akadémiai oktatás-kutatás és a gyakorlati marketing megoldások ötvözete. Közgazdász diplomáját 2006-ban szerezte a Budapesti Corvinus Egyetemen. Ezt követően, 2009-ben közgazdász tanári diplomát is szerzett. Jelenleg a Budapesti Corvinus Egyetem PhD hallgatója, fokozatát várhatóan 2024-ben szerzi meg. 10 évig non-profit szervezetek marketing menedzsereként dolgozott, majd az AC Nielsen-nél piackutatási tanácsadóként tevékenykedett. Zsuzsanna ezzel párhuzamos másfél évtizedes oktatói pályafutása során is széleskörű tapasztalatokat szerzett. Az általa az ország több egyetemén oktatott tantárgyak a marketing, a piackutatás és a kutatómódszertan spektrumát fedik le magyar és angol nyelven is.





PUSZTAI TAMÁS elismert conjoint szakértő, jelenleg értékesítés elemző az Allianz Hungária Zrt-nél. Diplomáját a Budapesti Corvinus Egyetemen szerezte, érdeklődése már ekkor a marketingkutatási módszertan irányába fordult. A hazai kvantitatív piackutatásban 18 évet töltött a Research International Hoffmann, a TNS Hoffmann, majd a Kantar Hoffmann színeiben, az utolsó években kutatómódszertani vezetőként. A versenyszférában betöltött pozíciója mellett hosszú évekig vett részt a Budapesti Corvinus Egyetem piackutatás és kutatómódszertan tárgyainak oktatásában és jelentős részt vállalt a tantárgyi továbbfejlesztésekben. 25 éve foglalkozik a conjoint módszertannal mind a gyakorlatban, mind az egyetemi oktatásban és a multinacionális piackutató cégeken belül a hazai és nemzetközi vállalati továbbképzésekben.

A kérdőív és az adatbázis bemutatása

A könyvben egy megkérdezéses vizsgálatban használt kérdőíven és a hozzá tartozó adatbázison mutatjuk be az egyes módszerek elvégzésének lépéseit. A kvantitatív megkérdezéses vizsgálatához felhasznált kérdőív kifejlesztésére egy diplomamunka keretében került sor, felhasználására a diplomamunka szerzőjétől engedélyt kaptunk. Az online megkérdezésben hallgatóink vettek részt, ennek eredményeként kaptuk a felhasznált adatbázist, amely a megjelölt YouTube csatornán szintén rendelkezésre áll.

A kutatás célja a fogyasztói magatartás vizsgálata volt, az omnichannel értékesítéssel kapcsolatos fogyasztói szokások és attitűdök, valamint a fenntarthatóság fogyasztói megítélése, különösen annak a kérdésnek a tükrében, mennyire egyeztethetők össze a többcsatornás vásárlás és a fenntarthatósági célok. A kutatás indításakor a cél és a kutatási kérdések mellett hipotézisek megfogalmazására is sor került, amelyet az adatbázis alapján lehetett tesztelni.

A kérdőívben a bevezető bemutatkozás és kutatási cél rövid bemutatása után a szűrőkérdéssel kerül a célcsoport kiválasztásra, akik az elmúlt egy évben a kiválasztott öt termék kategóriában vásároltak: ruházat, kozmetikum, szórakoztató elektronika, könyv és ruházati kiegészítők.

Ezt követően a kérdőív a következő blokkokat tartalmazza:

- az általános vásárlási szokások az 5 termék kategóriában
- az okos vásárlói magatartás tényezői
- a fenntartható/környezettudatos omnichannel viselkedés jellemzői

- az online vásárlási döntést befolyásoló tényezők preferenciájára vonatkozó conjoint kártyák
- a fenntartható/környezettudatos vásárlás tényezői
- az általános fenntartható/környezettudatos magatartás jellemzői
- szociodemográfiai blokk

A megkérdezés online felületen történt, a megkérdezésben döntően egyetemi hallgatók vettek részt, az ösztönző nyereményjátékban való részvétel volt. A mintavétel módszere önkényes volt, a minta elemszáma 390 fő.

A minta megoszlása a demográfiai jellemzők szerint: az életkort tekintve az egyetemi életkor a jellemző, az átlagos életkor 22,7 év. A minta 52%-a férfi, 48%-a nő. A megkérdezettek 39%-ának az állandó lakhelye Budapest, a többiek Budapesten kívüli. A megkérdezettek mintegy fele (49%) dolgozik a tanulmányok mellett.

Arra a kérdésre, hogy melyik termék kategóriában vásárolt utoljára, a megkérdezettek több, mint fele (55%) a ruházati kategóriát jelölte meg, kozmetikumot 14%, szórakoztató elektronikai terméket 7%, könyvet 8%, ruházati kiegészítőt 16% vásárolt.

A kérdőív további, vásárlásra vonatkozó kérdései erre az utoljára vásárolt termék kategóriára vonatkoztak.

Az adatbázisra elvégezhető elemzés néhány lehetőségét az egyes elemzési módszereknél mutatjuk be.

A feldolgozáshoz használt statisztikai program: az SPSS

A számítógépen történő adatfeldolgozás is többféle program használatával lehetséges, mi ebben a könyvben az SPSS programot használjuk, illetve a program használatát ismertetjük. Természetesen tudjuk, hogy a statisztikai adatelemzésre több program is rendelkezésre állhat, a többségük valamilyen üzleti modell szerint működik, általában éves előfizetési díjjal elérhető. Megjelentek a nyílt forráskódú (ingyenesen hozzáférhető) programok is, amelyek közül az R program egyre szélesebb körben terjed, ismereteink szerint elsősorban a tudományos kutatási felhasználása jelentős. Mi azért választottuk elemzésünkhöz továbbra is az SPSS-t, mert mind az oktatásban, mind a gyakorlati kutatásokban a legszélesebb felhasználói körrel rendelkezik Magyarországon is.

Az SPSS rövidítés a világ minden bizonnyal legszélesebb körben használt statisztikai adatelemző programrendszerének rövidítése. Ezt még 1968-ban fejlesztette ki az amerikai Stanford Egyetemen Norman Nie, Delae Bent és Hadlei Hull.

Az SPSS rövidítés eredetileg a „Statistical Package for the Social Sciences” (statisztikai csomag a társadalomtudományok számára) rövidítést jelentette. Később a „Superior Performing Software System” rövidítésre változott. Időközben a név „Statistical Product and Service Solutions”-re változott, és a statisztika és a szolgáltatás integrálását célozta.

Az SPSS 2010-es IBM általi felvásárlását követően az SPSS for Windows neve PASW Statistics-ra változott. A PASW a „Predictive Analytics Software” rövidítése. A 19. verziótól kezdve az SPSS nevet újra bevezették. Elődjéhez hasonlóan a jelenlegi verzió is IBM SPSS Statistics terméknév alatt fut.

Az SPSS egy nagy teljesítményű statisztikai adatelemző és adatkezelő rendszer, grafikusan egyszerű felülettel. Kezelése egyszerű és részben önmagyarázó menüvel és grafikailag világosan kialakított párbeszédpanelekkel történik. Ez lehetővé teszi a gyors statisztikai elemzést a megfelelő módszertani háttérismeretek birtokában. A programtól azonban nem szabad elvárni, hogy statisztikai módszereket írjon elő. Ezért az elemzőnek kell olyan statisztika alapismeretekkel rendelkeznie, hogy sikeresen elemezhesse az adatokat. Ebben a könyvben ezért minden módszernél összefoglaljuk a legszükségesebb elméleti tudnivalókat, a hangsúlyt azonban az alkalmazásra tesszük, ehhez az IBM SPSS Statistics 29.0.1 verziót használjuk.

Az SPSS program és használatának részletes bemutatását tartalmazza magyar nyelven Ketskemény László- Izsó Lajos – Könyves Tóth Előd: „Bevezetés az IBM SPSS Statistics programrendszerbe” című könyv.

01 A MEGKÉRDEZÉSES VIZSGÁLAT ELŐKÉSZÍTÉSE, A KÉRDŐÍV ÖSSZEÁLLÍTÁSA

A társadalomkutatásban alkalmazott empirikus kutatások módszertani alapjai széleskörűek és többféle szempont szerint csoportosíthatóak. A különböző szempontok szerinti besorolások gyakran megfelelnek egymásnak vagy összehasonlíthatóak, de soha nem lehet teljes egyenlőségjelet tenni. A kutatás jellege szerint megkülönböztethetünk feltáró, leíró és magyarázó, ok-okozati kutatásokat, az adatok jellege szerint kvalitatív és kvantitatív kutatást, az adatgyűjtés módja és ideje szerint szekunder és primer adatgyűjtési módszereket. A feltáró jellegű kutatások nagyrészt szekunder vagy primer kvalitatív kutatási módszer használatát jelentik, de nincs teljes egyenlőség, kvantitatív kutatás is eredményezhet feltáró jellegű eredményeket, az eredmények általánosíthatóságától függően. Mindezek a módszertani jellemzők a társadalomkutatás különböző területein általánosan érvényesek, megjegyezve, hogy az egyes alkalmazási területeken lehetnek speciális jellemzők is, amelyeket az adott terület szakmai közössége kialakított és elfogad.

Ebben a könyvben a társadalomkutatás egyik olyan területére, a marketingre vonatkozóan mutatjuk be a kvantitatív kutatás során alkalmazott legfontosabb módszertani jellemzőket és módszereket, amely terület elképzelhetetlen folyamatos empirikus vizsgálatok nélkül. A marketingkutatásra vagy piackutatásra is több definíció létezik, mi erre is, mint a könyvben a legtöbb helyen, a Malhotra-Simon (2009): Marketingkutatás című könyv meghatározásait használjuk elsősorban amiatt, mert ez az a magyar nyelven is elérhető könyv, amelyet a hazai felsőoktatási intézmények a marketing mester programokban használnak, és a módszertan egyes részei, elsősorban a többváltozós statisztikai módszertan is elsősorban a mester és a doktori programokra járó hallgatók számára releváns. Természetesen a bachelor programokba járók számára is javasoljuk a módszertan rájuk vonatkozó részének a tanulmányozását.

A könyvet azonban – a módszertan széleskörben való felhasználhatósága miatt is – nem csak a marketing alap-mester és doktori programok hallgatóinak és a marketingkutatással, piackutatással elméletben és gyakorlatban foglalkozó szakembereknek szánjuk, hanem a társadalomkutatás többi területével is foglalkozó elméleti és gyakorlati szakembereknek, akik empirikus vizsgálatokat folytatnak és a könyvben bemutatott módszereket alkalmazzák.

Ebben a könyvben a marketingkutatásban alkalmazott megkérdezéssel vizsgálatok során gyűjtött adatok feldolgozását fogjuk bemutatni.

A marketingkutató: „Az információk szisztematikus, objektív feltárása, összegyűjtése, elemzése, közlése valamint felhasználása, amelynek célja a marketingtevékenység során felmerülő problémák megoldására irányuló vezetői döntések elősegítése” (Malhotra-Simon, 2009)

Az AMA (American Marketing Association) definíciója:

„A marketingkutató az a funkció, amely összeköti a fogyasztót, az ügyfelet és a nyilvánosságot a gyártókkal és a forgalmazókkal olyan információk segítségével, amelyek marketingproblémák és lehetőségek azonosítására és értelmezésére, az intézkedések létrehozására, finomítására és értékelésére, a teljesítmény nyomon követésére és a folyamat jobb megértésére szolgálnak. Meghatározza az e kérdések kezeléséhez szükséges információkat, megtervezi az információgyűjtés módszerét, irányítja és végrehajtja az adatgyűjtési folyamatot, elemzi az eredményeket, és kommunikálja a megállapításokat és azok következményeit”. (Jóváhagyva 2017-ben, www.ama.org)

Jelen könyvben a megkérdezéses vizsgálatokban alkalmazható módszertanról foglalkozunk, a társadalomkutatási és marketingkutatási folyamat jellegéről, felépítéséről csak összefoglalóan teszünk említést, feltételezzük, hogy az Olvasó tájékozódik azokból a magyar és nemzetközi szakirodalmakból, amelyek rendelkezésre állnak. Az ajánlott szakirodalmak listáját a könyv végén közöljük.

A marketingkutató folyamatának áttekintése

A megkérdezéses vizsgálatok több szempont szerint is csoportosíthatók, egyik lehetséges csoportosítás a céljaik, jellemzőik, módszereik megkülönböztetése a kutatás jellege szerint, ezt tartalmazza a következő táblázat:

1.1. táblázat Alapvető kutatási módszerek összehasonlítása

	Feltáró	Leíró	Ok-okozati
Cél	ötletek és szempontok felfedezése	piaci jellemzők vagy funkciók leírása	ok-okozati kapcsolatok és hatásmechanizmusok meghatározása
Jellemzői	rugalmas, változékony	specifikus hipotézisek előzetes kialakítása	egy vagy több független változó manipulálása
	gyakran a kutatás kezdő- és végpontja	előre tervezett és strukturált felépítés	egyéb közvetítő változók kontrollálása

	Feltáró	Leíró	Ok-okozati
Módszerek	szakértői megkérdezés	szekunder adat (kvantitatív elemzéssel)	kísérletek
	próbakérdezés	megkérdezés	
	szekunder adatok (kvalitatív elemzéssel)	panelvizsgálat	
	kvalitatív kutatás	megfigyelésből származó adatok	

Malhotra-Simon (2009), 3.2. táblázat

Az adatforrások jellege szerint megkülönböztethetünk szekunder és primer adatokat, a szekunder adatot más, nem az adott probléma céljára gyűjtötték, számos formája és forrása lehetséges. A primer adat olyan, amelyet az adott problémával kapcsolatos kutatásban nyerünk, legfontosabb módszerei a megfigyelés, megkérdezés és a kísérlet. A primer marketingkutatási módszerek közül a leggyakrabban alkalmazott módszer a megkérdezéses kutatás, amelyen belül megkülönböztethetjük a kvalitatív és a kvantitatív kutatási módszereket. A kvantitatív megkérdezéses kutatás a standard kérdőívvel történő megkérdezés, amelyekkel a megkérdezettekkel történő kommunikáció módja szerint önkítöltő vagy személyes vagy telefonos megkérdezés útján kapjuk meg az adatokat.

Napjainkban elsősorban a technika fejlődése következtében jelentős változások tapasztalhatók mind a fogyasztók magatartásában, mind az ezzel foglalkozó kutatási módszerekben. Ezekről a hangsúlyeltolódásokról és tendenciákról részletesen olvashatunk a Szűcs és társai által írt Marketingkutatás 4.0 (Szűcs et al. 2023) könyvben. A változásokat vizsgálva is a szerzők megállapítják, hogy a kérdőíves megkérdezés még mindig jelentős szerepet játszik az empirikus kutatásokban, a kommunikáció módjában az online és mobil típus a meghatározó.

A kérdőív az adatgyűjtés strukturált módszere, amely olyan írásban vagy szóban feltett kérdések sorozatából áll, amelyekre a válaszadók válaszolnak (Malhotra-Simon, 2009). A megkérdezés során, az adatfelvételhez egyéb eszközök is kapcsolódhatnak, de ezek kiegészítik a kérdőívet, ha egyáltalán szükségesek. Ilyenek lehetnek például a személyes, kérdezőbiztos által végzett megkérdezésnél a kísérő kártyák, apró ajándékok a megkérdezett számára. A megkérdezés legfontosabb eszköze a kérdőív, amelyet nagyon gondosan kell összeállítani, ellentétben azzal az időnként hallható hiedelemmel, hogy a kérdőív összeállítás olyan rutin feladat, amelyben kérdéseket kell feltenni, a lényeg úgy is a válaszokban, az adatokban van. Az állításnak az a fele igaz is lehet, hogy az elemzéshez az adat a legfontosabb, viszont ha valóban olyan adatokat akarunk kapni, amelyek a legtöbb

információt hordozzák egy adott kutatási probléma szempontjából, akkor a kérdőív szerkesztés szabályait ajánlatos figyelembe venni.

Jelen fejezetben nem tárgyaljuk az összes szabályt, amelyek elsősorban a kérdőív tartalmát és formáját érintik. A hangsúlyt azokra a tényezőkre helyezzük, amelyeket a módszertan szempontjából, tehát a tervezett adatfeldolgozási módszerek szempontjából ajánlott figyelembe venni.

A kérdőív felépítése

A kérdőív szerkezete a kutatási kérdések és probléma megfogalmazását tükrözi, amelyet a gyakorlati kutatásokban a brief tartalmaz. A kérdések kiválasztásának és megtervezésének együtt kell járnia az elmélet alapján megfogalmazott kutatási kérdéssel és az értékeléssel kapcsolatos megfontolásokkal.

A kérdőív szerkesztésének lépéseit a következő szempontok befolyásolják:

- a szükséges információk köre, amelyet a kutatási brief és kérdések alapján határozhatunk meg
- a megkérdezettekkel való kommunikáció típusa, a kérdezési módszer választása
- az összegyűjtött adatok tervezett feldolgozási módszere (a tervezett egykét- és többváltozós elemzések milyen típusú adatokkal végezhető a legmegfelelőbbben)

A kérdőív szerkesztésekor meg kell határozni a kérdések tartalmát, a kérdések feltevésének módját, fontos ugyanis, hogy a megkérdezett tudjon és hajlandó legyen a kérdésre válaszolni, de nagyon fontos ebben a lépésben a kérdés típusáról és a skálázásáról is dönteni. A kérdés típusa szerint lehet nyílt vagy zárt kérdés, a kérdőív kérdéseinek többségében zárt kérdést használunk, részben a kvantitatív megkérdezés jellege miatt, részben a válaszadói hajlandóság növelése illetve fenntartása érdekében. Ez nem jelenti azt, hogy a kérdőívben egyetlen nyílt kérdést sem használhatunk, de érdemes a nyílt kérdések számát minimalizálni és csak ott alkalmazni, ahol feltétlenül szükségesnek tartjuk. A nyílt kérdések feldolgozása is hosszadalmasabb, de megoldható a feldolgozás, vagyis nem a feldolgozás nehézsége miatt nem javasolható a nagyobb számú nyílt kérdés alkalmazása.

A kérdések megfogalmazása után meg kell határozni a kérdések sorrendjét, itt szokás javasolni az ún. bemelegítő kérdéseket, amellyel a megkérdezettet bevezetjük a témába, szokás javasolni a tölcser elvet is, amely szerint az általános kérdésektől jutunk el a konkrétabb kérdésekig.

A könyvben használt, az online és offline vásárlói magatartásra, valamint a környezettudatosság mértékét és a vásárlói magatartásra gyakorolt hatását vizsgáló megkérdezés kérdőívében a fokozatosság elve és a bemelegítő kérdések elve is érvényesül.

A kérdések **funkciója** alapján a következő főbb **kérdéstípusokat** különböztetjük meg:

- Szűrő
- Bevezető
- Átvezető
- Tárgyra vonatkozó
- Ellenőrző
- Demográfiai

A szűrő kérdés is fontos a kérdőív készítésekor, a szűrő kérdés a megkérdezés célcsoportját választja ki. Ebben az esetben azok a megkérdezettek tartoznak a célcsoportba, akik az elmúlt évben vásárolták azt az öt termék kategóriát, amelyre a megkérdezésünk vonatkozik. A célcsoport gyakran a megkérdezettek szociodemográfiai jellemzői alapján kerül meghatározásra, ehhez sok esetben hasznos más szempontokat is kapcsolni. Ilyen szempont egy vásárlással, fogyasztással kapcsolatos megkérdezésben azt az időtartamot megjelölni, amelyre nagyobb valószínűséggel emlékeznek a válaszadók, ezzel nagyban növeljük a válaszok megbízhatóságára vonatkozó követelményt. A tárgyalt kérdőívben azt tételezzük fel, hogy az utóbbi egy évre vonatkozóan a megkérdezettek megfelelő pontossággal tudnak és akarnak válaszolni. Hasonló szempont lehet, amelyet meg kell határoznunk, hogy egy terméknek a fogyasztóját vagy/és vásárlóját akarjuk megkérdezni, ez ugyanis gyakran nem esik egybe, a szűrő attól függ, hogy a kérdőív vagy a kérdések egy csoportja melyik tevékenységre vonatkozik.

A megkérdezettel való kommunikáció módja meghatározhat egyes utasításokat, ez a kérdőív online kommunikációval megkérdezett, vagyis önkitöltő módon történt a megkérdezés. Ez azt jelenti, hogy szükséges egy bevezetővel kezdeni, amely röviden bemutatja a megkérdezést és fel kell tüntetni a válaszadás önkéntességére vonatkozó mondatot. A válaszadók bizalmát növelheti, ha leírjuk, hogy a válaszadás anonim, az adatokat bizalmasan kezeljük és harmadik személynek nem szolgáltatjuk ki. Az adatgyűjtésre és megőrzésre vonatkozó szabályokat a gyakorlati piackutatás számára az ESOMAR előírásai szabályozzák (<https://esomar.org/code-and-guidelines/icc-esomar-code>).

A kérdőív a szűrővel és az általános vásárlási szokásokra vonatkozó kérdésekkel kezdődik, amelyben az első kérdés a „mit vásárolt”-ra vonatkozik.

Szűrő kérdés

- I. Az elmúlt egy évben vásárolt-e ruházati (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába), kozmetikai vagy szórakoztató elektronikai (pl.: mobiltelefon, kamera, e-book stb.) termékeket?
- Igen
 - Nem ► kérdőív vége

Általános vásárlási szokások 3 termék kategóriát tekintve: Ruházat, kozmetikumok, szórakoztató elektronika.

Q01 – Mely terméket/termékeket vásárolt az alábbiak közül az elmúlt egy évben? Több választ is megjelölhet!

- Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)
- Kozmetikum
- Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book stb.)
- Könyv
- Ruházati kiegészítő (cipő, táska és egyéb kiegészítők)

A kérdések lehetnek olyanok, amelyre egy választ várunk a megkérdezettől, ebben az esetben egy válaszlehetőséget adunk számára. Ilyen a Q01 kérdés, a Q09, amely az online vásárlási kategóriára kérdez, vagy a Q19:

Q19 – Találkoztál-e már a vásárlásaid során olyan termékkel, amelyen valamilyen formában feltüntették a kézhez vétel (kiszállítás, bolti átvétel, csomagponti/csomagautomatás átvétel stb.) környezetbarát jellegét?

- Igen, viszonylag gyakran.
- Ritkán.
- Nem, még soha.

Ezek mind olyan kérdések, ahol minőségi (szövegesen megfogalmazott) alternatívák között választhat a megkérdezett és egyet kell kiválasztania. Ezen a kérdéstípuson belül szokás megkülönböztetni az alternatív és szelektív kérdéseket. Az alternatív kérdés olyan, ahol két alternatíva közül kell választani (dichotóm kérdésnek is szokás nevezni):

Szokott-e Ön élelmiszert online vásárolni?

- igen
- nem

A szelektív kérdéseknél felsorolunk válaszlehetőségeket, amelyek közül a megkérdezettnek vagy egyet kiválasztania, vagy többet is megjelölhet. A válaszlehetőségeket megfogalmazhatjuk elméleti megfontolások, szakirodalmi ajánlások alapján vagy kvalitatív kutatási eredményeket felhasználva, amelyek többnyire saját, primer kutatások.

Példa a több választ is megjelölhető szelektív kérdésre:

**Q02 – Hol szoktad vásárolni az alábbi termékkategóriákat (elmúlt egy évben)?
Több válasz is lehetséges!**

Q2-1 Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)

1. Hipermarketben (pl. Tesco, Auchan)
2. Szupermarketben (pl. Spar)
3. Diszkontban (pl. Aldi, Lidl)
4. Fizikai szaküzletben (pl. ruhabolt, könyvesbolt, drogéria)
5. Általános webáruházban (pl. EMAG, Alza)
6. Online szakáruház (pl. online könyvesbolt/ruhabolt/drogéria)
7. Egyéb helyen:

Ez a kérdés egyválaszos is lehetne, ha például azt kérdezzük, hogy **leggyakrabban** hol szokta vásárolni a terméket a válaszadó. Nagyon fontos hangsúlyozni az egyéb lehetőség felkínálását, akár egy- akár többválaszos a kérdés. Ez azért fontos, mert a lehetőségek legkörültekintőbb felsorolása esetén is előfordulhat, hogy egy adott válaszadó egyik alternatívát sem érzi magáénak, emiatt a válaszadási hajlandósága is lecsökkenhet, akár meg is szakíthatja a kitöltést. Ezzel együtt meg kell jegyeznünk, hogy az egyéb válaszok általában nem nagy számban fordulnak elő. Az egyéb lehetőséghez való beírás felkínálása miatt ezt a kérdést vegyes típusúnak is nevezik, vagy kvázi-nyitottnak is tekinthetjük, de a kérdést mégis csak a zártan felkínált válaszlehetőségek dominálják.

Hasonlóan érdemes gondolni egyes kérdések esetében a „nem tudom” vagy a „nem akarok válaszolni” lehetőségre is. Mindkettő valós alternatíva lehet a válaszadók számára, ugyanakkor mérlegelni szokás, különösen a „nem tudom” válaszlehetőség felkínálását is, attól tartva, hogy a válaszadót elterelheti a könnyebb megoldás felé, ugyanakkor sok esetben érdemes felkínálni, mivel ezzel szűkítjük a válaszadó lehetőségeit abban az esetben, ha valóban nem tudja a választ a kérdésre.

Az eddig említett kérdéseket mindet úgy kellett megválaszolni, hogy egy vagy több lehetőséget választ a megkérdezett, vagyis az egyes lehetőségekről eldönti, hogy igen vagy nem, választja vagy nem, és így jelöli-e azt a válaszlehetőséget.

Lehetséges olyan kérdés, ahol minőségi vagy mennyiségi ismérveket rangsorolni kell, vagyis általában számokkal jelölni (minőségi jelölés is lehetséges, de azt ritkábban alkalmazzák), hogy melyik válaszlehetőséget preferálom leinkább vagy melyikre illik leginkább a kérdésben megfogalmazott állítás.

Azt a kérdéstípust, amelyben egyes válaszlehetőségeket külön-külön többféle érték megjelölésével kell értékelni, értékelő skálának is nevezhetjük. Sokszor erre a kérdéstípusra használják a „skála-kérdés” megnevezést, amely nem hibás,

de a későbbiekben megmutatjuk, hogy minden zárt kérdés valamilyen skálatispusba sorolható. Az értékelő skálákat tartalmazó kérdésben egy vagy több állítást kell értékelnie a megkérdezettnek.

A Q04-es kérdésben az omnichannel vásárlási magatartást leíró magatartási típusok esetében kell értékelnie a megkérdezettnek minden egyes lehetőséget aszerint, hogy mennyire jellemző rá:

Q04 – Gondolj az előző kérdésben megnevezett kategóriára! Mennyire jellemzőek rád a következő tevékenységek az előző kérdésben megjelölt kategória vásárlása során? Kérlek értékeld 1-től 7-ig skálán, ahol az 1 jelenti, hogy egyáltalán nem jellemző, a 7 pedig, hogy teljes mértékben jellemző.

1. Ha az interneten keresek, legtöbbször meg is veszem a kiválasztott terméket valamelyik webáruházban.
2. Sokszor az interneten keresem meg a terméket, amit meg akarok venni, de utána mindig elmegyek egy boltba és ott vásárlom meg.
3. Ha a boltban keresek, majd meglátok egy nekem tetsző terméket, ott helyben meg is veszem.
4. Ha a boltban keresek, majd meglátok egy nekem tetsző terméket, többnyire megkeresem az interneten és online veszem meg.

A kérdésnél megemlítjük, hogy ha öt termék kategóriáról is érdeklődünk a kérdőívben, fontos ebbe a magatartási kérdésbe beleírni, melyik termék kategóriára vonatkozóan nyilatkozzon. Az egyes állításokat hétfokozatú Likert skálán értékelheti a megkérdezett, amelynél mindig bele kell írni a kérdésbe, hogy a skála végpontjai mit jelentenek. Az értékelő skála esetében többféle szempontot kell mérlegelnünk:

- páros vagy páratlan skálát használjunk: a páros skálának, mivel nincs „középpontja”, előnye lehet, hogy választásra kényszerítheti a válaszadót azzal, hogy nem lehetséges egy semleges válaszlehetőség jelölése. Ugyanakkor a gyakorlatban elterjedtebb a páratlan skálák használata, a magyar gyakorlatban például az ötfokozatú skála használata, mivel ez megfelel az iskolai osztályzatoknak, ahol – bár a kérdésfeltevésben itt is meg kell jelölni a végpontok jelentését – a válaszadók számára szinte automatikus, hogy az 5-ös érték a legnagyobb egyetértést, a legfontosabb, legjellemzőbb értékelést jelenti.
- az értékelő skála legyen kiegyensúlyozott: a pozitív és a negatív minősítésnek ugyanannyi értéke legyen, ha szöveget rendelünk az értékekhez, lehetőleg ugyanúgy fogalmazzunk a negatív és a pozitív értékek esetén.
- fontos, hogy a számokkal megjelölt értékelő skálánál melyek a számok: nem ugyanaz a hatás, ha negatív számok is előfordulnak, mint ha csak pozitív számok lennének. Bár a negatív és pozitív számok használatának

a feldolgozáskor vannak előnyei, de a kitöltő számára más lehet az érték percepciója.

A kérdések megfogalmazásakor számos szempontra kell a figyelmet felhívni, hogy ne kövessünk el hibát, csak néhány szempontot említünk:

- egyértelműen és lehetőleg egyszerűen fogalmazzuk meg a kérdést, kerüljük a szakszavak használatát. A hosszasan körülírt kérdések különösen online megkérdezésnél nem célszerűek.
- ügyeljünk arra, hogy egy kérdés ne tartalmazzon két állítást, amelyre válaszolni kell („egyérté- e azzal, hogy az adott termék jó minőségű és kedvező árú?” – melyikkel értsen egyet a megkérdezett).
- ne tegyünk fel tendenciózus kérdést, ne orientáljuk a válaszadót valamilyen irányba, ne tegyünk fel olyan kérdést, amely szinte sugallja a választ.

A megkérdezettel való kommunikáció módja szerint gyakran szó esik arról, hogy az online megkérdezésnél másképpen érdemes eljárni, különösen a mobiltelefonon történő megkérdezésnél hívják fel a figyelmet a kérdőív hosszára, a kérdések megfogalmazásának módjára. Valóban különbséget kell tenni már aszerint is, hogy a kérdőívet kérdezőbiztos által bonyolítottan kérdezzük, személyesen vagy telefonon, vagy a megkérdezett „kézbe kapja”, vagyis önkitöltő módon történik a megkérdezés. Az önkitöltő kérdőíveknel nagyobb a hiányos kitöltés lehetősége, amikor a megkérdezett egyszerűen abbahagyja a kitöltést. Emiatt is javasolható a nem nagyon hosszú kérdőív, de erről a mértékről eltérő véleményeket lehet hallani, az 5 perctől a 15 percig. Vannak eszközök, amelyekkel befolyásolni lehet a kitöltőt a teljes kitöltésre, az esetleges nyereményen kívül a kérdések „egyszerűségével”, vagyis jól érthető, nem hosszan fogalmazott zárt kérdésekkel és az is nagyon hasznos lehet, ha tájékoztatjuk arról, hogy mekkora részt töltött már ki és mennyi van hátra.

Még nem beszéltünk a kérdőívben a szociodemográfiai blokkról, amelyben a megkérdezett nemét, életkorát, lakhelyét, családi állapotát, foglalkozását, jövedelmi viszonyait (természetesen a téma jellegétől függően a kérdések összetétele változhat) kérdezzük, amely a kérdőív fontos alkotórésze az elemzés céljára. Klasszikusan az a felfogás, hogy ennek a blokknak a kérdőív végére kell kerülnie, mivel a megkérdezett személyére vonatkozó adatokat tartalmaz és egyrészt érdemes szétválasztani a megkérdezés tartalmi kérdéseit a személyes kérdésektől, másrészt egyes válaszadóknál elriasztó is lehet, hogy az elején úgymond mindjárt saját magáról „faggatják”. Ez a sorrend azonban megváltozni látszik, különösen azóta, amióta az online kérdőívszerkesztő programok, különösen az ingyenesen elérhetőek, az elején kínálják fel a szociodemográfiai blokkot. Emiatt sok felhasználó átveszi ezt a gyakorlatot, sőt, ezt tartja természetesnek. Feltehetően

nem befolyásol annyit negatív irányba a válaszadói hajlandóságon, mint amennyi előnyt az online kérdőív szerkesztők jelentenek a felhasználók számára.

A kérdőív összeállítását követően feltétlenül szükség van a tesztelésre, kipróbálásra, a próbakérdezésre. A próbakérdezést azon célcsoportból kiválasztott megkérdezettekkel kell elvégezni, akik a megkérdezésben a célcsoportot alkotják, valamint olyan módon kell velük kommunikálni, ahogyan majd a megkérdezés is történik. Nagyon fontos tesztelni, hogy minden kérdés érthető-e, a megkérdezett és a kutató ugyanazt ért-e egy-egy fogalmon, kérdésen, a megkérdezett tud-e válaszolni a kérdésre. A próbakérdezéskor kis mintán megpróbáljuk feltárni a lehetséges problémákat és megtalálni azt is, hogyan küszöböljük ki azokat. A próbakérdezés a kérdőív és a téma bonyolultságának függvényében lehet egy- és többlépcsős, a minta nagyságára vonatkozóan az ajánlás 15 és 30 fő között, de nyilván heterogénebb célcsoport esetében indokolt lehet nagyobb minta is. Nagyon fontos: a próbakérdezés a kérdőív tesztelésére szolgál, nem arra, hogy a válaszokat kiértékeljük és ennek alapján mintegy „elő-elemzést” végezzünk.

A kérdőív tartalmi véglegesítése magában foglalja azt a döntést is, hogy az egyes kérdéseket milyen módon értékeljük a válaszadókkal. vagyis milyen skálázottságúak legyenek a kérdéseink. A következőkben a skálázási lehetőségeket foglaljuk össze úgy, hogy nem a statisztikai jellemzőkkel foglalkozunk részletesen, hanem az adatfeldolgozás, a módszerek által támasztott igényeket domborítjuk ki.

Mérhetőség, skálázás

A kérdőívek alapján nyert adatok statisztikai elemzésének „nyersanyaga” a (rendelkezésre álló vagy még gyűjtendő) *adat*. Az adatok minőségét többek között a *mérés* módja határozza meg. Az adatok a mérési folyamatok eredményei. A mérés azt jelenti, hogy a tárgyak tulajdonságait bizonyos szabályok szerint számokban fejezik ki.

Az adatok és a mérés felveti az információk mérhetőségének problémáját. A megkérdezéses vizsgálatokban beleütközünk a mérhetőségi korlátba, amely elméleti és kutatásfilozófiai kérdéseket is felvet, amelyekkel nem foglalkozunk részletesen. Módszertani szempontból elsősorban az a kérdés, hogyan lehet a mérés megbízhatóságát a különböző mérési szintek alkalmazásával növelni, támogatni.

A mérések skálaszintje

Már említettük, hogy minden zárt kérdés valamilyen skálaszintnek felel meg.

A mérésre szolgáló eszközt, a „mérőrudat”, amelyen egy tulajdonság jellemzőit ábrázolják, *skálának* nevezzük. Attól függően, hogy egy tárgy tulajdonsága milyen módon fejezhető ki (mérhető) számokban, megkülönböztetünk különböző *skálaszintű* elsődleges mérési skálákat:

1. Névleges skála
2. Ordinális skála
3. Intervallum skála
4. Arányskála.

A skála szintje meghatározza mind az adatok információtartalmát, mind a matematikai műveletek alkalmazhatóságát.

1.2. táblázat

Skála		Jellemzők	Lehetséges műveletek
Nem metrikus	Nominális/névleges	Osztályozás minőségi ismérvekkel	gyakoriságszámítás
	Ordinális/sorrendi	Rangsorolás értékek hozzárendelésével	Medián, kvartilisek
Metrikus	Intervallum	Skála egyenlő nagyságú közökkel	Kivonás, átlagérték
	Arány	Skála egyenlő nagyságú szakaszokkal és természetes nullaponttal	átlagszámítás

A *nominális skála* a legegyszerűbb mérési szintet jelenti, példaként egy futóversenyhez kapcsolódó számozást mutatunk be. Vegyünk amatőr utcai futóversenyt, ahol az indulásra jelentkezők kapnak egy indulási számot, amely teljesen esetleges, sem a futó adataival, sem a későbbi teljesítménnyel nem függnek össze. Ez egy minőségű jellegű számhozzárendelés, amelyben a futók számai a nominális vagy névleges skálának felelnek meg. A megkérdésezés vizsgálatokban ilyen tipikus kérdések, amelyek nominális skálázottságúak: a megkérdézett neme: férfi-nő, a reklámhordozók: televízió – újságok – óriásplakátok.

A nominális skálák a minőségi jellemzők osztályozását jelentik. A feldolgozás során ezeket a jellemzőket kódolják, a kódolás önkényes módon történik, a kódszámokat másképp is ki lehetett volna osztani, amennyiben ez a kiosztás egyértelmű, vagyis amennyiben egy szám pontosan egy jellemzőt határoz meg. A nominális adatokkal a gyakoriságok csak a jellemző értékek (vagy az azokat reprezentáló számok) megszámlálásával határozhatók meg.

Az *ordinális skála* már lehetővé teszi a rangsor felállítását rangértékek (azaz ordinális számok) segítségével. A rangsort jelölhetjük rangértékkel vagy szöveges minősítéssel is.

A futóversenyek példában az, hogy valaki 1., 2., 3. stb. helyen ért célba, semmit nem mond az egymás közti eltérésről vagy az időteljesítményről, de a teljesítménykülönbséget mutatja, a rangértékek semmit sem mondanak az objektumok közötti távolságokról.

Az ordinális skálánál a gyakoriság mellett az eloszláshoz tartozó egyes statisztikai mutatókat is értelmezni lehet: módusz, medián, kvartilisek. Felhívjuk a figyelmet arra, hogy az SPSS az ordinális skálázottságú változóknak is kiszámítja az átlagát, ha – hibásan – beállítjuk a statisztikai mutatók opciójában, de ne elemezzük vagy értelmezzük ezt az átlagértéket.

A nominális vagy névleges és az ordinális vagy sorrendi skálák az un. nem-metrikus skálák kategóriájába tartoznak.

A következő, magasabb mérési szintet jelentő szint az *intervallum skála*, amely diszkrét értékekből áll, amelyek távolsága egyenlő. A futóverseny példáján: ha valaki skálán értékeli a futókat valamilyen szempont szerint, például a futásuk rugalmassága, szépsége szerint, és az értékelésnek az idejükhöz még nincs köze, de az értékelési szempont szerinti teljesítményekből már a helyezés mellett a teljesítmények távolságát is mérni tudjuk. Skála például lehet ötfokozatú, ebben az esetben mindig jelezni kell, hogy az 1-es és az 5-ös érték mit jelent. Tipikus példaként a hőmérsékleti skálákat szokás említeni, ahol is a Celsius, a Fahrenheit és a Kelvin skála is ugyanolyan távolságokra kalibrált, viszont az eltérő nulla pont következtében ha csak egy számot hallunk, nem biztos, hogy meg tudjuk állapítani a hőmérsékletet. Felhívjuk ugyanakkor a figyelmet a megszokás jelentőségére, vagyis ha egy adott területen, régióban, kulturában valamilyik skála használata a megszokott, például Európában a Celsius skáláé, akkor a hőmérséklet szóbeli közlésekor mindenki a Celsiusban mért fokra fog gondolni (írásban feltüntetésre kerül). A megszokás miatt is alkalmazzák gyakran az iskolai osztályzatok szerinti értékelő intervallum skálát.

Az *arányskála* abban különbözik az intervallum skálától, hogy van természetes nulla pontja. Ha a futóversenyen az egyes időeredményeket összehasonlítjuk, nem kell hozzátenni, hogy melyik vonatkozási rendszerben mértük, mert mindegyik értéknél azonos a nulla pont, ahonnan mérjük az időt. A megkérdésezés vizsgálatoknál arányskálában mérjük az életkort (ha nem kategóriákat adunk meg), megkérdézhetjük a költségeket és a jövedelmet is arányskálázott változóként. Az arányskálázott adatok esetében nemcsak a különbség, hanem az adatok hányadosa vagy aránya (innen az elnevezés) is információtartalommal bír a nullpont rögzítése miatt.

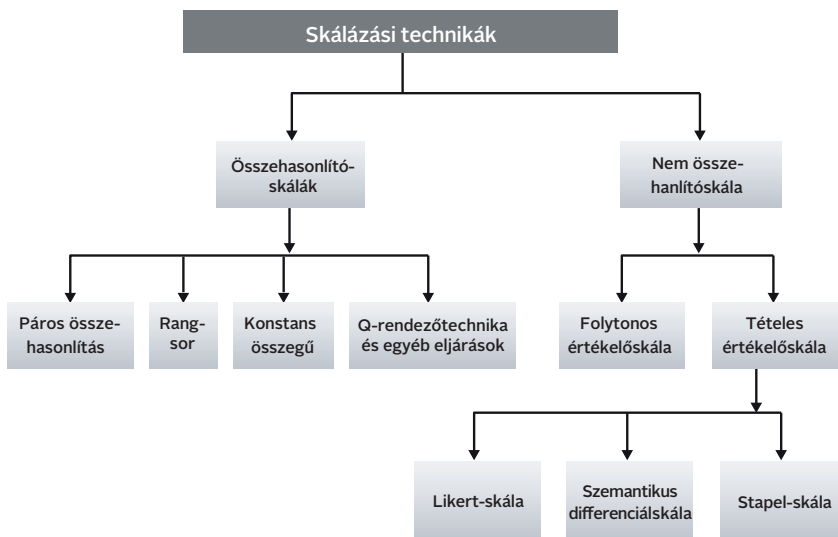
Az intervallum (interval scale) és az arányskála (ratio scale) a metrikus, mérhető skálák kategóriájába tartozik. Ezekkel az értékekkel már az átlagszámítás is elvégezhető, valamint több olyan statisztikai elemzési módszer is van, amely vagy metrikus adatokat vagy metrikus adatokat is igényel.

Minél magasabb a skála szintje, annál nagyobb az érintett adatok információtartalma, és annál több számtani művelet és statisztikai mérés alkalmazható az adatokra. Általában lehetséges az adatok magasabb skálaszintről alacsonyabb skálaszintre történő átalakítása, de fordítva nem. Ez hasznos lehet az adatok áttekinthetőségének növelése vagy elemzésük egyszerűsítése érdekében. Például gyakran képeznek jövedelmi vagy árkategóriákat. Ez magában foglalhatja az eredetileg arányskálájú adatok intervallum-, ordinális vagy nominális skálára történő átalakítását. Az alacsonyabb skálaszintre történő átalakítás természetesen mindig információvesztéssel jár.

Egy változó értékei lehetnek mérések vagy számítások eredményei (elsődleges vagy származtatott adatok). A változókat ezért a mérésük skálaszintje szerint is lehet kategorizálni (számos más osztályozás is létezik). A változókat megkülönböztethetjük aszerint, hogy minőségi (string) vagy mennyiségi (numeric) jellegűek, illetve a skála szintjük szerint metrikus vagy nem metrikus kategóriába tartozhatnak.

Az elsőrendű mérési skálák mellett a skálákat a *skálaképzési technikák szerint* is megkülönböztethetjük. Ezeket a skálaképzéseket, a skálák jellemzőit nem tárgyaljuk, de megemlítjük, hogy vannak a marketingkutatásban gyakrabban és kevésbé gyakran alkalmazott skálák. A *nem összehasonlítható skálák* közül a gyakorlati piackutatásban leggyakrabban a Likert skálát alkalmazzák, ritkábban a szemantikus differenciált, a Stapel skálát, feltehetőleg a negatív értékek problematikus alkalmazása miatt igen ritkán. Az *összehasonlítható skálák* közül szinte mindegyik alkalmazásával találkozhatunk, de nem nagyon gyakran és inkább speciális módszerek alkalmazása esetén, mint a conjoint módszernél is. A *folytonos értékelőskálák* használata főleg az online önkitöltő megkérdezésem módszerek terjedése következtében gyakoribb, mint korábban, ebben az esetben egy „csuszkaival” kell a megkérdezettnek megjelölnie az értékelését. Ez valóban árnyaltabb értékelést tesz lehetővé, mint a diszkrét skála, a feldolgozás szempontjából itt is egy értéket veszünk figyelembe, amely a jelölőszakasz végpontja.

1.1. ábra



Forrás: Malhotra, Simon (2009), 8.2. ábra

Az értékelő skála esetében többféle szempontot kell mérlegelnünk:

- páros vagy páratlan skálát használunk: a páros skálának, mivel nincs „központja”, előnye lehet, hogy választásra kényszerítheti a válaszadót azzal, hogy nem lehetséges egy semleges válaszlehetőség jelölése. Ugyanakkor a gyakorlatban elterjedtebb a páratlan skálák használata, a magyar gyakorlatban például az ötfokozatú skála használata, mivel ez megfelel az iskolai osztályzatoknak, ahol – bár a kérdésfeltevésben itt is meg kell jelölni a végpontok jelentését – a válaszadók számára szinte automatikus, hogy az 5-ös érték a legnagyobb egyetértést, a legfontosabb, legjellemzőbb értékelést jelenti.
- az értékelő skála legyen kiegyensúlyozott: a pozitív és a negatív minősítésnek ugyanannyi értéke legyen, ha szöveget rendelünk az értékekhez, lehetőleg ugyanúgy fogalmazzunk a negatív és a pozitív értékek esetén.
- fontos, hogy a számokkal megjelölt értékelő skálánál melyek a számok: nem ugyanaz a hatás, ha negatív számok is előfordulnak, mint ha csak pozitív számok lennének. Bár a negatív és pozitív számok használatának a feldolgozáskor vannak előnyei, de a kitöltő számára más lehet az érték percepciója.

Külön is érdemes megemlíteni az ún. **többszintű skálát**, amelyről a kérdőív-nél már volt szó. A többszintű skála olyan állításhalmaz, amely egy vagy több fogalom mérésére szolgál, amelyek látens változók, közvetlenül nem célszerű őket megkérdezni. A marketingben ezek a fogalmak általában a fogyasztói vagy vásárlási magatartásra, a fogyasztó attitűdjére vonatkoznak és elemzésük nagyon fontos a kutatás számára. Hasonlóan fontos a mérhető változóknak, amelyeket manifeszt vagy indikátor változóknak is szokás nevezni, a minél pontosabb mérése. A többszintű skálák alapja a fogyasztói magatartásnak valamelyik elmélete, a kifejlesztésnek több lépése van, amelyet statisztikai mutatók tesznek módszertanilag megalapozottá. A piackutató intézetek gyakran alkalmaznak olyan termékeket, amelyek valamilyen fontos marketing jelenséget mérnek, mint a márkaérték, a fogyasztói elégedettség, az életmódra vonatkozó mérések, ezekben olyan többszintű skálákat használnak, amelyeket a skáafejlesztés szigorú eljárásával fejlesztettek ki a cég fejlesztő részlegében. A gyakorlati kutatásokban gyakran alkalmaznak a kutató által a témakörhöz összeállított saját állításhalmazt, amelynek különösen fontos a skála értékelhetősége.

A skálák értékelése a *valódi értékmodell elve* alapján történik, amely azt foglalja képletbe, hogy a mérések esetében a mért érték eltér a valóságos értéktől, az eltérést véletlen és szisztematikus hiba okozhatja. A skála értékelésbe a megbízhatóság, az érvényesség és az általánosíthatóság tartozik. A megbízhatóság arra vonatkozik, hogy a skála mennyiben ad konzisztens eredményt ismételt mérés esetében. A skála belső konzisztenciájának mérésére többféle módszer használhatunk, a leggyakrabban alkalmazott a Cronbach alfa mutató számítása és értékelése. A Cronbach alfa mutató a skálaértékek összes lehetséges kettéosztásából adódó korrelációs koefficiensnek átlaga. Az együttható értéke 0 és 1 közé eshet, a 0,6 alatti értékek esetén (gyakorlati kutatásokban) nem megfelelő a belső konzisztencián alapuló megbízhatóság.

Két, tapasztalatból adódó ajánlás:

- a fogyasztói magatartásra, attitűdre vonatkozó állításoknál az elmélet egyik ajánlása, hogy részben a megkérdezett figyelmének fenntartására, részben a válaszok konzisztenciájának ellenőrzése céljából is érdemes az állítások előjelét időnként változtatni, ami azt jelenti, hogy a pozitív állítások mellett negatív előjelű (fordított megfogalmazású) állításokat is használni. Ezt a hazai gyakorlatban nem javasoljuk követni, részben tudományos kutatásaik alapján szolgáltatásmarketinggel foglalkozó szakemberek is óvatosságot javasolnak (Kenesei, Kolos, 2014), részben saját tapasztalataink is azt erősítik meg, hogy nálunk a válaszadók általában nem figyeltek az állítás megváltozott jellegére, emiatt sok olyan válasz született, ami figyelmetlenségből nem vette észre az állítás fordított jellegét. Mivel ezt

azonban az adatokból nem tudjuk biztonsággal kiszűrni, emiatt inkább nem javasoljuk a negatív, a fordított állítások alkalmazását

- a másik gyakori törekvése egy kutatónak a kérdőív összeállításakor, hogy mivel minél több árnyalt információt szeretne, ezért túl sok intervallum vagy arány skálát helyez el a kérdőívben, amitől hosszabb is lesz a kitöltési idő, valamint a válaszadó elvesztheti a kedvét és félbehagyja a kérdőívet. Emiatt gondos mérlegelés és a próbakérdéskor az intervallum és arány-skálák számának a tesztelése is szükséges.

A marketingkutatási folyamatban a kérdőív véglegesítése és a kommunikációról való döntés mellett azt is meg kell határoznunk, hogy milyen kört fogunk megkérdezni, vagyis hogy a kitűzött célcsoport szerint meghatározott alapsokaság minden tagját meg tudjuk-e kérdezni vagy mintát veszünk az alapsokaságból. Az előbbi lehetőség nagyon ritka, jellemző eset a mintavételes megkérdezés. A mintavételnél a marketingkutatás a statisztikai mintavételi eljárásokat tekinti olyan alapnak, amelynek a törvényszerűségeit figyelembe veszi. Mivel ebben a könyvben az adatelemzési módszerekre fókuszálunk, ezért nem foglalkozunk részletesen a mintavételi eljárásokkal, amelyeket két nagy csoportba sorolhatunk, a véletlen és nem véletlen eljárásokba, mindkét nagy kategórián belül többféle mintavételi módszert különböztethetünk meg. A statisztikai törvények alapján a minta eredményeinek általánosíthatósága a véletlen eljárások esetében lehetséges, más mintavételi módszerek esetében az általánosíthatóság korlátozott vagy nem lehetséges. A marketingkutatásban gyakran nem lehetséges vagy nem célszerű statisztikai véletlen mintavételt alkalmazni, ez nem helytelen eljárás abban az esetben, ha mintavételünk korlátait ismerjük és az eredmények értelmezésénél alkalmazzuk. Ez gyakran azt eredményezi, hogy a kvantitatív megkérdezésünk eredménye korlátozottan általánosítható.

02 AZ ADATBEVITEL

Az adatbevitel előkészítésekor a legfontosabb teendő annak meghatározása, hogyan kerülnek be a kérdőív kérdései az adatmátrixba, amely egy többnyire Excelben előállított adattáblázat.

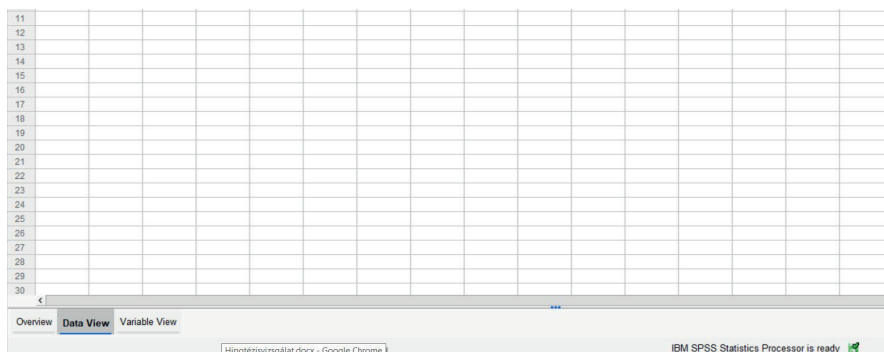
Ebben a folyamatban a kérdőív kérdéseire adott válaszok transzformációja történik olyan formátumba, amely lehetővé teszi a számítógépen SPSS segítségével történő adatfeldolgozást. A transzformációra többféle lehetőség adódhat, javaslatunk szerint a lehető legegyszerűbb és legelemibb módszert érdemes alkalmazni, a továbbiakban ezt követjük.

Az SPSS program adatbeviteli ablaka az Excel táblázatban szerkesztett adatbevitőnek megfelel, illetve az Excelbe felvitt adatokat be tudjuk olvasni, emiatt is az adatok kiindulásként Excelben való megjelenítési lehetőségeit mutatjuk be.

Nézetek és fájlok az SPSS-ben

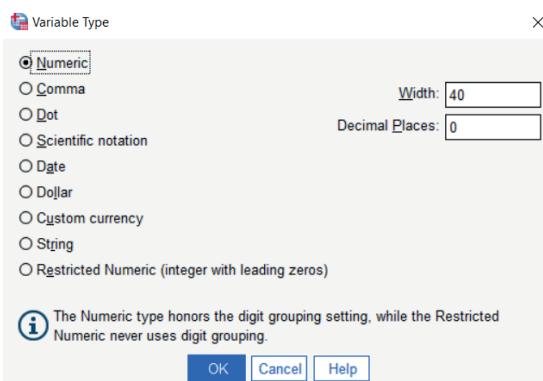
Mielőtt azonban belekezdenénk a tényleges bevitelbe, szeretnénk néhány technikai információt adni az SPSS kezeléséről, hogy megkönnyítsük a dolgunkat. Az SPSS többféle ablakot és fájlformátumot alkalmaz.

Az adatokat a .sav kiterjesztésű fájlba menthetjük és az adatszerkesztő (*data editor*) ablakban láthatjuk őket. Az SPSS indítását követően megjelenik egy üdvözlő ablak, ahol kattintson a „Mégse” gombra, hogy az üres adatszerkesztőhöz (üres adatmátrix) lépjen. Ezt követően megjelenik egy táblázatkezelő-szerű eszköz, de csak első pillantásra tűnik „egy másik Excelnek” Az összegyűjtött adatokat ebbe a mátrixba kell elmenteni. Ezen ablakon belül három fül található: *Overview*, *Data View*, *Variable View*. Ezen utóbbi kettőt mutatjuk be a könyv keretein belül.

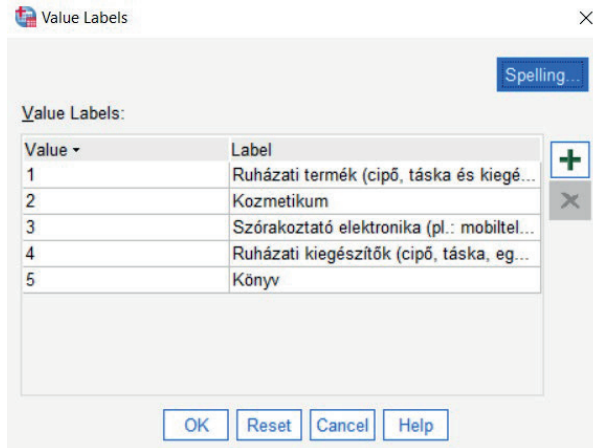


A *Data View* az adatok táblázatszerűen rögzített formáját mutatja be. Ebben az adatmátrixban minden sor egy kitöltött kérdőív, azaz egy kitöltő válasza, míg az oszlopok a változók. A *Variable View* a változókról rögzít információkat, az elemzés során ezt az nézetet használjuk gyakrabban. Olyan információkat rögzítünk itt, mint:

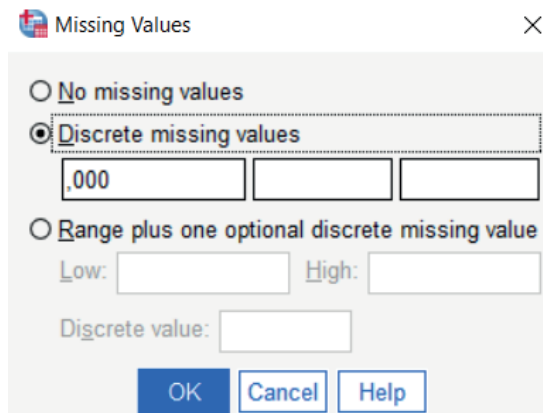
- *Name*: a változó neve, ami nem tartalmazhat szóközt és speciális karaktereket. Érdekes rövid neveket választani, ami később segíti a navigációt a sok változó között. Például az egy kérdéshez, vagy témához kapcsolódó változókat azonos kóddal kezdeni, és alul vonással indexálni, pl: Q01_1, Q01_2, Q01_3. A helyes névválasztás meggyorsítja a későbbiekben a változók kiválasztását, ezáltal az elemző munkát.
- *Type*: az adat típusát lehet beállítani. Számszerű adatnál a „*numeric*” kategória, míg szöveges adatnál a „*string*” kategória fog itt megjelenni.



- *Label*: a változó szöveges tartalmát jelöli, ez fog megjelenni az ábrákon és táblázatokon, ezért ide olyan szövegeket érdemes írni, amelyek rövidek ugyan, de segítik az elemzés olvasóját, hallgatóját, hogy megértse a vizsgált változóink tartalmát.
- *Values*: itt van lehetőségünk a korábban numerizált kategóriák mögé kódolni a szöveges tartalmakat. Ez azért fontos, mert a későbbi outputokon így nem a számkódok, hanem a szöveges tartalmak lesznek majd olvashatóak, ezzel is megkönnyítve az elemzés megértését. Például a Q9-es kérdésnél a legutóbb vásárolt online kategóriát ismerhetjük meg.

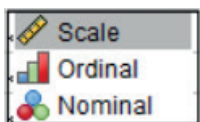


- *Missing*: a hiányzó adatok kódolásának többféle opciója állítható itt be. Lehetséges egy érték (*discrete*), vagy akár egy értéktartomány (*range*) hiányzó adatként való értelmezése.



- *Measure*: a változó mérési skálája rögzíthető itt, melyeket két okból is érdemes jól beállítani. Egyrészt bizonyos elemzések csak megfelelően beállított mérési skála esetén futnak le (pl. a nem paraméteres tesztek). Másrészt az egyes mérési szintekhez tartozó piktogramok szintén tudják segíteni a nagy mennyiségű változó esetében a navigációt, keresést. Érdemes megfigyelni, hogy míg négy mérési skála van (nominális, ordinális, intervallum, arány), addig az SPSS csak három mérési skála közül enged választani. A metrikus, számszerűsített intervallum és arányskála esetében csak annyit jelez, hogy „scale”. Ennek oka, hogy statisztikai értelemben

az intervallum skálával is elvégezhetőek azok a műveletek és hipotézisvizsgálatok, amik az arányskála esetében elvégezhetőek.



Az outputok egy másik ablakban, az ún. *Statistics Viewer* nevű ablakban jelennek meg, amelyet az SPSS .spv kiterjesztésű fájlként ment el. A képernyője kétféle osztott. Bal oldalon az addig lefuttatott elemzések listája van egy fa szerkezetbe rendezve, melynek elemei átnevezhetőek, illetve a +/- jelek segítségével összecsuksukhatóak, kibonthatóak. Az egyes outputok kijelölése is elvégezhető innen. A jobb oldalon magukat az eredményeket (outputok) látjuk. Egy-egy eredmény másol-beilleszt parancsokkal, billentyűkombinációkkal átemelhetőek másik szoftverekbe (pl: office alkalmazások) és ott akár tovább is formázhatóak. Meg kell említenünk, hogy az SPSS rendelkezik egy harmadik ablakkal is, ahol a programozott futtatásokat ún. *syntax*-okon keresztül szerkeszthetőek és futtathatóak, ez a könyv azonban nem tér ki a *syntax* írás és szerkesztés részleteire.

The screenshot shows the IBM SPSS Statistics Viewer interface. The left pane displays a tree view of the analysis outputs. The main window is divided into two sections: 'Frequencies' and 'Explore'.

Frequencies

[DataSet1] C:\Users\user\Downloads\adatbazis_elemzesre.sav

Statistics

Mely termék/termékeket vásárolt az elmúlt egy évben? Könyv

N	Valid	272
	Missing	118

Mely termék/termékeket vásárolt az elmúlt egy évben? Könyv

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Könyv	272	69,7	100,0	100,0
Missing System	118	30,3		
Total	390	100,0		

Explore

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Hol szoktad vásárolni? - Kozmetikum - Szupermarketben (pl. Spar)	38	9,7%	352	90,3%	390	100,0%

Descriptives

		Statistic	Std. Error
Hol szoktad vásárolni? - Kozmetikum - Szupermarketben (pl. Spar)	Mean	2,00	,000
	95% Confidence Interval for Mean	Lower Bound	2,00
		Upper Bound	2,00
	5% Trimmed Mean	2,00	

A kérdőívből az SPSS fájlba

A következő lépés az adatok elemzésre való előkészítése. Hogyan lesz a kitöltött kérdőívből szerkeszthető adatfájl, amely az SPSS segítségével történő elemzés alapjául szolgál? Az adatbevitelkor a kérdőívek változókká való átalakítása történik.

Fontos megjegyzések az adatok bevitele előtt

Mielőtt elkezdenénk az egyes változók meghatározását a változó nézetben, el kell végezni néhány fontos előkészítő munkát, hogy megkönnyítse az adatok kezelését.

Amennyiben papíralapú kérdőívet dolgozunk fel, számozzunk meg minden kérdőívet egy folyamatos számmal (ún. ID no. – identification designation vagy identification number). Ez biztosítja, hogy az eredeti kérdőívet bármikor ellenőrizni tudjuk, ha az adatsorban rendellenességeket vagy hiányzó értékeket találunk. Ha online módon történt a megkérdezés, akkor az adatbázisban automatikusan számozottak a kérdőívek.

A kérdőív kérdéseit változókká transzformáljuk, amelyhez meg kell határozni, hogy a kérdés hány változónak feleltethető meg. Ez követően mivel az SPSS numerikus megközelítéssel dolgozik, ezért a felvehető értékeket számkódolni szükséges, amennyiben az nem metrikus, számszerű adat. Ezt a folyamatot kódolásnak is nevezik. Ezt legegyszerűbben még Excel formátumban érdemes megtenni, azért is, mert az online kérdőívszerkesztők szinte mindegyike támogatja a .xlsx, vagy .csv formátumba való letöltést. Az alább bemutatott numerizálási átalakításokat legegyszerűbben a keresés-csere funkciókkal lehet megtenni az excelben, kijelölve az adott oszlopo(ka)t.

A transzformáció a kérdések megválaszolási módja szerint történik.

Ha egy kérdésre *egy válasz* adható, akkor azt egy változóba transzformáljuk, ha több, akkor annyi változóba, ahány válasz adható.

Ha a kérdésre *egy válasz* adható, akkor ennek a kérdésnek egy változót feleltetünk meg, például:

Q03 – Válassz ki egyet az alábbi kategóriák közül, amelyet a legutóbb vásároltál! (A további kérdések erre a kategóriára fognak vonatkozni.)

_Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába) → 1

_Kozmetikum → 2

_Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book stb.) → 3

_Könyv → 4

_Ruházati kiegészítő (cipő, táska és egyéb kiegészítők) → 5

Ez a kérdés egy változót fog jelenteni az adatbázisban: Q03, amely 1-től 5-ig vehet fel értékeket, ahol a kódszámok az egyes termék kategóriákra vonatkoznak.

Amennyiben egy kérdésben a válasz alternatívák közül *több is megjelölhető* (a válaszadók egy választ jelölnek vagy tetszés szerinti választ, akár az összeset is megjelölhetik), a legegyszerűbb adatbeviteli mód az, ha minden egyes alternatíva egy változót jelent, a változók értékei pedig kódszámok: jelölte a választ (1, vagy az opció sorszáma) vagy nem jelölte (0, vagy hiányzó adat).

Kérdőívünkben az első kérdésnél a megkérdezett több termék kategóriát is megjelölhetett:

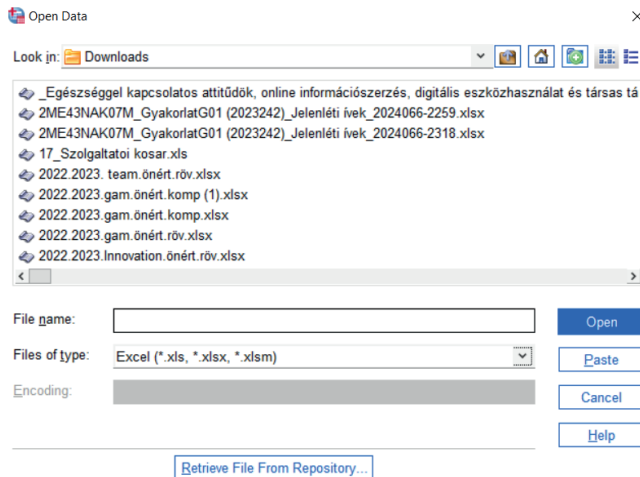
Q01 – Mely terméket/termékeket vásároltad az alábbiak közül az elmúlt egy évben? Több választ is megjelölhetsz!

1. Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)
2. Kozmetikum
3. Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book stb.)
4. Könyv
5. Ruházati kiegészítő (cipő, táska és egyéb kiegészítők)

Ennek megfelelően az adatbázisban ezt a kérdést 5 változó fogja megjeleníteni: Q01_1, Q01_2, Q01_3, Q01_4, Q01_5, ahol minden változó azt jeleníti meg, hogy az adott opciót választották-e.

A számformátummal megválaszolt kérdések esetében (pl: D2 – Melyik évben született?) nem szükséges további transzformáció.

Az így számokká alakított adatbázist importáljuk excelből az SPSS-be. Ezt a *file / open / data* menüpontokon keresztül tehetjük meg. Ahhoz, hogy megtaláljuk az elmenett Excel fájlt, a *Files of type* mezőt át kell állítsuk SPSS-ről (.sav) Excel formátumra (.xlsx).



Azt követő párbeszédablakot (*Read Excel File*) egyszerűen okézzuk, és betölti az adatokat az SPSS. Ezt követően a *Variable View*-ban még tegyük meg a szükséges beállításokat, ellenőrizve, hogy a beolvasás megfelelő módon történt-e meg. A „*label*”-be az excel fejlécből kerülnek a szövegek és ebből tömörít és rövidít egy változó nevet az SPSS, amit mindenképpen érdemes átkódolni könnyebben navigálható kódokra. Szintén szükséges beállítás még a „*values*” mezők beállítása, hogy a kategorizált változók esetében tisztában legyünk azzal, hogy melyik szám kód milyen szöveges tartalmat képvisel az adatbázisban. Ezt követően menthetjük az adatbázist, hiszen készen áll az elemzés első lépéseire.

Adatbázis műveletek

Szinte elképzelhetetlen, hogy az adatelemzés során ne lenne szükségünk a meglévő változók átalakítására, összesítésére, szűrésére. Az SPSS számtalan adatbázis műveletet tesz lehetővé, ezek közül a leggyakrabban használt négy műveletet mutatjuk be az SPSS alkalmazása során.

- *Select cases*: a változók szűrése, egyfajta filter funkció
- *Compute*: számítási műveletek a változókkal,
- *Recode*: a változók újrakódolása, átkategorizálása,
- *Split file*: az eredmények csoportosítása.

Data / Select cases A legtöbb elemzés során felmerül az igény, hogy az adatokat valami mentén szükséges lenne leszűrni, például csak azok a válaszadók érdekelnek minket, akik kollégiumban laknak (D4), vagy csak a személygyűjtést

szelektíven gyűjtők érdekelnek bennünket (Q20_03=5, 6, 7). A *data / select cases* menüpont alatt az „*If condition is satisfied*” opcióra kattintva a megfelelő változó kívánt értékeit, vagy értéktartományát választhatjuk ki. Az elemzendő értékeket kell kiválasztani, a többi kategóriát / értéket az elemzés következő szakaszában az SPSS nem veszi figyelembe. Azokat az eseteket, akiket nem kívánunk vizsgálni törölhetjük is az adatbázisból (*delete unselected cases*) vagy egyszerűen ideiglenesen kiszűrhetjük (*filter out unselected cases*). Arra is lehetőségünk van, hogy a leválogatott adatokat egy új adatbázisba helyezzük át (*copy selected cases to a new dataset*). Ezen három szűrési lehetőség között a legjellemzőbben az ideiglenes megoldást jelentő filterező megoldást használjuk. Több kategória egy változón belüli kiválasztásához VAGY kapcsolat kell, amit az „|” logikai jel használatával tudunk alkalmazni. Több feltétel egyidejű teljesüléséhez, amelyek több változó szerint vannak definiálva ÉS kapcsolat szüksége, amelyhez az „&” jel alkalmazására van szükség. A megadott szűrés egész addig érvényben marad, ameddig azt ki nem kapcsoljuk az „*All cases*” opcióval a menüpontban.

Transform / Compute Variable Gyakorta megesik, hogy az elemzés folyamán matematikailag kell kombinálnunk a változókat. Például szeretnénk összeadni őket, átlagolni, az arányukra vagyunk kíváncsiak, stb. Ehhez egy új változót kell létrehozni a *Transform / Compute Variable* menüpontban. A létrehozni kívánt új változó nevét a „*Target Variable*” mezőbe kell tenni. Itt érdemes olyan nevet és label-t választani, ami utal arra, hogy mely változókból és milyen művelettel képeztük (pl: name: Q05_8_9_MEAN, label: az árazás átlagos szerepe). A „*Numeric Expression*” menüben kell definiálni a matematikai műveletet. A function group alkalmazásával akár az előre beépített függvények közül választhatunk (pl: statistical: max, mean, mode, median, sum, variance, min). A számítással létrehozott új változó az adatbázis utolsó változója lesz. A folyamat befejezéseként még szükség lehet az új változó felcímkezésére a későbbi elemzések egyszerűsítése céljából.

Transform / Recode A meglévő változók átalakítására is szükség lehet, például a 7 elemű likert skálánkat át szeretnénk transzformálni egy 5 elemű likert skálává. Vagy szeretnénk létrehozni egy olyan változót, amiben életkorok alapján életkori csoportokat hozunk létre. Ezekben az esetekben segíthet az átkódolás, azaz az újrakódolás. A menüpontok között választhatunk a *recode into same variable* és a *recode into different variable* opciók közül. Javasoljuk, hogy a második lehetőséget válasszuk, mert ebben az esetben az eredeti változó, amit át kívánunk kódolni, nem változik meg, az első opció azonban átírja a kiinduló változót.

Az így létrehozandó új változó esetében meg kell adjuk annak a nevét és címkejét (label) is, ahol az elnevezések során szintén érdemes utalni a kiinduló változóra, valamint a transzformáció jellegére. Szükséges ezt követően a „*change*” gombot megnyomni, hogy aktivizálódjon az „*old and new values*” lehetőség. A bal

oldali beviteli mező (*old values*) a kiinduló értékeket, míg a jobb oldali beviteli mező (*new values*) az új kódokat tartalmazza. Az *old* ► *new* mezőben a már beállított változtatási logikát tudjuk nyomon követni. A létrehozott új változó ismét a változó lista végén lesz megtalálható, melynek a felcímzése feltétlenül hasznos.

Data / Split file Előfordulhat, hogy az eredményeket csoportosítva szeretnénk látni, például külön a férfiak és külön a nők eredményei érdekelnek. A csoportosítás során érdemes a csoportképző mérési szintű változókat használni (nominális vagy ordinális). A beállítást a *data / split file* menüpont alatt lehet megtenni, ahol a csoportképző változót a „*groups based on*” mezőbe kell tenni. Az egy outputon belüli csoportosítást a „*Compare groups*” lehetőség valósítja meg, míg az „*Organize output by groups*” opció minden kategória esetében létrehozza az outputot külön-külön. Amennyiben beállítottuk valamelyik opciót, az adattábla jobb alsó sarkában látható lesz a „*Split by*” felirat. A csoportok szerinti riportolás egész addig fennmarad, ameddig azt ki nem kapcsoljuk, amihez az „*Analyze all cases, do not create groups*” opciót szükséges választani.

03 LEÍRÓ STATISZTIKA

Az adatelemzés kezdetén a leíró statisztikai vizsgálatokat érdemes lefuttatni. A leíró statisztikai vizsgálatok során mindig egy-egy változó szerint ismerjük meg a vizsgálni kívánt mintát, amiből több előnyünk is származhat.

- A leíró statisztikai vizsgálat során könnyen fény derülhet adatfelvételi, vagy értelmezési hibákra. Ez segít megakadályozni, hogy hibás, vagy félreérthető eredményeket mutassunk be a vizsgálat későbbi szakaszaiban.
- A kiugró értékek (outlierek) a leíró statisztikai elemzés során egyértelműen meghatározhatóak grafikus ábrázolás segítségével (boxplot). Egyéb vizuális eszközök is segíthetnek megérteni az egyes változók alapvető jellegét, illetve gyakorisági táblák, leíró statisztikai mutatószámok további információkat adhatnak.

Az alkalmazható leíró statisztikai eszköztárat meghatározza a vizsgálat tárgyát képező változó mérési szintje.

- Nominális skála: gyakorisága tábla alkalmazható, hogy lássuk, az egyes opciókat hányan, és a minta milyen százalékában választották. A mutatószámok közül a módusz értelmezhető.
- Ordinális skála: gyakorisága tábla alkalmazható, hogy lássuk, az egyes opciókat hányan, és a minta milyen százalékában választották. A mutatószámok közül a módusz és a medián (a kvartilisek is) alkalmazhatóak
- Intervallumskála és arányskála: az arányskálán és intervallumskálán mért változók általában nagyon sokféle értéket vehetnek fel, ezért a gyakorisági tábla alkalmazható ugyan, de nem sűríti eléggé össze az információt. Ezért a táblázat helyett inkább az eloszlás ábrázolása terjedt el a gyakorisági hisztogram (sűrűségfüggvény) formájában. Az összes leíró statisztikai mutatószám is kalkulálható és értelmezhető az alább bemutatottak szerint.
 - Módusz: a minta leggyakrabban előforduló értéke vagy kategóriája.
 - Medián: a (növekvő sorrendbe állított) minta középső eleme, aminél a megfigyelések 50%-a kisebb, 50%-a pedig nagyobb értéket vesz fel.
 - Átlag: az adott változó mintabeli összegét osztjuk az elemszámmal.
 - Szórás: az átlagtól vett átlagos eltérés.
 - Variancia: a szórás négyzete.
 - Minimum: a minta legkisebb értéke.
 - Maximum: a minta legnagyobb értéke.
 - Tartomány: a maximum és a minimum különbsége.

- Kvartilisek: negyed részre osztó pontok. Az első kvartilis (Q1) esetén a megfigyelések 25%-a ennél kisebb, míg 75%-a ennél nagyobb értéket vesz fel; a második kvartilis (Q2) megegyezik a mediánal; a harmadik kvartilis (Q3) a megfigyelések 75%-ánál helyezkedik el, azaz háromnegyede a mintának ennél kisebb, míg negyede ennél nagyobb értéket vesz fel.
- Ferdeség (*skewness*): az eloszlás szimmetriáját vizsgáló mutatószám; 0 esetén az eloszlás szimmetrikus, pozitív szám esetén jobbra elnyúló (balra ferde), míg negatív esetben balra elnyúló (jobbra ferde) az eloszlás alakja.
- Csúcsosság (*kurtosis*): az eloszlás csúcsosságát, a szélső értékek elhelyezkedését vizsgáló mutatószám. Az értéke normális eloszlás esetén 0. Normális eloszlásnál lapultabb, ha az értéke negatív (azaz kevesebb szélsőértéket tartalmaz). Normális eloszlásnál csúcsosabb, ha az értéke pozitív (a tipikushoz hasonló több értéket tartalmaz).

1. Gyakorlati példa

Vizsgáljuk meg, hogy mely termékkategóriákat vásárolták legutóbb online a kitöltők (Q9).

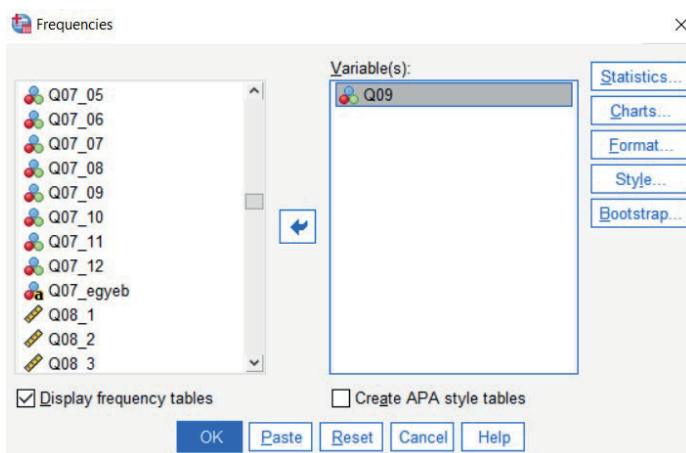
Ez a kérdés egy egyválasztós kérdés, ahol összesen ötféle termékkategóriából választhattak a kitöltők. A kérdésre adott válasz ezért egy nominális változóba rendeződik az adatbázisban, melyet gyakorisági táblával elemzünk.

Elérés útvonala: Analyze / Descriptives / Frequencies

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Frequencies' dialog box is open, and the 'Display' section is expanded to show 'Bar Charts' and 'Residuals'. The 'Display' section is also expanded to show 'Bar Charts' and 'Residuals'. The 'Display' section is also expanded to show 'Bar Charts' and 'Residuals'.

Name	Type	Width	Label	Values	Missing
59	Q06_13	Numeric	40		
60	Q06_1	Numeric	40		
61	Q06_2	Numeric	40		
62	Q06_3	Numeric	40		
63	Q06_4	Numeric	40		
64	Q06_5	Numeric	40		
65	Q07_01	Numeric	40		
66	Q07_02	Numeric	40		
67	Q07_03	Numeric	40		
68	Q07_04	Numeric	40		
69	Q07_05	Numeric	40		
70	Q07_06	Numeric	40		
71	Q07_07	Numeric	40		
72	Q07_08	Numeric	40		
73	Q07_09	Numeric	40		
74	Q07_10	Numeric	40		
75	Q07_11	Numeric	40		
76	Q07_12	Numeric	40		
77	Q07_égyéb	String	2000		
78	Q08_1	Numeric	40		
79	Q08_2	Numeric	40		
80	Q08_3	Numeric	40		
81	Q08_4	Numeric	40		
82	Q08_5	Numeric	40		
83	Q09	Numeric	40		
84	Q10_1	Numeric	40		
85	Q10_2	Numeric	40		
86	Q10_3	Numeric	40		
87	Q10_4	Numeric	40		
88	Q10_5	Numeric	40		
89	Q10_6	Numeric	40		
90	Q10_7	Numeric	40		

A párbeszédablakban számos beállítási lehetőségünk van, de itt csak a változó lista alatti „Display frequency tables” opciót választjuk.



Értelmezés

A kérdést összesen 255 fő válaszolta meg, míg 135 válaszadó nem töltötte ki (*missing*). Az egyes kategóriák jelölésének számosságát, gyakoriságát a „Frequency” oszlopban láthatjuk. A legutóbbi alkalommal a kitöltők közül 126 fő vásárolt valamilyen ruházati terméket. Ők a kutatásban részt vevő összes válaszadó 32,3%-át jelentik (*Percent*), ez azonban majdnem a fele (49,4%) azok közül, akik megválasztották ezt a kérdést (*Valid Percent*). A *Percent* tehát az összes résztvevőhöz képest mutatja a százalékos megoszlást, azaz a relatív gyakoriságot, míg a *Valid Percent* csak azokhoz viszonyít, akik az adott kérdésre választ adtak.

Az alábbi termékkategóriák közül, melyiket vásároltad online a legutóbbi alkalommal?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)	126	32,3	49,4	49,4
	Kozmetikum	27	6,9	10,6	60,0
	Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book)	29	7,4	11,4	71,4
	Ruházati kiegészítők (cipő, táska, egyéb kiegészítők)	48	12,3	18,8	90,2
	Könyv	25	6,4	9,8	100,0
	Total	255	65,4	100,0	
Missing	System	135	34,6		
Total		390	100,0		

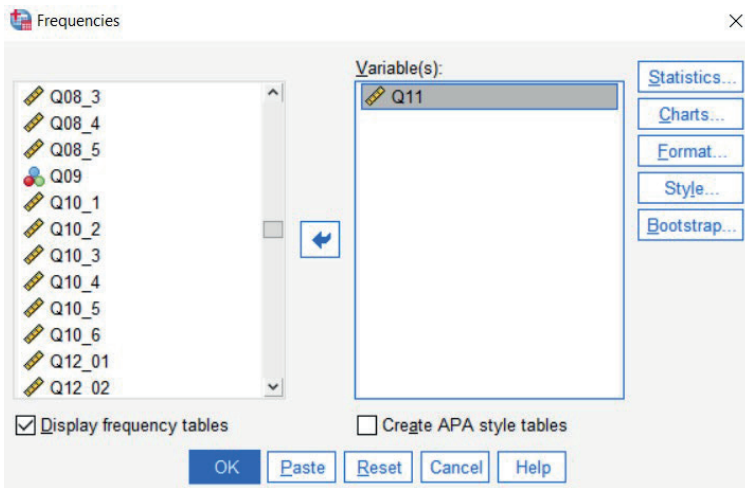
2. Gyakorlati példa

Vizsgáljuk meg, hogy legfeljebb hány percet hajlandóak utazni a csomagátvevő pont-hoz a kitöltők (Q11).

A változó percben rögzíti a kitöltők utazási hajlandóságát, ezért arányskálájú változónak tekinthető. A vizsgálathoz leíró statisztikai mutatókat és a vizuális ábrázoláshoz hisztogramot alkalmazunk.

Elérés útvonala: Analyze / Descriptives / Frequencies (ld. előző feladat).

A beállítások során tegyük át a Variables-be a vizsgálni kívánt Q11 változót.



Lekért adatok a menüpontokból:

Statistics:

- ✓ Percentile Value:
 - Quartiles*: kvartilisek, Q1, Q2, Q3
- ✓ Dispersion:
 - Std. deviation*: szórás
 - Minimum*: minimum
 - Maximum*: maximum
 - Range*: terjedelem
- ✓ Central Tendency
 - Mean*: átlag
 - Median*: medián
 - Mode*: módusz

✓ Distribution

Skewness: ferdeség

Kurtosis: csúcsosság

Frequencies: Statistics

Percentile Values

- Quartiles
- Cut points for: 10 equal groups
- Percentile(s):

Add
Change
Remove

Central Tendency

- Mean
- Median
- Mode
- Sum

Values are group midpoints

Dispersion

- Std. deviation
- Variance
- Range
- Minimum
- Maximum
- S.E. mean

Distribution

- Skewness
- Kurtosis

Continue Cancel Help

Charts:

- ✓ *Chart Type – Histograms* és *Show normal curve on histogram*. Ennek a segítségével lesz látható, hogy eltolódik-e az eloszlás a normális eloszláshoz képest.

Frequencies: Charts

Chart Type

- None
- Bar charts
- Pie charts
- Histograms:
 - Show normal curve on histogram

Chart Values

- Frequencies
- Percentages

Continue Cancel Help

Értelmezés

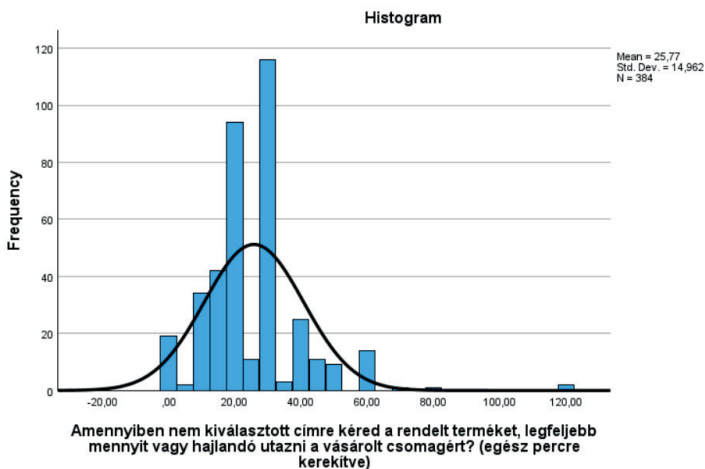
384 fő válaszolta meg a kérdést (*Valid*), 6 fő esetében nem kaptunk választ (*Missing*). Az átlagos utazási hajlandóság 25,77 perc (*Mean*), amitől átlagosan 14,96 perccel térnek el a válaszadók (*Std. Deviation*). A leggyakrabban 30 perccel jelöltek vállalható utazási időként a kitöltők (*Mode*). A kitöltők negyede 15 percnél kevesebb (*Q1*), fele 25 percnél kevesebb (*Q2, Median*), míg háromnegyede 30 percnél kevesebb (*Q3*) utazási időt tart vállalhatónak. A legrövidebb utazási idő 0 perc (*Minimum*), a leghosszabb pedig 120 perc (*Maximum*). Az eloszlás csúcsosabb a normális eloszlásnál (pozitív a *Kurtosis* mutató), valamint balra dől (negatív a *Skewness* mutató), azaz a rövidebb utazási idők dominálják a mintát.

Statistics

Amennyiben nem kiválasztott címre kéred a rendelt terméket, legfeljebb mennyit vagy hajlandó utazni a vásárolt csomagért? (egész percre kerekítve)

N	Valid	384
	Missing	6
Mean		25,7656
Median		25,0000
Mode		30,00
Std. Deviation		14,96217
Variance		223,867
Skewness		1,743
Std. Error of Skewness		,125
Kurtosis		7,912
Std. Error of Kurtosis		,248
Range		120,00
Minimum		,00
Maximum		120,00
Percentiles	25	15,0000
	50	25,0000
	75	30,0000

Vizuálisan a hisztogram segít érteni az eloszlás természetét, ahol a fekete görbe jeleníti meg a normális eloszlást.



04 HIPOTÉZISVIZSGÁLAT

Gyakran felmerül a kérdés, hogy miért van szükség hipotézisvizsgálatokra, amikor a mintában már látszanak bizonyos összefüggések, például a mintabeli nők átlagfizetése magasabb, mint a mintabeli férfiak átlagfizetése. Akkor miért is ne vonjuk le azt a következtetést, hogy a nők jobban keresnek a férfiaknál?

Ennek a kérdésnek a megválaszolásához meg kell értenünk a populáció és a mintavétel logikáját. Tegyük fel, hogy van egy üzleti kérdésünk, például szeretnénk megérteni, hogy van-e különbség a vegetáriánusok és a vegánok között abban, hogy milyen mértékig fontos számukra a kozmetikai cikkek kiválasztása során, hogy az bizonyítottan állatkísérletektől mentes legyen, vagy azonos mértékben tartja ezt fontosnak ez a két csoport. Kozmetikai termék gyártóként az összes vegetáriánus és vegán véleménye foglalkoztat minket, azonban jóformán lehetetlenség lenne minden ilyen étrendet követő fogyasztó szokására egyesével rákérdezni. Ezen két csoport tagjai jelentik a kutatásunk szempontjából a populációt, vagy más néven teljes sokaságot. Mit lehet tennünk egy ilyen esetben? Egy mintát veszünk a két csoport képviselői közül, és az ő szokásaikra kérdezzük rá. Ezen a módon elérjük a sokaság egy megfelelően kiválasztott részét, de valójában nem csak az ő (minta) szokásaik érdekelnek minket, hanem arra vagyunk kíváncsiak, hogy az ő válaszaik alapján milyen következtetést vonhatunk le az összes vegánra és vegetáriánusra (sokaság) az állatkísérletek megtétele kapcsán. Ez a példa körvonalazza a statisztikai indukció lényegét, amikor konkrét megfigyelésekből vagy adatokból (mintákból) általános következtetéseket vonunk le a teljes populációra (sokaságra) vonatkozóan.

Jelen könyvnek nem célja a mintavételi lehetőségeket részletezni, azonban annyit mindenképpen meg kell jegyezni, hogy az indukció megbízhatóságát (azaz, hogy helyes következtetést vonunk le a mintából a sokaságra) javítja, (1) ha növeljük a minta elemszámát, (2) véletlenszerű mintát veszünk, (3) reprezentatív mintát veszünk, azaz bizonyos sokaság szerinti jellemzőket (pl: demográfia megoszlást) a mintában is követünk.

Mivel mintáról van szó, ezért szem előtt kell tartanunk, hogy mindig van esélye annak, hogy a vett minta torz a valósághoz képest. Ezért soha nem lehetünk 100%-ig biztosak abban, hogy amit a mintában látunk, az megfelel a valóságnak (sokaságnak). A hipotézisvizsgálatok során ezt a bizonytalanságot „kezeljük”, azaz annak az esélyét latolgatjuk, hogy amit a mintában látunk, az milyen valószínűség mellett lehet igaz a sokaság tekintetében. Tehát a mintában látottak milyen valószínűséggel általánosíthatóak a sokaság esetében.

A statisztikai hipotézis tehát egy olyan állítás vagy feltételezés, amelyet egy sokaságról teszünk, és amelyet statisztikai módszerekkel ellenőrzünk adott minta alapján. A hipotézis lehet egy paraméter feltételezett értékének vizsgálata, egy kapcsolat két (vagy több) változó között, vagy egy változó hatásának vizsgálata egy másik változóra. A statisztikai hipotézisrendszer két hipotézisből áll, melyek egymást kizárják, és együtt a teljes eseményteret lefedik (azaz valamelyik mindenképpen bekövetkezik, de csak az egyik következhet be).

A **nullhipotézis (H₀)** mindig matematikai egyezőséget tartalmaz, ami jellemzően tartalmát tekintve azt jelenti, hogy nincs kapcsolat két változó között, vagy nincs eltérés két vizsgált paraméter között, illeszkedik egymásra két eloszlást. Az ún. **alternatív, H₁ hipotézis** pedig épp ennek az ellenkezőjét állítja. Ebből fakad, hogy az esetek többségében elvetni kívánjuk a H₀ hipotézist és a H₁-et igazolni, hiszen itt van az eltérés, itt van a kapcsolat megfogalmazva. Ettől eltérő esetek is előfordulnak ugyanakkor a statisztikában, amikor H₀ hipotézis tartalmazza a számunkra kedvező, bizonyítandó dolgot. Ilyen eset például a szórás-homogenitás igazolására vonatkozó Levene próba, ahol a H₀ tartalmazza a csoportok szórásának azonosságát (azaz itt van megfogalmazva a variancia-homogenitás). A hipotézist a kutatónak kell felállítania a statisztikai módszertan szabályai szerint, hiszen az SPSS ezeket nem fogalmazhatja meg a kutató helyett. Ezért, hogy olvasóink dolgát megkönnyítsük, minden bemutatott hipotézisvizsgálati módszer esetében megfogalmazzuk a H₀ hipotézist.

A nullhipotézis és az alternatív hipotézis közötti választás alapját a próba-függvény vagy más néven próbastatisztika kiszámítása (illetve SPSS általi kiszámíttatása) adja az adataink alapján. Ez a számérték segít meghatározni, hogy a megfigyelt adataink mennyire valószínűek a nullhipotézis feltételezése mellett. A próbastatisztikát mindig egy adott eloszláson értelmezzük. A statisztikai eljárások során leggyakrabban alkalmazott eloszlások a normális eloszlás (Z), a Student-féle t-eloszlás, az F-eloszlás és a khi-négyzet eloszlás. Az adott eloszlás típusa a vizsgálat típusától és a minta jellemzőitől függ. A statisztika értéket az SPSS output táblák jellemzően az adott eloszlás betűjelével, vagy a „statistic” kifejezéssel jelölik.

Ehhez a próbastatisztikához az adott eloszláson tartozik egy valószínűség, amelyet p értéknek nevezünk. Ez az a valószínűség, ahol a H₀ hipotézist éppen már elvetjük és a H₁ hipotézist fogadjuk el. Minél kisebb a p érték, annál erősebb bizonyítékot szolgáltat az alternatív hipotézis javára. Ha a p érték kisebb, mint az előre meghatározott szignifikanciaszint (ezt jelöljük alával, például 0,05), akkor a nullhipotézist elutasítjuk és az alternatív hipotézist elfogadjuk. Általában 5%-os (0,05) szignifikanciaszintet használunk, ami azt jelenti, hogy 95%-os biztonsággal döntünk a hipotézisekről.

Jelen fejezetben a hipotézisvizsgálat alapvető gyakorlati megközelítésére koncentráltunk. Annak mélyebb elméleti hátterét a Hunyadi, Mundruczó, Vita (1996): Statisztika c. tankönyv részletesen tartalmazza.

A vizsgálati módszer megválasztása

A megfelelő vizsgálati módszer megválasztása során figyelembe kell venni a kutatási kérdést, a változók típusát és a vizsgálat célját. Az alábbiakban egy összefoglaló táblázatot és egy leírást készítettünk, iránytűt adva ezzel a megfelelő statisztikai eljárás megválasztásához.

4.1. táblázat: Segédlet a megfelelő statisztikai módszer megválasztásához

	Vizsgálat jellege	Input változók		Output változók
Gyakorisági eloszlások	leíró	nem metrikus		–
Leíró statisztikai mutatók	leíró	metrikus és nem metrikus		–
Kereszt tábla	kapcsolat	nem metrikus		–
Varianciaelemzés	ok-okozat	független: nem metrikus	függő: metrikus	–
Korreláció – Pearson-féle	kapcsolat	mindkettő metrikus		
Korreláció – Spearman-féle	kapcsolat	mindkettő ordinális, vagy az egyik ordinális a másik metrikus		–
Regresszió	ok-okozat	független metrikus	függő metrikus	–
Faktorelemzés	struktúra feltáró	metrikus		metrikus (standardizált)
Klaszterelemzés	struktúra feltáró	nem metrikus és metrikus		nem metrikus (kategorizált)

A vizsgálati módszerek statisztikai szempontból lehetnek egy- két és többváltozós statisztikai elemzések. Az egyváltozós vagy leíró statisztika a gyakorisági vizsgálat a hozzá tartozó statisztikai mutatókkal. A kétváltozós vizsgálatok két metrikus vagy nem metrikus változó közötti viszonyt vizsgálnak (mindegyik módszer lehet többváltozós is). Leggyakrabban a közöttük lévő kapcsolat

létét vizsgáljuk, amihez statisztikai próbákat használunk. A statisztikai próbák hipotézisvizsgálat formájában lehetővé teszik annak a vizsgálatát, hogy két (vagy több) változó között a kapcsolat vagy hatás csak esetleges vagy statisztikailag is szignifikáns-e.

A marketingkutatókban az összefüggések vizsgálatára három fő módszert szokás használni, amelyek mind a változók közötti kapcsolatra, mind a mérési szintjükre eltérő követelményeket írnak elő:

- A keresztábra két (vagy több) változó között úgy vizsgálja a kapcsolatot, hogy nem feltétlenül kell feltételezni oksági összefüggést és mindkét változó lehet nominális vagy ordinális (nem metrikus, kategorizált változó). A változók közötti statisztikai összefüggés létét Khi-négyzet próbával ellenőrizzük.
- A variancia elemzésnél meg kell fogalmaznunk a változók közötti oksági kapcsolatot, a függő változónak metrikusnak kell lennie, a függetlennek pedig nominálisnak vagy ordinálisnak, a varianciaelemzés egy vagy több független és függő változó közötti összefüggést mér.
- A korreláció- és regresszióelemzésnél a két vagy több metrikus változó közötti kapcsolatot mérjük. A regresszióelemzésnél meg kell fogalmaznunk a változók közötti oksági kapcsolatot és jellemzően metrikus változókat kell használnunk. A regresszióelemzésnél a változók közötti összefüggést függvényel írjuk le, a kapcsolat erősségét, intenzitását a korrelációs együtthatóval mérjük.

A három összefüggésvizsgáló módszer közül a legerősebb kritériumot a változók mérési szintjére a regressziós elemzés fogalmazza meg, a legenyhébb a keresztábra és Khi-négyzet próba. Amennyiben például két metrikus változónk van, akkor legcélszerűbb a regressziós elemzést alkalmazni, de a változók kategorizálásával (RECODE menü segítségével) alkalmazhatjuk a variancia vagy a Khi-négyzet próbás elemzést is, ez azonban valamennyi információ veszteséssel járhat.

05 KERESZTTÁBLA ELEMZÉS

A kereszttábla elemzés a legegyszerűbbnek tűnő kapcsolatvizsgálati módszer a statisztikában, pontos alkalmazása és értelmezése azonban körültekintést igényel.

Elméleti bevezetés

Az elemzés során két kategorizált változó kapcsolatát vizsgáljuk, melyek lehetnek akár nominális-nominális, nominális-ordinális vagy ordinális-ordinális felállásban. A kereszttáblás vizsgálat statisztikai értelemben nem vizsgál ok-okozati összefüggést a két változó között. Gyakran emlegetik a vizsgálatot függetlenségvizsgálat néven is, mivel két kategorizált változó összefüggését kívánjuk ellenőrizni.

Alább egy példa (5.1. táblázat) látható a kereszttáblák, vagy kontingencia táblák általános formájára. F_{1A} azon elemek számosságát jelzik, akik az „egyik változó” első kategóriájába tartoznak, miközben a „másik változó” A kategóriájába is tartoznak, ezek az ún. keresztyakoriságok. Az n_1 azokat számosságát, elemszámát mutatja, aki az „egyik változó” első kategóriájának tagjai, ezeket hívjuk peremgyakoriságoknak. A táblázat jobb alsó csücskében az n jelölés a teljes mintaelemszámot mutatja.

5.1. táblázat: A kereszttábla általános sémája

		Másik változó			total
		A kategória	B kategória	C kategória	
Egyik változó	1. kategória	f_{1A}	f_{1B}	f_{1C}	n_1
	2. kategória	f_{2A}	f_{2B}	f_{2C}	n_2
total		n_a	n_b	n_c	n

Az elemzés során tulajdonképpen három főbb kérdésre keressük a választ:

1. Szignifikáns kapcsolat van-e a két változó között?

Amennyiben nincs szignifikáns kapcsolat, akkor az elemzés ezen tény riportolásával be is fejeződik. Ha szignifikáns a kapcsolat akkor további két megválaszolendő kérdés merül fel:

2. Milyen szoros a kapcsolat?
3. Mit tudunk elmondani a kapcsolat természetéről?

A kapcsolat szignifikanciáját a Khi-négyzet próba segítségével vizsgálhatjuk meg. A függetlenség feltételezése szerint, ha nincs kapcsolat a két vizsgált változó között, akkor az egyik változó kategóriáinak eloszlása nem befolyásolja a másik változó kategóriáinak eloszlását. Ezt a feltételezést a nullhipotézis fogalmazza meg, amely kimondja, hogy a két változó független egymástól.

A kereszttáblás vizsgálat hipotézisei:

H₀: nincs szignifikáns kapcsolat a két változó között, a két változó független egymástól

H₁: szignifikáns kapcsolat van a két változó között, a két változó nem független egymástól

Matematikailag, ha a két változó független egymástól, akkor a kereszttáblában az egyes cellákban várható gyakoriságok (függetlenség mellett feltételezett gyakoriságok) a peremgyakoriságokból számíthatók ki a következőképpen:

$$f_{ij}^* = \frac{f_i^* \cdot f_j}{n} \text{ ahol:}$$

f_{ij}^* az i-edik sor és j-edik oszlop függetlenség mellett várható gyakorisága

f_i^* az i-edik sor peremgyakorisága,

f_j a j-edik oszlop peremgyakorisága

n a teljes mintaelemszám

A kereszttáblás elemzés alapkérdése, hogy a megfigyelt gyakoriságok (valós adatok) mennyire térnek el a várható gyakoriságoktól (függetlenség esetén feltételezett, azaz számított adatok). Az eltérések mértékét a Khi-négyzet próba segítségével vizsgáljuk. Minél nagyobb a statisztika értéke, annál nagyobb az eltérés, és annál valószínűbb, hogy a két változó között szignifikáns kapcsolat van.

A Khi-négyzet statisztika a következőképpen számítható:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - f_{ij}^*)^2}{f_{ij}^*} \text{ ahol:}$$

f_{ij} a megfigyelt (valós) gyakoriság az i-edik sor és j-edik oszlop cellájában,

f_{ij}^* a függetlenség mellett várható (számított) gyakoriság az i-edik sor és j-edik oszlop cellájában,

r a sorok száma

c az oszlopok száma

Fontos feltétel, hogy a Khi-négyzet próba és a keresztábla csak akkor értelmezhető megbízhatóan, ha a cellákban lévő elvárt gyakoriságok (azaz az f_{ij}^* értékek) legfeljebb 20%-a kisebb 5-nél. Azaz azokban az esetekben, ha sok a (szinte) üres kategória, akkor ez az elemzés nem alkalmazható. Ilyen eseteken például a kategóriák logikus összevonása segíthet.

Összeségében tehát elmondható, hogy amennyiben a Khi-négyzet próba p értéke 0,05-nél kisebb, akkor a H_1 hipotézis lép életbe, azaz szignifikáns kapcsolat van a két változó között.

Az SPSS többféle *szorossági mutatót* is számít, amelyek segítenek meghatározni a két kategorizált változó közötti kapcsolat szorosságát. Ezek a mutatók különböző esetekben használhatók attól függően, hogy milyen típusú adatokat elemzünk.

Phi (ϕ) mutató akkor használható, ha a mindkét változó két kategóriát tartalmaz (pl: nem – Férfi / Nő és Egyetértés – Igen/Nem), azaz egy 2×2 -es vizsgálatról van szó. A Phi értéke -1 és 1 között lehet, ahol 0 azt jelenti, hogy nincs kapcsolat a változók között, és a szélső értékek (1 vagy -1) nagyon erős kapcsolatot jeleznek, ezt abszolút értékében értelmezzük.

Cramer-féle V (Cramer's V) bármennyi kategória esetén alkalmazható, tehát nem korlátozódik a 2×2 -es elemzésekre. A Cramer-féle V értéke 0 és 1 közé esik, ahol 0 azt jelenti, hogy nincs kapcsolat, és 1 azt jelenti, hogy függvényszerű kapcsolat van a változók között, azaz az egyik változó értékének ismeretében pontosan tudjuk, hogy a másik változó melyik kategóriát jelenti (pl: aki kötelező sorkatonaságot teljesített, az biztosan férfi volt).

A kapcsolat természetét a feltételes megoszlások elemzésével lehet feltárni, azaz megérteni azt, hogy aki az egyik változó szerint valamely kategóriába tartozik, az a másik változó szerint nagyobb eséllyel tartozik-e egy bizonyos kategóriába. Ennek elemzésére a százalékos megoszlások és a csoportosított oszlopdiagram vannak a segítségünkre.

1. Gyakorlati példa

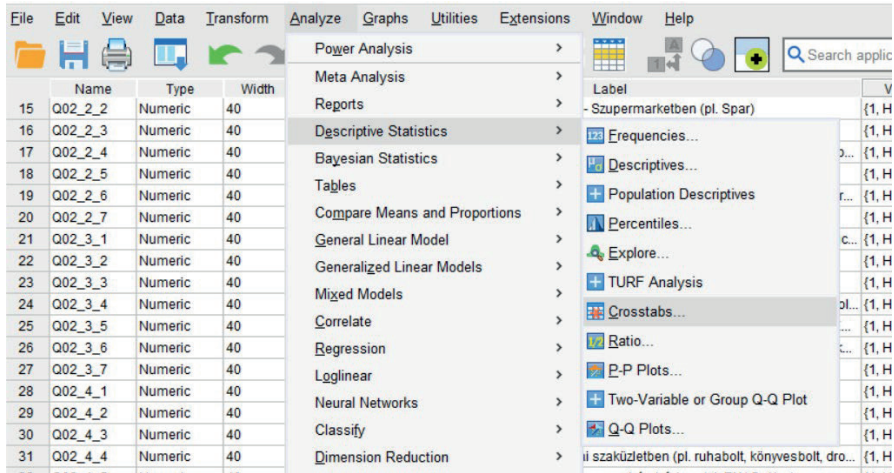
Van-e kapcsolat a kitöltők neme ($D1$ kategóriák: férfi / nő) és annak a ténye között, hogy az illető vásárol-e az Amazonon keresztül ($Q07_06$ kategóriák: vásárol Amazonon / nem vásárol).

Mivel mindkét változó nominális változó, azaz kategorizált változó, ezért a közöttük lévő kapcsolatot keresztáblás vizsgálattal elemezzük.

H_0 : Az Amazon használata független a nemtől

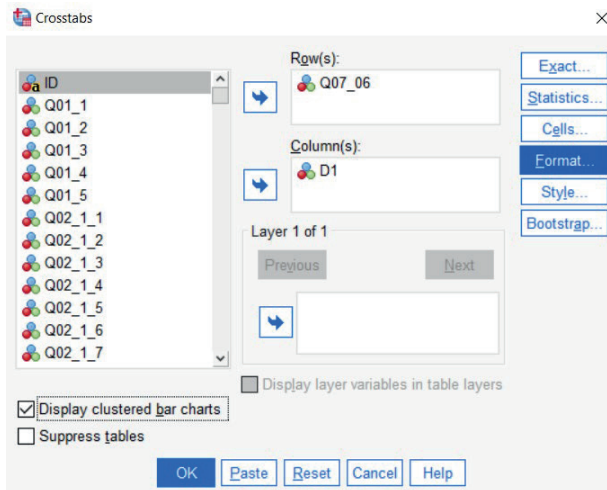
H_1 : Az Amazon használata és a nem összefüggést mutat

Elérési út: Analyze/Descriptive Statistics / Crosstabs



A beállítások során az egyik változót a sorokba (*Rows*), a másik változót pedig az oszlopokba (*Columns*) tesszük. Mivel a keresztábra elemzés során a kapcsolat irányáról nem mondunk semmit, ezért a változók sorokban (függő) és oszlopokban (független) való elhelyezése a kutató ízlésére van bízva. Érdeemes azonban végig gondolni, hogy vajon melyik változó lehet a független és függő változó. Itt most úgy történik meg a beállítás, hogy a független változó a nem, azaz azt feltételezzük, hogy a nem lesz hatással az Amazon használatára.

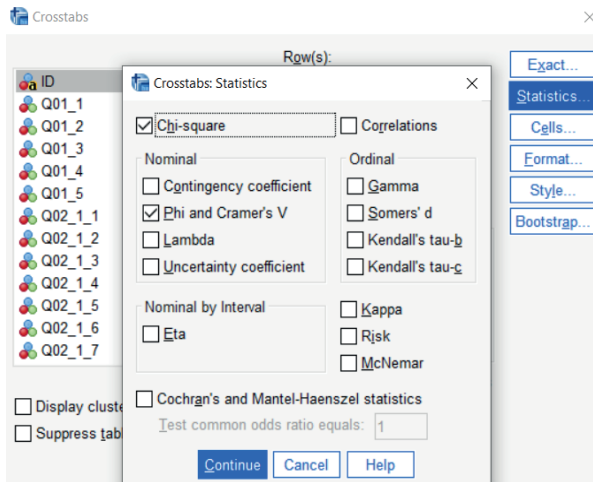
Ugyanitt állítjuk be a vizuális értelmezést segítő „*Display clustered bar charts*” lehetőséget.



Lekért adatok a menüpontokból:

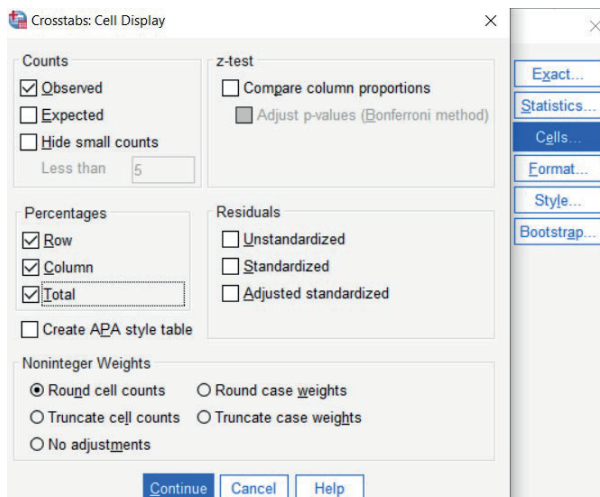
Statistics:

- ✓ *Chi-Square* → hipotézisvizsgálat Khi-négyzet próbáját számítja
- ✓ *Phi and Cramer V* → kapcsolat szorossági mutatószámokat számolja



Cells:

- ✓ *Percetanges (Row, Column, Total)* → relatív (százalékos) gyakoriságok. Amennyiben van logikailag azonosítható független változó, akkor az annak megfelelő százalékot szokás lekérni, de a másik változó szerinti százalékolás is értelmezhető. Itt most mind a három megoldást megnézzük.



Értelmezés

Első lépésként meg kell állapítanunk, hogy összefüggést mutat-e egymással az Amazonon való vásárlás és a nem. Ezt a Khi-négyzet próba segítségével állapíthatjuk meg. Mivel a Pearson-féle khi-négyzet értékhez tartozó szignifikancia-szint kisebb, mint 0,05 ($p < 0,001$), ezért el kell vessük a H_0 hipotézist és a H_1 hipotézist fogadjuk el, azaz megállapítható, hogy az Amazonon való vásárlás szignifikáns összefüggést mutat a nemmel.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	12,439 ^a	1	<,001		
Continuity Correction ^b	11,614	1	<,001		
Likelihood Ratio	12,682	1	<,001		
Fisher's Exact Test				<,001	<,001
Linear-by-Linear Association	12,407	1	<,001		
N of Valid Cases	390				

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 44,83.

b. Computed only for a 2x2 table

A cellákban lévő elemszám egyetlen esetben sem kisebb 5-nél, azaz 0%-ban (kisebb, mint 20%) fordul elő (szinte) üres cella. ezért a próba és a keresztábla megbízható eredményt ad.

Mivel az összefüggés igazolódott, ezért tovább elemezzük magát az összefüggést, a kapcsolat szorosságát. A Cramer-féle V mutató értéke 0,179, ami gyenge kapcsolatot jelez az Amazonon való vásárlás és a nem között.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-,179	<,001
	Cramer's V	,179	<,001
N of Valid Cases		390	

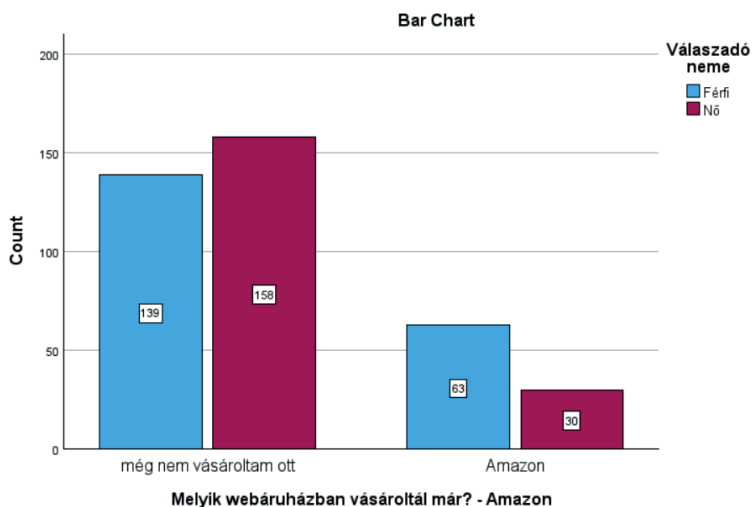
Az összefüggés természetét a keresztábla elemzésén keresztül érthetjük meg. Az összes, 390 kitöltő közül 297 fő még nem vásárolt az Amazonon keresztül, 93 fő már igen (peremgyakoróságok Amazon használat szerint). Az összes 390 kitöltő közül 202 férfi, míg 188 nő (peremgyakoróságok a nem szerint). A 202 férfi közül 63 fő (31,2%) vásárolt már az Amazon, míg a 188 nő közül 30 fő (16,0%) (keresztgyakoróságok). Azaz megállapítható, hogy a férfiak közel kétszer akkora arányban használják az Amazont, mint a nők.

Az összefüggés a másik változó irányából is érdekes. A 93 fő közül, akik az Amazonon vásárolnak 63 fő (67,7%) férfi, és 30 fő (32,3%) nő. Azaz az Amazon használóinak esetében a férfiak aránya több, mint kétszerese a nők arányának.

Melyik webáruházban vásároltál már? - Amazon * Válaszadó neme Crosstabulation

		Válaszadó neme			
		Férfi	Nő	Total	
Melyik webáruházban vásároltál már? - Amazon	még nem vásároltam ott	Count	139	158	297
		% within Melyik webáruházban vásároltál már? - Amazon	46,8%	53,2%	100,0%
		% within Válaszadó neme	68,8%	84,0%	76,2%
		% of Total	35,6%	40,5%	76,2%
Amazon		Count	63	30	93
		% within Melyik webáruházban vásároltál már? - Amazon	67,7%	32,3%	100,0%
		% within Válaszadó neme	31,2%	16,0%	23,8%
		% of Total	16,2%	7,7%	23,8%
Total		Count	202	188	390
		% within Melyik webáruházban vásároltál már? - Amazon	51,8%	48,2%	100,0%
		% within Válaszadó neme	100,0%	100,0%	100,0%
		% of Total	51,8%	48,2%	100,0%

Az utóbbi összefüggést szemlélteti az alábbi oszlopdiagram is (*clustered bar chart*).



2. Gyakorlati példa

Összefügg-e a kitöltő neve (D1) azzal, hogy melyik termékkategóriát vásárolta legutóbb online (Q09)?

Van-e kapcsolat a kitöltők neve (D1 kategóriák: férfi/nő) és annak a ténye között, hogy az illető melyik termékkategóriát vásárolta utoljára az interneten keresztül (Q09 kategóriák: ruházati termék, kozmetikum, szórakoztató elektronika, ruházati kiegészítők, könyv).

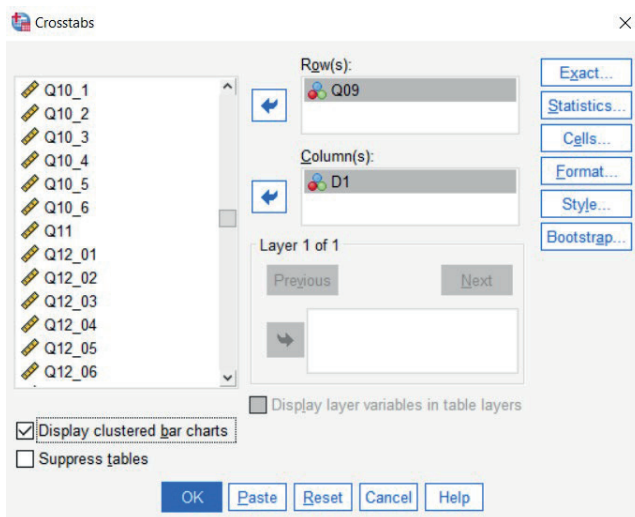
Mivel mindkét változó nominális, azaz nem metrikus változó, ezért a közöttük lévő kapcsolatot vizsgálatára a keresztábrás elemzés alkalmas.

Ho: Az utoljára online vásárolt termékkategória független a nemtől

H1: Az utoljára online vásárolt termékkategória és a nem összefüggést mutat

Elérési út: Analyze/Descriptive Statistics/Crosstabs

A sorok és oszlopok beállításánál a független változó a nem, azaz azt feltételezzük, hogy a nem lesz hatással az utoljára vásárolt termékkategóriára, tehát a nem (D1) kerül az oszlopokhoz, a termékkategória (Q9) pedig a sorhoz. Egyúttal beállítjuk a „Display clustered bar charts” opciót is.

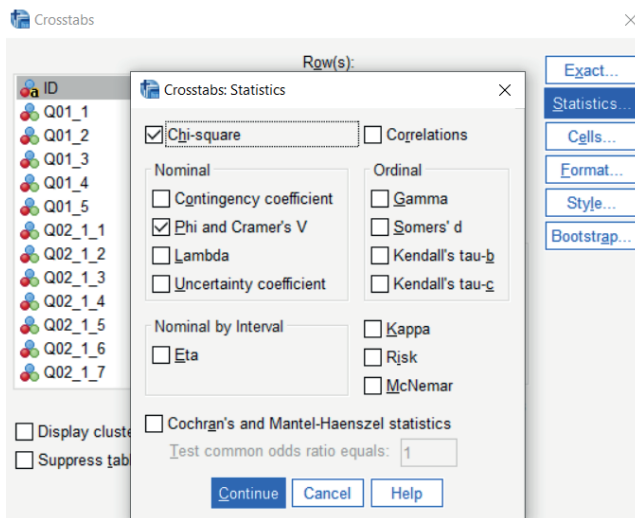


Lekért adatok:

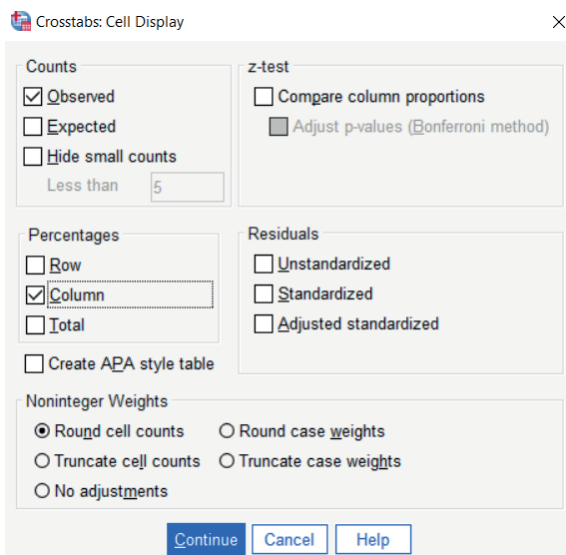
Statistics:

✓ Chi-Square → hipotézisvizsgálat Khi-négyzet próbáját számítja

✓ Phi and Cramer'V → kapcsolat szorossági mutatószámokat számolja



Cells: ✓ *Percetanges (Row, Column, Total)* → relatív (százalékos) gyakoriságok. Most a független változó szerint, tehát az oszlop alapján kérjük le a relatív gyakoriságokat.



Értelmezés

Első lépésként meg kell állapítanunk, hogy szignifikáns-e az összefüggés az utoljára vásárolt online termékkategória és a nem között. Ezt a Khi-négyzet próba segítségével dönthetjük el, amelyhez tartozó szignifikanciaszint kisebb, mint 0,05 ($p < 0,001$), ezért el kell vetnünk a H_0 hipotézist és a H_1 hipotézist fogadjuk el, azaz megállapítható, hogy az utoljára vásárolt termékkategória szignifikáns összefüggést mutat a nemmel.

A cellákban lévő elemszám egyetlen esetben sem 5-nél kisebb érték, azaz 0%-ban (kisebb, mint 20%) fordul elő (szinte) üres cella. Tehát a Khi-négyzet próba és a keresztábra megbízható eredményt ad.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	26,215 ^a	4	<,001
Likelihood Ratio	29,124	4	<,001
Linear-by-Linear Association	1,283	1	,257
N of Valid Cases	255		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 12.06.

Az összefüggés igazolódott, tehát annak természetét tovább elemezhetjük. Először a kapcsolat szorosságát vizsgáljuk meg a Cramer-féle V mutató segítségével, melynek értéke 0,321, tehát a két változó között a kapcsolat közepesnél valamivel gyengébbnek mondható.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	,321	<,001
	Cramer's V	,321	<,001
N of Valid Cases		255	

Az összefüggés tényleges jellegét a keresztábra elemzésén keresztül érthetjük meg. Az első, ami szembeötlő, hogy a ruházati termékkategóriát a nők 53%-a vásárolta online, míg a férfiaknak csupán 45,5%-a jelezte, hogy ez volt a legutóbb online vásárolt termékkategória. A kozmetikumok is nagyobb mértékben jellemzik a női vásárlókat (15,2%). A férfiak jelentősen a szórakoztató elektronika tekintetében jelölték nagyobb arányban (21,1%) hogy ezt a termékkategóriát vették legutoljára online, míg a nők esetében ugyanez az arány a legkisebb a nők esetében, csupán 2,3%-uk vásárolta a kategóriát legutóbb.

Az alábbi termékkategóriák közül, melyiket vásároltad online a legutóbbi alkalommal? *
Válaszó neme Crosstabulation

		Válaszó neme		Total	
		Férfi	Nő		
Az alábbi termékkategóriák közül, melyiket vásároltad online a legutóbbi alkalommal?	Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)	Count	56	70	126
		% within Válaszó neme	45,5%	53,0%	49,4%
	Kozmetikum	Count	7	20	27
		% within Válaszó neme	5,7%	15,2%	10,6%
	Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book)	Count	26	3	29
		% within Válaszó neme	21,1%	2,3%	11,4%
	Ruházati kiegészítők (cipő, táska, egyéb kiegészítők)	Count	23	25	48
		% within Válaszó neme	18,7%	18,9%	18,8%
	Könyv	Count	11	14	25
		% within Válaszó neme	8,9%	10,6%	9,8%
	Total	Count	123	132	255
		% within Válaszó neme	100,0%	100,0%	100,0%

Gyakorló feladat

Összefügg-e a kitöltő neme (D1) azzal, hogy figyeli-e a kozmetikumok esetében a termékösszetételt (Q18_o2)?

06 VARIANCIAELEMZÉS

Környezetünk megértése során felmerülhet a kérdés, hogy két vagy több különböző csoport viselkedése, attitűdje szignifikánsan eltér-e egymástól. Vajon különbözik-e az egyes vásárlói szegmensek újra vásárlási hajlandósága? Eltérő fogyasztói kosárértékkel rendelkeznek-e a különböző korcsoportok? Szignifikánsan különbözik-e az eltérő egyetemi szakokon tanulók hallgatói elégedettsége?

Ezeknek a kérdéseknek a megválaszolásában segít a varianciaelemzés. Emellett a kísérleti kutatások esetében is közkedvelt a módszer, amikor is a különböző kísérleti csoportok, scenariók átlagainak alakulását és eltérését kívánják megvizsgálni.

Elméleti bevezetés

A varianciaelemzés (ANalysis Of VAriance – ANOVA) célja megvizsgálni, hogy van-e különbség két vagy több csoport átlaga között. Az elemzés során megkülönböztetünk függő és független változókat. A független változó egy nem metrikus, kategorizáló változó, mely alapján az összehasonlítani kívánt csoportok egyértelműen azonosíthatóak. Ezeket szokás faktoroknak is nevezni. A függő változó mindig metrikus mérési szintű, mely alapján a különböző csoportok átlagai meghatározhatóak.

A varianciaelemzés során megkülönböztetünk egyszempontos (one-way ANOVA) és többszempontos (N-way ANOVA) varianciaelemzést. Az egyszempontos varianciaelemzés esetében egy függő metrikus és egy független nem metrikus változó szerepel az elemzésünkben, míg a többszempontos varianciaelemzésben kettő vagy több független nem metrikus változóval és egy metrikus függő változóval dolgozunk. A többszempontos varianciaelemzés használatakor a független változók közös, ún. interakciós hatása is értelmezhető. Erről a későbbiekben írunk részletesebben. Az ismételt méréses varianciaelemzéssel és az ANCOVA elemzéssel jelen könyvben nem foglalkozunk.

A varianciaelemzés tesztje az F-próba, melynek nullhipotézise a következő:

A varianciaanalízis (ANOVA) hipotézisei

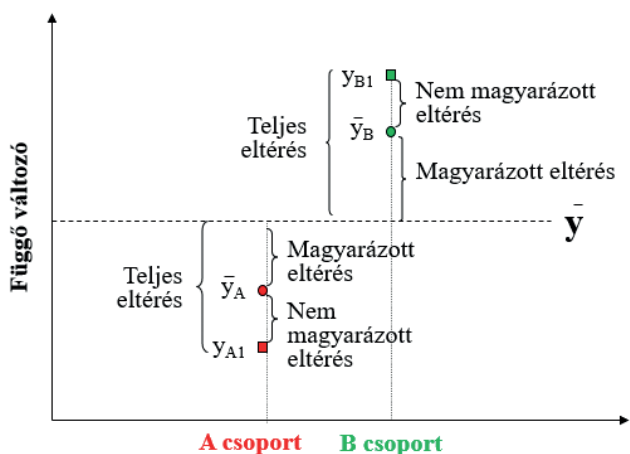
Ho: A csoportok átlaga nem tér el szignifikánsan egymástól.

H1: Legalább egy csoport átlaga szignifikánsan különbözik.

Amennyiben az F-próba eredménye alapján H_0 -t elvetjük, akkor állapítható meg szignifikáns különbség a független változó által meghatározott csoportok között. Az F-próba számításakor a csoportok közötti, vagyis a független változónak tulajdonítható átlagos eltérésnégyzetösszeg (MS_x) és a csoporton belüli, vagyis a hibának tulajdonítható átlagos eltérésnégyzetösszeg (MS_{hiba}) hányadosát vesszük figyelembe.

Egy egyéni érték (pl. y_{A1} , amely az A csoport első megfigyelésének konkrét érték jelöli) eltérése a teljes minta átlagától felosztható magyarázott és nem magyarázott eltérésre (6.1. ábra).

6.1. ábra: A varianciaelemzés alapjai



A magyarázott eltérés az adott csoport átlagának eltérését méri a teljes minta átlagától. Amennyiben ennek a különbségnek a négyzetét megszorozzuk az adott csoport elemszámával, akkor megkapjuk a külső eltérésnégyzetösszeget (SS_x vagy $SS_{külső}$).

$$SS_x = \sum_{j=1}^c n(\bar{Y}_j - \bar{Y})^2$$

A nem magyarázott eltérés az adott csoportba tartozó egyéni érték eltérését mutatja az adott csoport átlagától. A csoport minden egyes megfigyelésére szummázva ezt a négyzetes különbséget kapjuk meg a belső eltérésnégyzetösszeget ($SS_{belső}$ vagy SS_{hiba}).

$$SS_{hiba} = \sum_j \sum_i^n (Y_{ij} - \bar{Y}_j)^2$$

Az átlagos négyzetösszeg számítása során – mely az F próba értékének kiszámításához szükséges – a különböző eltérésnégyzetek összegét osztjuk el a megfelelő szabadságfokkal.

A külső eltérésnégyzetösszeg (SS_x vagy $SS_{\text{külső}}$) esetében a szabadságfok a kategóriaszám eggyel csökkentett értékével egyenlő, míg a belső eltérésnégyzet összeg ($SS_{\text{belső}}$ vagy SS_{hiba}) esetében a megfigyelések száma mínusz a kategóriaszám adja ezt az értéket.

$$F = \frac{\frac{SS_x}{(c-1)}}{\frac{SS_{\text{hiba}}}{(N-c)}} = \frac{MS_x}{MS_{\text{hiba}}}$$

Az SPSS program használatakor a kiszámított F-próba értéke mellett a belső, külső és teljes eltérés négyzetösszegeket, a szabadságfokokat, illetve az átlagos négyzetösszeg értékeket is leolvashatjuk majd az eredménytáblából.

Amennyiben szignifikáns különbség figyelhető meg a csoportok átlaga között, akkor annak mértékét is vizsgálni tudjuk az η^2 (η^2) mutatóval, mely 0 és 1 közötti értéket vehet fel. η^2 megmutatja, hogy a függő változó varianciájának hány százaléka magyarázott a független változó által.

A varianciaelemzés feltételei

A varianciaelemzés során az alábbi feltételek teljesülése szükséges:

1. A függő változó metrikus mérési szintű, a független változó nem metrikus vagy kategorizált legyen
2. A hibatag normál eloszlású legyen 0 átlaggal és konstans varianciával, vagyis a függő változó normális eloszlást kövessen. A normális eloszlás grafikus ábrázolással, Kolgomorov-Smirnov vagy Shapiro-Wilk teszttel, vagy hisztogrammal is ellenőrizhető. Nagyobb mintaelemszám esetén azonban ($N > 30$) a centrális határeloszlás törvény alapján a megsértése figyelmen kívül hagyható, és az F-próba eredménye megbízhatónak tekinthető.
3. Egyenlő variancia, vagyis szóráshomogenitás a csoportok között.

Teljesülése a Levene-teszttel ellenőrizhető, melynek nullhipotézise:

A Levene-teszt hipotézisei

H₀: A szórásnégyzetek egyenlők.

H₁: Legalább egy szórásnégyzet különbözik.

A szóráshomogenitás nem teljesülése az F-próba eredményét befolyásolja, így megsértése esetén a Welch-teszt használata javasolt inkább.

4. A megfigyelések, a válaszadók egymástól függetlenek, kizárólagosak, vagyis nincsenek ismételt mérések, válaszadók a mintában.
5. Hasonló cellagyakoriság az összehasonlítani kívánt csoportokban.

EGYSZEMPONTOS VARIANCIAELEMZÉS (ONE-WAY ANOVA)

Az egyszempontos varianciaelemzés során egy független/faktor változónk és egy függő metrikus változónk van, vagyis azt szeretnénk megvizsgálni, hogy a független változó által képzett csoportok átlagai vajon szignifikánsan különböznek-e egymástól. Amennyiben a független változónak két értéke van, pl. férfi / nő, akkor az F-próba helyett t-próbával is összehasonlíthatóak a csoportok. Kettőnél több csoport esetében a varianciaelemzés, vagyis az F-próba szükséges.

Gyakorlati példa

Nézzünk meg egy olyan példát, melyben a független változónak kettőnél több kimenete van.

A webáruház szeretné megvizsgálni, hogy a vásárolt termék jellege befolyásolja-e azt, hogy hajlandóak-e a vásárlók esetleges magasabb kiszállítási költséget fizetni azért, hogy a visszaküldés ingyenes legyen.

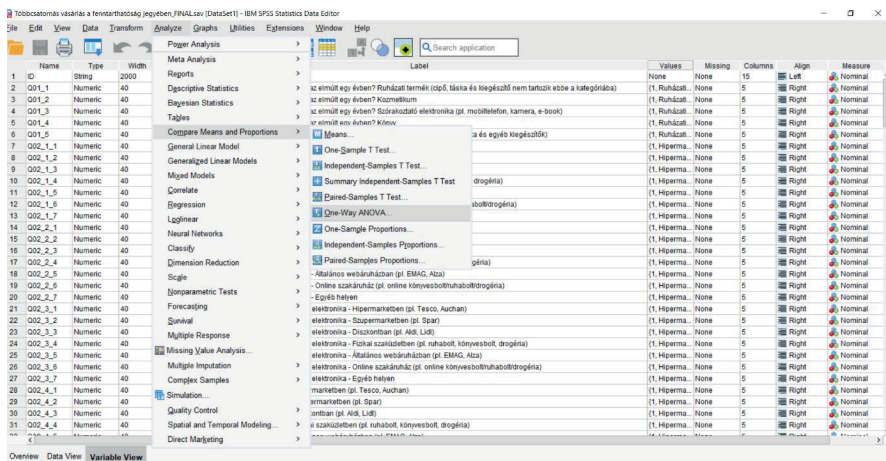
A kutatási kérdés megértéséhez a Q₀₉-es (*Az alábbi termékkategóriák közül, melyiket vásároltad online a legutóbbi alkalommal?*) valamint a Q_{13_3} (*Mennyire jellemző ONLINE vásárláskor (előzőleg jelölt termékkategória esetén)? – Hajlandó vagyok többet fizetni a termék kiszállításáért, ha cserébe fennáll az ingyenes visszaküldés lehetősége.*) változókat használjuk fel.

A független/faktor változónk a nem metrikus Q09-es nominális mérési szintű változó lesz, mely öt értéket vehet fel (1 – ruházati termék (kivéve cipő, táskák, kiegészítő), 2 – kozmetikum, 3 – szórakoztató elektronika, 4 – könyv, 5 – ruházati kiegészítő), tehát elemzésünk során öt különböző termék kategóriát tudunk összehasonítani a függő változó alapján.

A függő változónk a Q13_3-as változó, mely egy 1-től 7-ig Likert-skálán mért (1 – egyáltalán nem jellemző, 7 – teljes mértékben jellemző) metrikus változó. Habár a Likert-skálák ordinális skálák, mégis bizonyított, hogy legalább 5 érték használata esetében intervallum skálaként kezelhetőek, így marketingkutatók szempontból mi is intervallum skálának tekintjük őket.

Ezek alapján a kérdés, melyre az elemzésünk során választ kapunk, hogy vajon szignifikánsan különbözik-e a vizsgált termék kategóriák esetében az átlagos egyetértés azzal, hogy a megkérdezettek hajlandóak többet fizetni a termék kiszállításáért, amennyiben cserébe fennáll az ingyenes visszaküldés lehetősége, vagy a termék kategória ezt nem befolyásolja.

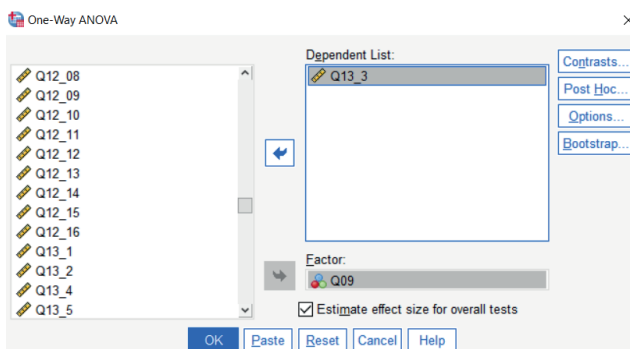
Elérési út: Analyze / Compare Means and Proportions / One-Way ANOVA



A felugró ablakban első lépésként a függő és független/faktor változókat kell meghatározni. A függő változó (*Dependent List*) a Q13_3 metrikus változó, míg a független változó (*Factor*) a Q09-as nem metrikus változó, mely az összehasonlítható csoportokat határozza meg.

Megjegyzés: A parancsot használva lehetséges egy független változó hatását több függőre vizsgálni. Ilyenkor a függő változókhöz több metrikus változó is definiálható. Az elemzés során minden egyes függő változóhoz tartozóan újabb és újabb, elkülönített egyváltozós varianciaelemzés kerül lefuttatásra.

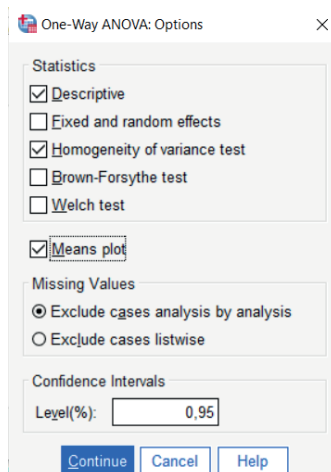
Az *Estimate effect size for overall test* parancs is legyen kiválasztva, ugyanis szignifikáns kapcsolat esetében ez adja meg számunkra a kapcsolat erősségét mérő mutatókat, többek között az η^2 mutatót is.



Lekért adatok a menüpontokból:

Options:

- ✓ *Descriptives* → független változó által képzett csoportokhoz, valamint a teljes mintához tartozó leíró statisztikai információkat tartalmazza (válaszadói számot, átlagot, szórást, minimumot, maximumot).
- ✓ *Homogeneity of variance test* → szóráshomogenitás tesztje, Levene-teszt
- ✓ *Means plot* → a különböző átlagok grafikus megjelenítése



Ezek után következik az adatok ellenőrzése. Első körben a feltételek teljesülését érdemes megvizsgálnunk.

- A hibatag normál eloszlású legyen 0 átlaggal és konstans varianciával. A leíró statisztikai mutatókat tartalmazó táblázatból (*Descriptives*) láthatjuk, hogy a mintaelemszám 30-nál nagyobb (N=255 fő), így a normál eloszlás megsértése a centrális határeloszlás törvény alapján figyelmen kívül hagyható, és az F-próba eredménye robusztusnak tekinthető. Ezek alapján a teszt elvégzésétől eltekinthetünk.

Descriptives

Mennyire jellemző ONLINE vásárláskor (előzőleg jelölt termékkategória esetén)? - Hajlandó vagyok többet fizetni a termék kiszállításáért, ha cserébe fennáll az ingyenes visszaküldés lehetősége.

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Ruházati termék (cipő, táská és kiegészítő nem tartozik ebbe a kategóriába)	126	4,45	1,695	,151	4,15	4,75	1	7
Kozmetikum	27	3,67	1,617	,311	3,03	4,31	1	7
Szórakoztató elektronika (pl. mobiltelefon, kamera, e-book)	29	4,17	1,583	,294	3,57	4,77	1	7
Ruházati kiegészítők (cipő, táská, egyéb kiegészítők)	48	4,06	1,731	,250	3,56	4,57	1	7
Könyv	25	3,52	1,686	,337	2,82	4,22	1	7
Total	255	4,17	1,700	,106	3,96	4,38	1	7

- Egyenlő variancia, vagyis szóráshomogenitás a csoportok között. A Levene-teszt értéke 0,277, a hozzá tartozó szignifikancia szint pedig 0,893. Ez alapján a H_0 hipotézist, vagyis, hogy a szóráshomogenitás fennáll, nem vetjük el, így az F-próba értéke megfelelő megbízhatósággal értelmezhető, nincs torzítás. Minden olyan esetben, amikor a Levene-teszthez tartozó szignifikancia szint 0,1-nél magasabb, a H_0 -t nem vetjük fel, vagyis a feltevél teljesül.

A szóráshomogenitás megsértése esetében *Welch-teszttel* lehet a csoportokat összehasonlítani, melynek kiszámítása az *Options* menüpontban kérhető le.

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
Mennyire jellemző ONLINE vásárláskor (előzőleg jelölt termékkategória esetén)? - Hajlandó vagyok többet fizetni a termék kiszállításáért, ha cserébe fennáll az ingyenes visszaküldés lehetősége.	Based on Mean	,277	4	250	,893
	Based on Median	,203	4	250	,936
	Based on Median and with adjusted df	,203	4	238,225	,936
	Based on trimmed mean	,266	4	250	,900

- A megfigyelések, a válaszadók egymástól függetlenek, vagyis nincsenek ismételt mérések, válaszadók a mintában feltétel teljesülését esetünkben az adatfelvétel módja biztosítja.
- Hasonló cellagyakoriság az összehasonlítani kívánt csoportokban.

A leíró statisztikai mutatókat tartalmazó táblázatban látható az egyes csoportok nagysága. A vizsgált mintában legtöbben a ruházati termék vásárlásához kötődően értékelték a kijelentéseket (N=126 fő), a többi csoport nagysága viszont nagyon hasonlóan alakul, így ez a feltétel is teljesül.

Az *F-próba* eredményét az *ANOVA táblázatban* találhatjuk. Ebből a táblázatból olvashatóak le a belső, külső, valamint a teljes eltérésnégyzetösszegek (*Sum of Squares / Between / Within / Total*), a szabadságfokok (*df*), illetve az átlagos eltérésnégyzetösszegek (*Mean Square*), melyek alapján kerül az *F-próba* értéke kiszámításra. Az *F-próba* értéke 2,478, a hozzá tartozó szignifikanciaszint pedig 0,045. Ez alapján az elemzésbe bevont csoportok szignifikánsan különböznek egymástól, vagyis megállapítható, hogy a termék típusa szignifikánsan befolyásolja azt átlagos egyetértés azzal, hogy a megkérdezettek hajlandóak-e többet fizetni a termék kiszállításáért, amennyiben cserébe fennáll az ingyenes visszaküldés lehetősége. Menedzseri szempontból ez azt jelenti, hogy a különböző termékkategóriák esetében eltérő árazás használható az ingyenes visszaküldés biztosítása érdekében, hiszen az elfogadási hajlandóság különbözik.

Az η^2 (η^2) mutató alapján láthatjuk, hogy ez azonban egy nagyon gyenge kapcsolatnak számít, mivel $\eta^2=0,038$, vagyis a függő változó varianciájának 0,38%-át magyarázza mindössze a független változó.

ANOVA

Mennyire jellemző ONLINE vásárláskor (előzőleg jelölt termékkategória esetén)? - Hajlandó va

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	28,003	4	7,001	2,478	,045
Within Groups	706,405	250	2,826		
Total	734,408	254			

ANOVA Effect Sizes^{a,b}

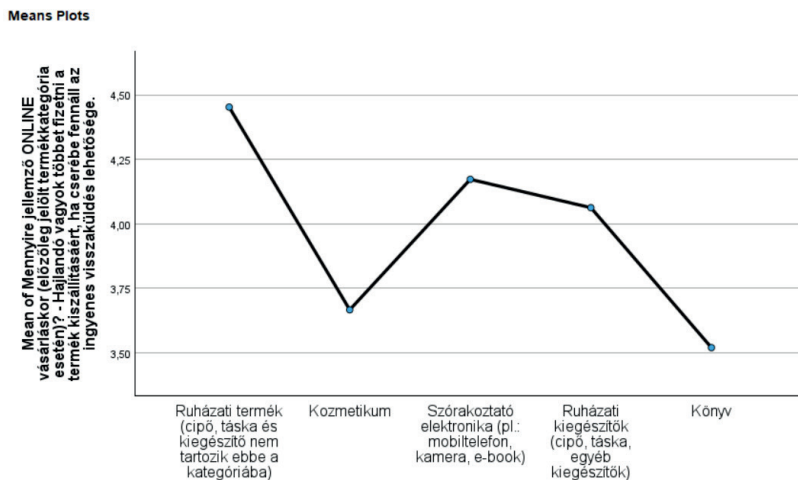
		Point Estimate	95% Confidence Interval	
			Lower	Upper
Mennyire jellemző ONLINE vásárláskor (előzőleg jelölt termékkategória esetén)? - Hajlandó vagyok többet fizetni a termék kiszállításáért, ha cserébe fennáll az ingyenes visszaküldés lehetősége.	Eta-squared	,038	,000	,081
	Epsilon-squared	,023	-,016	,066
	Omega-squared Fixed-effect	,023	-,016	,066
	Omega-squared Random-effect	,006	-,004	,017

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

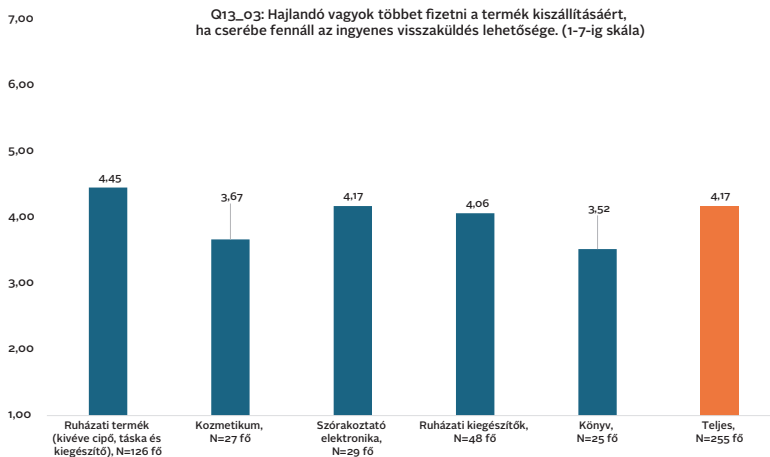
Az eredmények értékelése azonban itt még nem állhat meg. Fontos, hogy már látjuk, hogy szignifikáns különbség figyelhető meg az egyes elemzett termékkategóriák között, azonban látnunk kell azt is, és prezentálnunk is a kutatási riportban, hogy ezek a különbségek hogyan alakulnak. Ezt a leíró statisztikai adatok alapján (*Descriptives tábla*), valamint az átlagok grafikus megjelenítése alapján tudjuk megtenni, melyre az SPSS által elkészített *Means plot* (6.2. ábra), vagy bármilyen egyéni ábrázolás is megfelelő lehet (6.3. ábra).

6.2. ábra: A csoportok átlagainak alakulása (SPSS ábra)



Az alábbi termékkategóriák közül, melyiket vásároltad online a legutóbbi alkalommal?

6.3. ábra: A csoportok átlagainak alakulása (egyéni szerkesztés)



Az eredmények alapján megállapítható, hogy az összes válaszadó átlagosan 4,17-es (szórás=1,7) szinten ért egyet azzal, hogy online rendelés esetében hajlandó többet fizetni a kiszállítáért az ingyenes visszaküldésért cserébe, mely átlag a különböző termékkategóriák esetében szignifikánsan különbözik. Leginkább a ruházati termékek esetében jellemző ez a vélemény (átlag=4,45, szórás=1,69), míg legkevésbé a könyvek esetében (átlag=3,52, szórás=1,69). Ez az eredmény valószínűleg a termékkategóriában levő kockázatként is értelmezhető.

Gyakorló feladat

Eltérően jellemzi-e a férfiakat és nőket (D1), hogy lecseréltek már termékeket környezetvédelmi okból (Q18_o8)?

TÖBBSZEMPONTOS VARIANCIAELEZMÉS (N-WAY ANOVA)

A világ megértése során gyakran feltételezzük, hogy a viselkedést nem csak egy, hanem több független változó is együttesen befolyásolja. Vajon a vásárlók neme és jövedelmi státusza együttesen befolyásolják-e az elköltött pénz mennyiségét? A hallgatók neme és az érettségi minősítése együttesen befolyásolja-e az egyetemi tanulmányi átlagot?

Ezen kérdések esetében a független változók nem metrikus, csoportképző változók, a függő pedig metrikus, így a többszemponos varianciaelemzés módszerével válaszolhatóak meg.

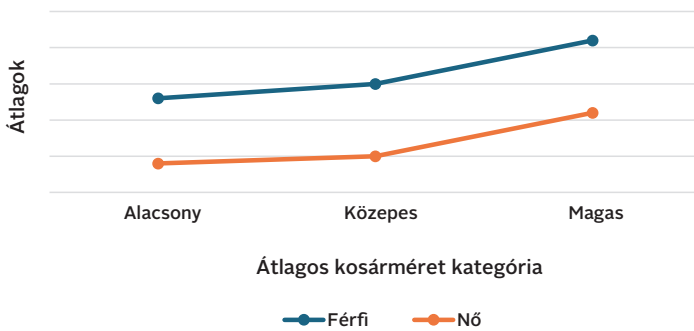
A többszemponos varianciaelemzés esetében legalább kettő független, nem metrikus változó hatását vizsgáljuk egy függő metrikus változóra. Ebben az esetben a teljes eltérésnégyzetösszeg a külső eltérésnégyzetösszegek és a hibának tulajdonított eltérésnégyzetösszeg mellett a független változók interakciójából származó eltérésnégyzetösszeggel is kiegészül.

$$SS_y = SS_{x_1} + SS_{x_2} + SS_{x_1x_2} + SS_{\text{hiba}}$$

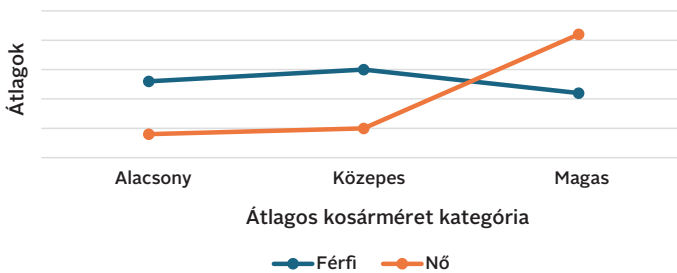
A többszemponos varianciaelemzés eredményeinek értékelése során először mindig a főhatás szignifikanciáját kell vizsgálni, mely megmutatja, hogy az elemzésbe bevont csoportok esetében létezik-e valamilyen szignifikáns különbség.

Amennyiben igen, akkor a következő lépésben az interakciós hatást kell górcső alá vetni. Az interakció megléte azt jelenti, hogy az egyik független változó hatása egy másik független változó hatásával együttesen érvényesül, vagyis a független faktor hatása a függőre függ egy másik faktor kategóriájától. Minél erősebb két független változó interakciója, annál nagyobb értéket vesz fel az interakciós eltérésnégyzet-összeg. A főhatások vizsgálatának, melyek a független változók hatását mutatják meg külön-külön, csak akkor van értelme, ha az interakciós hatás során a független változók erősítik egymás hatását (ordinális interakció – *ordinal interaction*, 6.4. ábra). Amennyiben létezik szignifikáns interakciós hatás, de a független változók átlagai ellentétes mozgást vagy ellentétes sorrendiséget mutatnak (nem ordinális interakció – *disordinal interaction*, 6.5. ábra és 6.6. ábra), akkor a főhatások nem értelmezhetőek. Ennek eldöntése legegyszerűbben grafikus ábrázolással történhet.

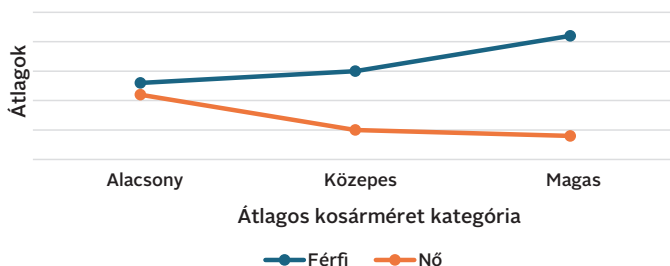
6.4. ábra: Ordinális interakció



6.5. ábra: Nem ordinális interakció keresztező „vonalakkal”



6.6. ábra: Nem ordinális interakció eltérő sorrend



Gyakorlati példa

A többszemponτος varianciaelemzés esetét az alábbi kutatási kérdés mentén nézzük meg részletesen.

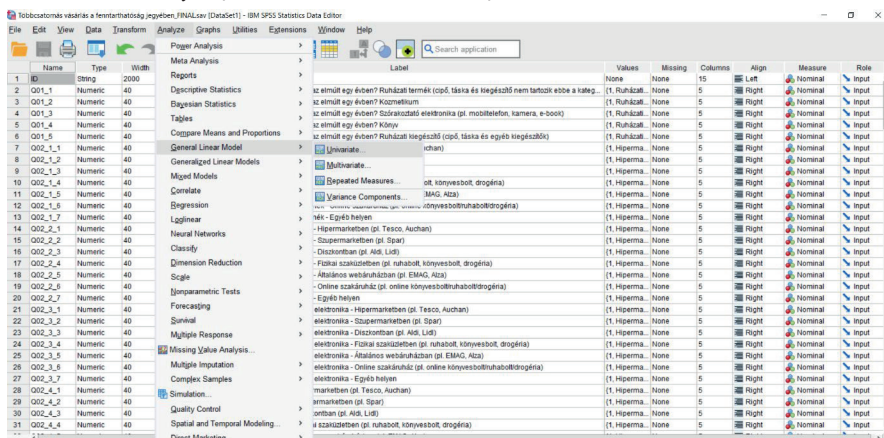
A piackutatócégünket megbízó vállalat szeretné megvizsgálni a vásárlók fizetési hajlandóságát a társadalmilag felelős termékek esetében. A vállalat, az online rendelt termékek kézhezvételéhez köthető környezettudatossági tapasztalatok alapján három szegmenst különít el. Kérdésként merült fel, hogy a vásárlók neme, illetve a szegmensbe tartozása közösen befolyásolja-e a vásárlók fizetési hajlandóságát a társadalmilag felelős termékekhez köthetően.

Elemzésünk során most az adatbázisból három változót választunk ki, hisz a kutatási kérdés kétszemponτος varianciaelemzéssel válaszolható meg.

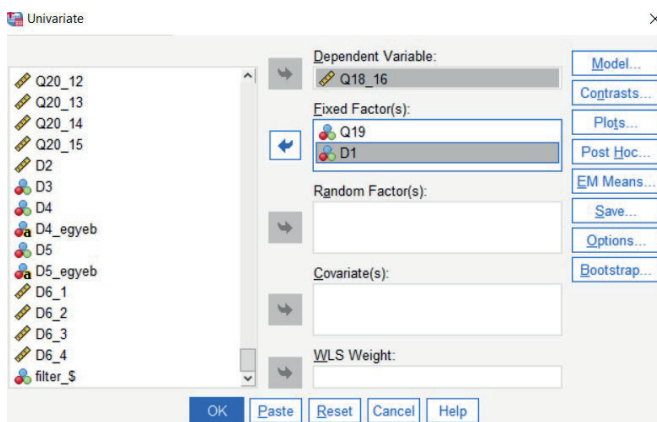
A függő változónk az adatbázis Q18_16-os Likert-skálán mért változója lesz (*Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? – Hajlandó vagyok többet fizetni a társadalmilag felelős termékekért akkor is, ha van olcsóbb alternatíva. / 1 – Egyáltalán nem jellemző, 7 – teljes mértékben jellemző*), és két független változóval dolgozunk. Az egyik a D1-es, nominális mérési szintű változó, mely a válaszadó nemét mutatja (1 – férfi, 2 – nő), míg a másik a Q19-es változó, mely a vállalat szegmenseit azonosítja (*Találkoztál-e már a vásárlásaid során olyan termékkel, amelyen valamilyen formában feltüntették a kézhez vétel (kiszállítás, bolti átvétel, csomagponti/csomagautomatás átvétel stb.) környezetbarát jellegét?*):

- 1 – akik gyakran találkoznak vásárlásaik során olyan termékekkel, amelyek kézhez vételéhez köthető környezetbarát jellegét feltüntetik,
- 2 – akik találkoztak már ilyen termékekkel, de nem túl gyakran,
- 3 – még sohasem találkoztak ilyen termékekkel.

Elérési út: Analyze / General Linear Model / Univariate



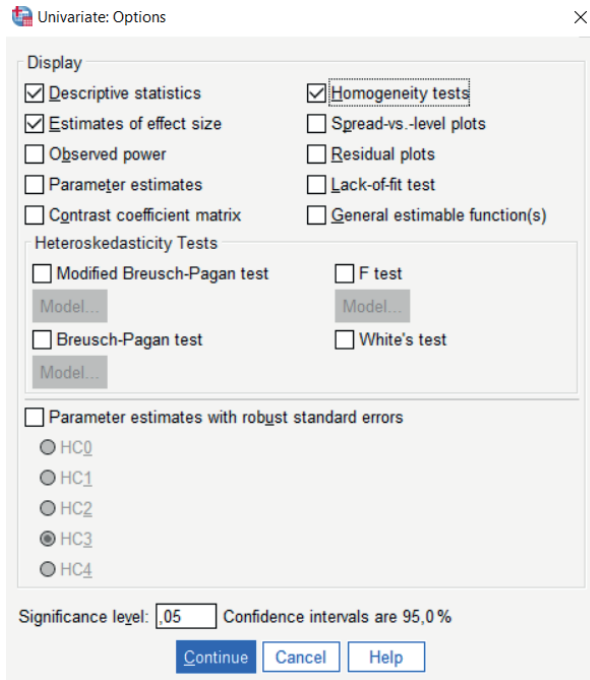
A felugró ablakban megadjuk az elemzésbe bevont változókat. A függő változónk (*Dependent Variable*) a Q18_16, míg a független változónk (*Fix factor(s)*) a Q19 és D1.



Lekért adatok a menüpontokból:

Options:

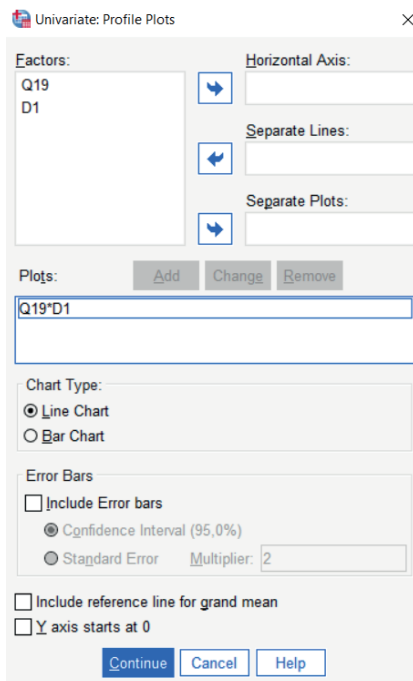
- ✓ *Descriptive Statistics* → a független változók által képzett csoportokhoz valamint a teljes mintához tartozó leíró statisztikákat foglalja össze (válaszadói számot, átlagot, szórást, minimumot, maximumot).
- ✓ *Estimate of effect size* → kapcsolat szorosságának mutatóját adja meg (E^2)
- ✓ *Homogeneity test* → szóráshomogenitás tesztje, Levene-teszt



Amennyiben az elemzésbe bevont független változók interakciója szignifikáns, érdemes őket grafikusán is ábrázolni, hogy az interakció jellegét (ordinális vagy nem ordinális interakció) megismerjük.

Ehhez a **Plots** menüpontból állítsunk elő egy olyan ábrát, mely a különböző csoportok átlagait mutatja.

- ✓ A vízszintes tengelyen (*horizontal axis*) érdemes azt a független ábrázolni, melynek kevesebb kimenete van, jelen esetben ez D1.
- ✓ A független vonalak (*separate lines*) sorba a másik független változót húzzuk be.
- ✓ Ezután adjuk hozzá az ábrát a lefuttatáshoz (*Add* parancs).
- ✓ Ezzel a *Plots* cellában létrejön az általunk definiált ábra (Q19*D1).



A parancsok lefuttatása után ismét a feltételek teljesülésével kezdjük az elemzésünket.

- A hibatag normál eloszlású legyen 0 átlaggal és konstans varianciával. Hasonlóan az előző, egyszempontos varianciaelemzéshez a leíró statisztikai mutatókat tartalmazó táblázatból (*Descriptives*) láthatjuk, hogy a mintaelemszám 30-nál nagyobb ($N=390$ fő), így a normál eloszlás megsértése a centrális határeloszlás törvény alapján figyelmen kívül hagyható, és az F-próba eredménye robusztusnak tekinthető. Ezek alapján a teszt elvégzésétől eltekinthetünk.

Descriptive Statistics

Dependent Variable: Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? -

Találkoztál-e már a vásárlásaid során olyan termékkel, amelyen valamilyen formában feltüntettek a kézhez vétel (kiszállítás, bolti átvétel, csomagponti/csomagautomatás átvétel stb.) környezetbarát jellegét?	Válaszadó neme	Mean	Std. Deviation	N
Igen, viszonylag gyakran.	Férfi	3,50	1,836	28
	Nő	3,77	1,655	44
	Total	3,67	1,720	72
Igen, de ritkán	Férfi	3,19	1,594	115
	Nő	3,77	1,462	100
	Total	3,46	1,558	215
Nem, még soha	Férfi	2,58	1,545	59
	Nő	3,95	1,555	44
	Total	3,17	1,687	103
Total	Férfi	3,05	1,640	202
	Nő	3,81	1,524	188
	Total	3,42	1,628	390

- Egyenlő variancia, vagyis szóráshomogenitás a csoportok között. A Levene-teszt értéke 1,205, a hozzá tartozó szignifikancia szint pedig 0,306. Ez alapján a H_0 hipotézist, vagyis, hogy a szóráshomogenitás fennáll, nem vetjük el, így az F-próba értéke megfelelő megbízhatósággal értelmezhető, nincs torzítás.

Levene's Test of Equality of Error Variances^{a,b}

		Levene Statistic	df1	df2	Sig.
Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Hajlandó vagyok többet fizetni a társadalmilag felelős termékekért akkor is, ha van olcsóbb alternatíva.	Based on Mean	1,205	5	384	,306
	Based on Median	,933	5	384	,459
	Based on Median and with adjusted df	,933	5	369,870	,459
	Based on trimmed mean	1,183	5	384	,317

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Hajlandó vagyok többet fizetni a társadalmilag felelős termékekért akkor is, ha van olcsóbb alternatíva.

b. Design: Intercept + Q19 + D1 + Q19 * D1

- A megfigyelések, a válaszadók egymástól függetlenek, kizárólagosak, vagyis nincsenek ismételt mérések, válaszadók a mintában, melyet esetünkben az adatfelvétel módja biztosít.

- Hasonló cellagyakoriság az összehasonlítani kívánt csoportokban.

A leíró statisztikai mutatókat tartalmazó táblázatban látható az egyes csoportok nagysága, mely alapján megállapíthatjuk, hogy a feltétel teljesül.

A feltételek teljesülésének köszönhetően az F-próba eredményei megbízhatóak.

Első körben a teljes hatáshoz (*Corrected Model*) tartozó F-próbát és szignifikancia szintet kell megvizsgálnunk. Jelen esetben az F-próba értéke 6,328, a szignifikancia szint pedig $<0,001$, így a modellünk létezik, vagyis létezik szignifikáns különbség a csoportok között, és lehetséges a további eredmények értékelése.

Ezután az *interakciós hatást* értelmezzük (Q19*D1). Az F-próba értéke 3,103, míg a hozzá tartozó szignifikancia szint 0,046, így a modellünkben a két független változó együttes hatása, interakciója érvényesül, vagyis a vásárlók szegmense és neme együttesen befolyásolja a fizetési hajlandóságukat a társadalmilag felelős termékek esetében. Az η^2 alapján egy gyenge kapcsolatról beszélünk (0,016).

Tests of Between-Subjects Effects

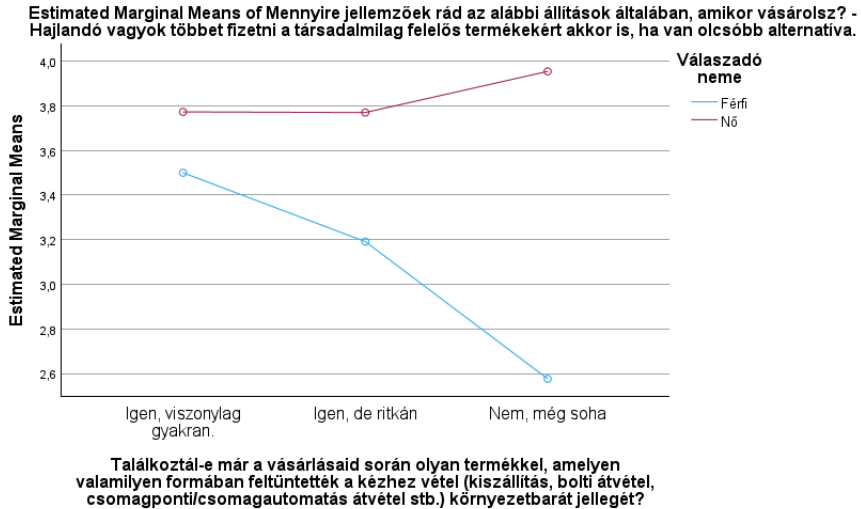
Dependent Variable: Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Hajlandó vagyok töt

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	78,491 ^a	5	15,698	6,328	<,001	,076
Intercept	3691,169	1	3691,169	1488,024	<,001	,795
Q19	6,012	2	3,006	1,212	,299	,006
D1	42,560	1	42,560	17,157	<,001	,043
Q19 * D1	15,393	2	7,696	3,103	,046	,016
Error	952,544	384	2,481			
Total	5594,000	390				
Corrected Total	1031,036	389				

a. R Squared = ,076 (Adjusted R Squared = ,064)

Mivel az átlagok ábrázolása alapján látszik (6.7. ábra), hogy nem ordinális interakciós hatásról beszélünk, mivel az egyes csoportokhoz tartozó átlagok sorrendje különbözik, így a főhatásokat (Q19 és D1 sorok) nem tudjuk értelmezni.

6.7. ábra: A részcsoportok átlagainak alakulása. (SPSS output)



Ebben az esetben is érdemes az eredményeket a csoportok és interakció szintjén értelmezni és prezentálni. Az eredmények alapján láthatjuk, hogy a nők fizetési hajlandósága alapvetően magasabb, azonban a szegmensbe tartozás ellenkező irányba befolyásolja a férfiak és nők hajlandóságát, ezért fontos a két változó interakciójának figyelembevétele.

Gyakorló feladat

Befolyásolja-e a válaszadó neme (D1) és a legutóbb vásárolt termék típusa (Q03) együttesen azt, hogy a vásárló számára mennyire fontos, hogy a vásárlásai ne okozzanak számára sok gondot (Q05_02)?

07 KORRELÁCIÓ ELEMZÉS

A korrelációelemzés a kapcsolatvizsgálati módszerek gyakran alkalmazott megközelítése a statisztikában. A marketingkutatók során azonban valamivel ritkábban fordul elő, hiszen magas mérési szintű változókat követel, amely a kérdőíves megközelítések során jellemzően kisebb mértékben áll a rendelkezésre.

Elméleti bevezetés

Az elemzés során két metrikus, vagy ordinális szinten mért változó kapcsolatát vizsgáljuk, melyek lehetnek akár metrikus-metrikus, ordinális-metrikus vagy ordinális-ordinális felállásban. A korrelációelemzés nem vizsgál ok-okozati összefüggést a két változó között. Az ugyenezen mérési szinten lévő változók ok-okozati összefüggését a regresszióelemzés keretein belül végezhetjük el.

Amennyiben mindkét változó metrikus mérésű, alkalmazható az ún. Pearson-féle, paraméteres korrelációs eljárás. Ha legalább az egyik vizsgálni kívánt változó azonban ordinális mérési szintet követ, akkor már a nem paraméteres Spearman-féle korrelációt kell alkalmazni a kapcsolat vizsgálatához.

A Pearson-féle korreláció az alábbi képlet alapján számolható ki.

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

ahol X_i és Y_i a változók értékei, valamint \bar{X} és \bar{Y} pedig az átlaguk az átlaguk.

Az elemzés során a keresztábrához hasonlóan itt is három kérdésre keressük a választ:

1. Szignifikáns kapcsolat van-e a két változó között?

Amennyiben nincs szignifikáns kapcsolat, akkor az elemzés ezen tény riportolásával be is fejeződik. Ha szignifikáns a kapcsolat akkor további két megválaszolendő kérdés merül fel:

2. Milyen szoros a kapcsolat?
3. Mit tudunk elmondani a kapcsolat irányáról?

A korreláció elemzés hipotézisei:

Ho: nincs szignifikáns kapcsolat a két változó között, a két változó nem mozog együtt

H1: szignifikáns kapcsolat van a két változó között, a két változó együtt mozog

A p-érték mutatja meg, hogy a megfigyelt együtt mozgás véletlenszerű-e vagy statisztikailag szignifikáns, azaz létezik a kapcsolat a két változó között. Ha a p-érték kisebb, mint a választott szignifikancia szint (pl. 0,05), akkor a korreláció szignifikánsnak tekinthető.

A kapcsolat szorosságát a Pearson-féle r mutató, vagy Spearman-féle rho mutatók mérik, amelyek segítenek meghatározni a két kategorizált változó közötti kapcsolat szorosságát.

A kapcsolatszorossági mutatók általános jellemzőjéhez hasonlóan az r és a rho mutatók értéke -1 és 1 között lehet, ahol 0 azt jelenti, hogy nincs kapcsolat a változók között, és a szélső értékek (1 vagy -1) nagyon erős, tökéletes kapcsolatot jeleznek. A szorosság mértékét abszolút értékben értelmezzük.

A mutatók előjele a kapcsolat irányáról árulkodik. A „+” előjel azt jelenti, hogy a két változó egyenesen arányos egymással, azaz egy irányba mozognak, ha az egyik változó növekszik, növekszik a másik változó értéke is. A „-” előjel ennek az ellentétét jelenti, a változók fordított arányosságát jelzi, azaz ellentétes irányba mozognak. Amennyiben az egyik változó növekszik, a másik változó értéke csökken.

Gyakorlati példa

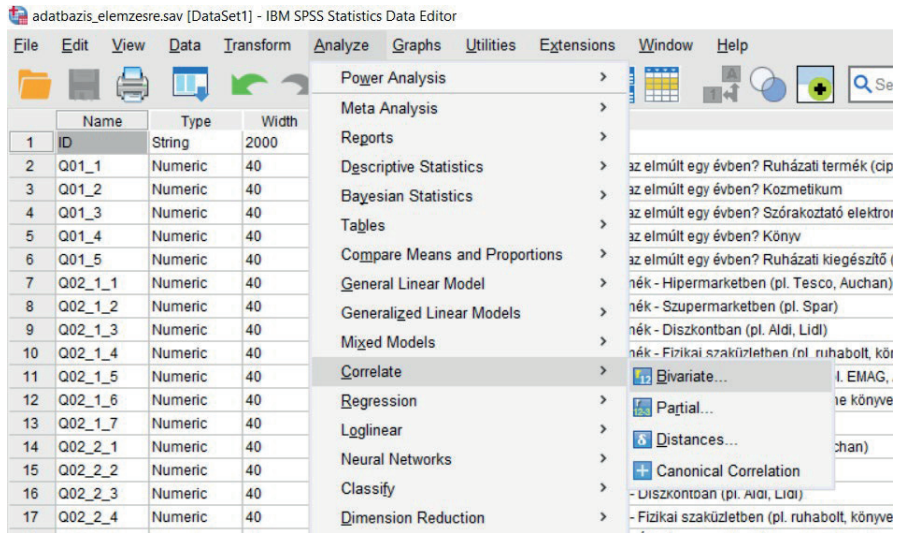
Milyen kapcsolat figyelhető meg az egyes kategóriák között a havi költségek tekintetében (D6_1 – D6_4)?

Mivel mind a négy változó metrikusnak tekinthető, ezért a közöttük lévő kapcsolatot korrelációelemzés segítségével vizsgáljuk. A vizsgálat során végső soron minden kategóriát párosítunk minden kategóriával, így összesen hat változópár keletkezik. Az alábbi hipotézist ezért minden kategória pár esetében meg kell fogalmazni. Tehát végső soron nem egy, hanem hat hipotézist vizsgálunk majd az elemzés során.

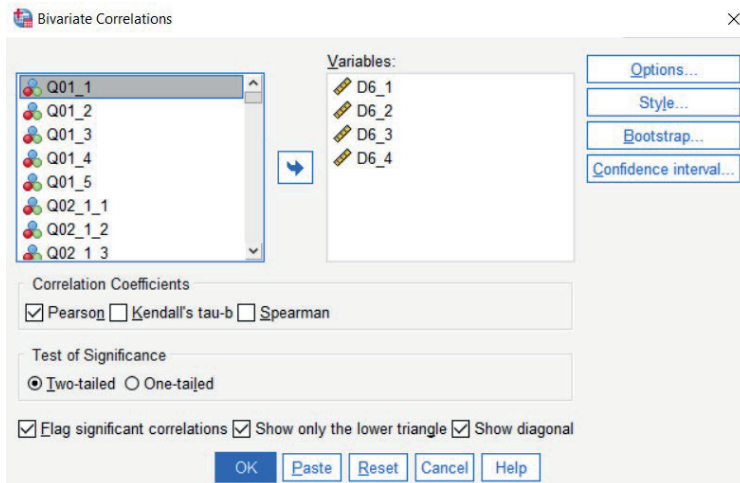
Ho: Az adott két kategória költései nem mozognak együtt

Hr: Az adott két kategória költései együtt mozognak

Elérési út: Analyze / Correlate / Bivariate



A beállítások során a vizsgálni kívánt változót *Variables* mezőbe helyezük. Mivel minden változónk metrikus, ezért a *Correlation coefficients* lehetőségek közül a *Pearson*-t választjuk. A kényelmesebb értelmezés miatt érdemes még jelölni a „*Flag significant correlations*” és a „*Show only the lower triangle*” opciókat.



Értelmezés

A keletkezett outputot korrelációs mátrixnak nevezzük, ahol sorokban és oszlopokban is megtalálhatóak a vizsgálni kívánt változók, ezáltal biztosítva a sorok és oszlopok találkozásánál, hogy minden változópár külön elemezhető legyen. Egy-egy ilyen találkozásnál, metszetenél mindig három adatot találunk. A *Pearson Correlation* a korrelációs mutatószámot, a *Sig. (2-tailed)* a p értéket, míg az N az elemszámot jeleníti meg.

Első lépésként meg kell állapítanunk, hogy szignifikáns kapcsolat van-e az adott változópár között, amihez a p értéket hívjuk segítségül. Azokban a cellákban, ahol a p érték 0,05-nél kisebb, el kell vessük a H_0 hipotézist, tehát szignifikáns kapcsolat igazolható. Ezt az SPSS program csillag szimbólum megjelenítésével is hangsúlyozza a Pearson Correlation sorok mutatószámainál. Röviden tehát, azon változó párok között lesz szignifikáns a kapcsolat, ahol megjelenik a * jel. Ez jelen példánkban minden változó pár esetében igaz lesz.

Ezt követi a szignifikáns változópárok kapcsolatának elemzésre, amiben a Pearson Correlation r mutatója segít bennünket. A mutató nagysága a kapcsolat szorosságát, míg előjele annak irányát jelzi.

Néhány példa az alábbi korrelációs mátrix eredményeiből:

A legerősebb szignifikáns kapcsolat havi költsékek esetében a szórakozás és az élelmiszer jellegű költsékek között van, a kapcsolat ebben az esetben pozitív és közepesen erős. Azaz a magasabb szórakozási költsékek együtt járnak a magasabb élelmiszer költsékekkel is ($r(388)=0,435$, $p<0,001$).

A leggyengébb szignifikáns kapcsolat a havi költsékek esetében szórakozás és a lakhatás között figyelhető meg. Aki többet költ lakhatásra, kis mértékben többet költ szórakozásra is, és fordítva ($r(390)=0,166$, $p<0,001$).

Correlations

		Mekkora összeget költesz átlagosan havonta az alábbi tétélekre? - Lakhatás (F/hó)	Mekkora összeget költesz átlagosan havonta az alábbi tétélekre? - Étkezés, élelmiszerek (F/hó)	Mekkora összeget költesz átlagosan havonta az alábbi tétélekre? - Utazás (F/hó)	Mekkora összeget költesz átlagosan havonta az alábbi tétélekre? - Szórakozás - (F/hó)
Mekkora összeget költesz átlagosan havonta az alábbi tétélekre? - Lakhatás (F/hó)	Pearson Correlation	--			
	N	390			
Mekkora összeget költesz átlagosan havonta az alábbi tétélekre? - Étkezés, élelmiszerek (F/hó)	Pearson Correlation	,266**	--		
	Sig. (2-tailed)	<,001			
Mekkora összeget költesz átlagosan havonta az alábbi tétélekre? - Utazás (F/hó)	Pearson Correlation	,272**	,338**	--	
	Sig. (2-tailed)	<,001	<,001		
Mekkora összeget költesz átlagosan havonta az alábbi tétélekre? - Szórakozás - (F/hó)	Pearson Correlation	,167**	,435**	,402**	--
	Sig. (2-tailed)	<,001	<,001	<,001	
	N	390	388	388	390

** . Correlation is significant at the 0.01 level (2-tailed).

Gyakorló feladat

Milyen a kapcsolat a környezettudatossággal kapcsolatos attitűdök tekintetében (Q20_1 – Q20_15)?

08 REGRESSZIÓELEMLZÉS

A többváltozós statisztikai elemzés legerjedtebb és legtöbbet alkalmazott módszere napjainkban a regresszióelemzés. Ez több okra vezethető vissza. Egyrészt, a regresszióelemzés során egy vagy több független (magyarázó) változó egy adott függő változó értékére kifejtett hatásait tudjuk megbecsülni. Az elemzésbe bevont változók száma lényegében tetszőleges lehet, annak csak a mintanagyság (és az elérhető változók köre) szab korlátokat. Másrészt, a regresszióelemzés – később ismertetésre kerülő – előfeltevései sok esetben feloldhatók vagy nem teljesülésük kezelhető, így pedig lényegében minden kutatási kérdés esetén alkalmazhatóvá válik a módszer. Végezetül, az elemzésbe bevont változók mérési szintje is tetszőleges lehet, bár ebben a fejezetben kizárólag azt az esetet tárgyaljuk, amikor a függő változó metrikus (a magyarázó változó viszont lehet metrikus vagy kategorikus is). A következő fejezetben azonban bemutatjuk, hogyan kezelhető az az eset, amikor a függő változó sem metrikus.

Fontosnak tartjuk kiemelni, hogy a regresszióelemzés előbb ismertetett előnyei következtében a módszertan nagyon szerteágazó, ebben a könyvben mindössze ízelítőt, rövid betekintőt tudunk nyújtani. Akit mélyebben, részletesebben érdekel a regresszióelemzés, annak további, kimondottan a témával foglalkozó szakkönyveket javasolunk, például *Békés Gábor és Kézdi Gábor „Adatalemzés – Üzleti, közgazdasági és szakpolitikai döntésekhez”* (Alinea Kiadó, 2024) vagy *Jeffrey M. Wooldridge „Introductory Econometrics: A Modern Approach”* (Cengage, 2020) című műveit.

Elméleti bevezetés

A lineáris regresszióelemzés során a függő változót a magyarázó változók lineáris függvényeként írjuk fel, vagyis a következő becslőfüggvénnyel dolgozunk:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u, \quad (8.1)$$

ahol y jelöli a függő (vagy eredmény-) változót, x_1, x_2, \dots, x_k a magyarázó változókat, $\beta_1, \beta_2, \dots, \beta_k$ pedig a magyarázó változók függő változóra való hatásait mérő paramétereket (β_0 egy igazodást segítő konstans), míg u az ún. reziduomot vagy hibtagot. Utóbbi tehát a megfigyelések azon része, amelyet a regressziós modell nem magyaráz meg, vagyis a valós értékek regressziós becsléstől való eltérései.

A regresszióelemzés során célunk a paraméterek megbecslése, tehát annak meghatározása, hogy az egyes magyarázó változók milyen mértékben befolyásolják a függő változó értékét. Az eredmények alapján tehát nem csak a változók közötti kapcsolat irányára, hanem annak pontos mértékére is válaszokat kapunk, ez a regresszióelemzés egyik fő értéke, erőssége.

Egyváltozós regresszió esetén egyetlen magyarázó változóval, míg **többsváltozós regresszió** során több magyarázó változóval magyarázzuk a függő változót. Ezt leszámítva érdemi eltérés nincs a kétféle eset között, ezért a továbbiakban nem különböztetjük meg ezeket.

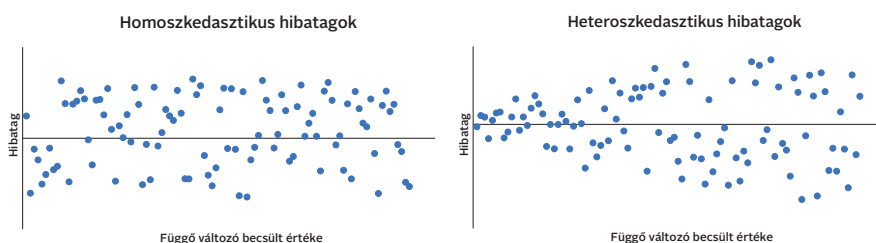
Általánosságban hat olyan feltétel van, amelyek teljesülése garantálja a lineáris regressziós becslés torzítatlanságát, illetve azt, hogy a lehetséges egyéb torzítatlan becslőfüggvények közül a legjobb lesz (hatásosság), emellett pedig az eredmények vizsgálatához szükséges tesztek is kellő biztonsággal elvégezhetők legyenek. Ezen hat feltétel az alábbi:

1. **Linearitás**, vagyis a változók közötti lineáris kapcsolat (amelyet a modell felírásakor feltételezünk) a valóságban is megállja a helyét. Amennyiben a (8.1) egyenletben felírt lineáris becslőfüggvény nem igaz a valóságban (a valóságban nem így generálódnak az adatok), akkor a becsült paraméterek sem tudják megfelelően leképezni a valóságot.
2. **Véletlen minta**, vagyis az adatfelvétel során az alapsokaságból véletlenszerűen kerültek kiválasztásra a kitöltők/megfigyelt alanyok, minden egyednek ugyanakkora volt a bekerülési valószínűsége. Ez a feltevés garantálja azt, hogy a mintavételi hiba statisztikai módszerekkel kezelhető, vagyis statisztikai következtetések levonhatók a vizsgált alapsokaságra vonatkozóan.
3. **Tökéletes multikollinearitás hiánya**, vagyis a magyarázó változók között nincs tökéletes lineáris kapcsolat (+1 vagy -1 korreláció). Ugyanakkor a magyarázó változók közötti erős, de még nem tökéletes multikollinearitás is problémás lehet, ez növeli a paraméterbecslések bizonytalanságát. Erről a későbbiekben még szó lesz.
4. **Exogenitás**, vagyis a magyarázó változók és az elméleti hibatarag (u) közötti lineáris kapcsolat, korreláció hiánya. Ez lényegében azt jelenti, hogy a modell teljesnek tekinthető, minden fontos változót figyelembe vettünk, vagy legalábbis csak olyan változók maradtak ki a modelltől, amelyek a modellben szereplő magyarázó változóktól lineárisan függetlenek. Ez a feltevés okozza egyébként a regressziós becslés torzítottságát a legtöbb esetben, éppen ezért a regresszióelemzéssel foglalkozó szakirodalom egyik kulcskérdése, hogy az exogenitási feltevés nem teljesülése (tehát endogenitás) esetén milyen módszerrel biztosítható mégis a becslések torzítatlansága. Fontossága ellenére jelen fejezet kereteit

meghaladja ennek részletes tárgyalása, azonban a harmadik esettanulmányban részben kitérünk erre a problematikára és lehetséges kezelésére is.

5. **Homoszkedaszticitás**, vagyis a reziduumok szórásának állandósága, annak függetlensége a magyarázó változók aktuális értékeitől. Ha ez nem teljesül, akkor heteroszkedaszticitásról beszélünk. A homoszkedaszticitási feltevés lényege az, hogy a magyarázó változók kisebb és nagyobb értékei esetén sem válik bizonytalanabbá a becslés, ahogy azt a 8.1. ábra is mutatja.

8.1. ábra: Homoszkedasztikus és heteroszkedasztikus hibataragok illusztrációja



6. **Reziduumok normális eloszlása**, amit nagy mintánál ($n > 100$) a centrális határeloszlás tétele alapján adótnak veszünk, kisebb mintánál eloszlástesztetekkel ellenőrizhető.

A regressziós becslés torzítatlanságához az 1–4. feltételek teljesülése szükséges, a hatásos (lehető legpontosabb) becsléshez ezen felül az 5. feltétel is elengedhetetlen, a 6. feltétel pedig ahhoz kell, hogy a következőkben tárgyalt hipotézisvizsgálatok elvégezhetőek legyenek, pontosabban azok eredménye megbízható és helyes legyen.

A regressziós modell megfelelésének vizsgálata

Egy regresszió lefuttatása napjainkban már nem tekinthető nehéz feladatnak, többféle statisztikai és ökonometriai programcsomag áll rendelkezésre hozzá, ezek közül több ingyenesen is elérhető (pl. R). Az elemző legfontosabb feladata az eredmények áttekintése és értelmezése. Az eredmények értelmezése előtt azonban szükséges átgondolni, hogy az előző alfejezetben bemutatott feltevések vélhetően teljesülnek-e a modell esetében, vagyis számíthatunk-e arra, hogy az eredmények torzítatlanok és hatásosak lesznek. Ez azért kiemelkedően

fontos, mert elképzelhető, hogy a modell teljesíti a következőkben leírt teszteket, tehát alkalmazhatónak tűnik, azonban a becslési eredmények mégis torzítottak, vagyis valójában nem vagy csak korlátozottan alkalmasak érdemi következtetések levonására vagy éppen döntéstámogatásra. Erről részletesebben a gyakorlati példa során lesz szó.

A modell becslése után fontos megvizsgálni, hogy a felépített regressziós modell megfelelően illeszkedik-e az adatokra, van-e magyarázóereje, illetve – amennyiben igen, akkor – minden modellben szereplő magyarázó változóra szükség van-e. Mindkét kérdés megválaszolásához hipotézisvizsgálatokat végzünk.

A **modell globális tesztelését**, vagyis annak vizsgálatát, hogy a modell egésze jól leírja-e a vizsgált adathalmazban rejlő összefüggéseket ún. *Wald F*-próbával végezzük. Ennek nullhipotézise szerint a modell minden becsült paramétere zérus, vagyis a magyarázó változók egyike sem rendelkezik érdemi magyarázóerővel, nem hat a függő változóra. Az alternatív hipotézis azt mondja ki, hogy legalább az egyik (akár mindegyik, de nem feltétlenül) magyarázó változónak van szignifikáns hatása a függő változóra.

A modell globális tesztelésének hipotézisei:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \exists \beta_i \neq 0 (i=1, 2, \dots, k)$$

A nullhipotézis elfogadása esetén a modell adott formájában nem alkalmas a további elemzésre, ilyenkor más függvényformával vagy más változókészlettel érdemes továbbhaladni. Amennyiben a nullhipotézis elutasításra kerül, akkor szükséges annak vizsgálata, hogy az elemzésbe bevont mindegyik magyarázó változó szükséges-e, van-e érdemi magyarázóereje. Ebben az esetben tehát **a változók egyedi tesztelését** végezzük el *t*-próbával. A *t*-próba célja megvizsgálni, hogy egy adott regressziós együttható különbözik-e nullától (azaz szignifikáns-e a modellben), vagyis szükség van-e az adott magyarázó változóra a regresszióban.

A változók egyedi tesztelésének hipotézisei:

$H_0: \beta_i = 0$, vagyis az *i*-edik regressziós együttható nem különbözik szignifikánsan 0-tól, azaz az *i*-edik magyarázó változót el kell távolítani a modellből;

$H_1: \beta_i \neq 0$, vagyis az *i*-edik regressziós együttható szignifikánsan különbözik 0-tól, azaz az *i*-edik magyarázó változó érdemben hat a függő változóra.

Amennyiben egyes magyarázó változók modellből való eltávolítása szükségessé válik (tehát elfogadjuk az adott magyarázó változó esetén a nullhipotézist), úgy azt egyesével érdemes elvégezni, mindig azzal a magyarázó változóval kezdve, amelyik esetén a t -próba empirikus szignifikanciaszintje (p -értéke) a legnagyobb volt. Ezt hívják *backward eliminációnak*. Ennek oka, hogy általában a magyarázó változók is összefüggnek egymással (van valamilyen szintű *multikollinearitás*), tehát egy magyarázó változó modellből való elhagyása adott esetben szignifikánssá tehet egy korábban inszignifikáns magyarázó változót.

A már kizárólag szignifikáns magyarázó változókat tartalmazó modell megszületése után válik szükségessé annak értelmezése. Ennek során a modell magyarázóerejét, illetve az egyes becsült paramétereket értelmezzük.

A *modell magyarázóerejét* az R^2 mutató méri, ennek értéke 0 és 1 közé eshet. Az R^2 azt mutatja meg, hogy a függő változóban lévő információ (variancia) hány százalékát magyaráztuk meg a modellel, vagyis a modell ismeretében mennyivel csökkent a függő változó előrejelzésének bizonytalansága. Az R^2 mutató nem minden esetben egyformán fontos, alapvetően akkor szoktuk használni, ha cél a minél pontosabb előrejelzés (pl. keresletbecslés esetén).

A becsült paraméterek értelmezése a regresszióelemzés egyik kulcsa, hiszen ezáltal tudunk képet alkotni arról, hogy a magyarázó változók (minden más változatlansága mellett) hogyan befolyásolják a függő változót. Az adott (i -edik) magyarázó változóhoz kapcsolódó becsült paraméter (β_i) azt mutatja meg, hogy ha a magyarázó változó értéke minden más változatlansága mellett egy egységgel nő, akkor várhatóan mennyivel változik a függő változó értéke.

Dummy változók

Ezen a helyen szükséges megemlíteni, hogy a regresszióelemzésbe kvalitatív (nem metrikus skálán mért) változók is bekerülhetnek, ezeket ún. *dummy* (magyarul vak-) változókkal kell kódolni. Egy dummy változó mindig két értéket vehet fel: 0-t vagy 1-et. 0-t akkor vesz fel, ha az adott tényező nem teljesül az adott megfigyelés esetén (ezt lehet referenciacsoportnak is hívni), míg 1-et akkor, ha teljesül. Amennyiben egy kvalitatív változó kettőnél több értéket is felvehet, akkor a felvehető értékek számánál eggyel kevesebb számú új dummy változóra lesz szükség a regresszióban. Ekkor a kihagyott érték lesz a referenciacsoport, a többi értéket ehhez viszonyítjuk majd. A dummy változókhoz kapcsolódó becsült paraméterek pedig azt mutatják meg, hogy ha az adott tulajdonság/szempont teljesül, akkor minden más változatlansága mellett várhatóan mennyivel változik (nő vagy csökken) a függő változó értéke (a referenciacsoportéhoz képest).

Gyakorlati példa

E rövid elméleti felvezetés és áttekintés után nézzük meg egy konkrét példán keresztül, hogy a regresszióelemzés hogyan alkalmazható a gyakorlatban.

Egy kozmetikai kiskereskedő vállalat azon gondolkozik, hogy bevezesse termékei online értékesítését. A megfelelő döntés meghozatalához azt szeretné megvizsgálni, hogy milyen tényezők befolyásolják azt, hogy valaki milyen gyakran vásárol kozmetikumokat online.

A kérdés megválaszolásához a Qo6_2 (*Milyen gyakran szoktad az adott termék kategóriát online vásárolni (10 alkalomból hányszor)? – Kozmetikumok*) változóra ható tényezőket kell elemeznünk. Tehát a függő változó a Qo6_2 lesz.

Az adatbázisban lévő változók közül elsősorban a demográfiai változókat (D1, D3, D4, D5, D7) vesszük figyelembe, emellett pedig néhány Likert-skálás kérdésre adott választ (Q18_02, Q18_04, Q18_05, Q18_07, Q18_15, Q18_16), valamint anyagi lehetőségeket mérő válaszokat (D6_2, D6_4), illetve szintén szerepeltetjük a kérdőív felvételét megelőző egy hónapban online végzett kozmetikumvásárlások számát (Qo8_2). Mivel kizárólag azok érdekelnek minket, akik vásárolnak (az elmúlt egy évben vásároltak) kozmetikumokat, ezért ez bekezdés szűrőfeltételnek (Qo1_2 = 2).

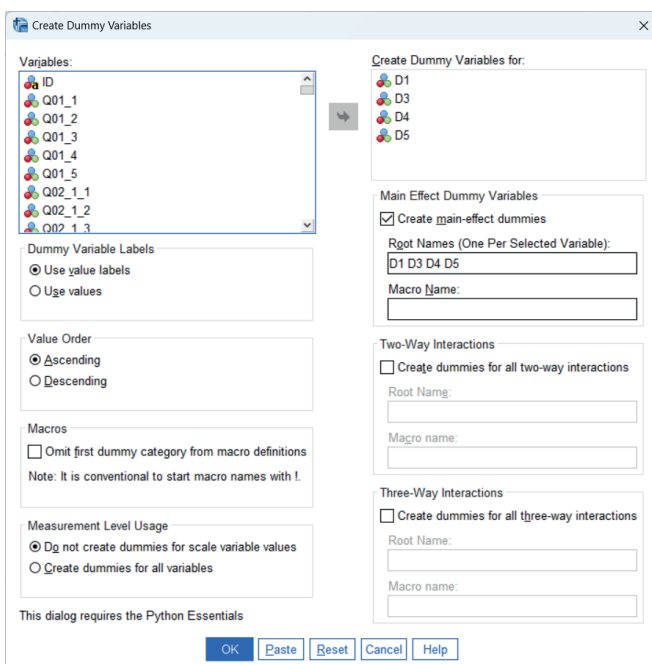
Első lépésként fontos áttekinteni az egyes változók mérési szintjét, és a nem metrikus mérési szintű változókat dummy változókká kell alakítani:

- Qo6_2: Arányskálán mért változó, átalakítást nem igényel.
- D1: Nominális skálán mért változó, szükséges átkódolni dummy változóvá (új változó: D1_ferfi). Az új változó 1 értéket vesz fel, ha valaki férfi, egyébként (tehát nőknél) 0-t, tehát a nők alkotják a referenciacsoporthoz.
- D3: Nominális skálán mért változó, szükséges átkódolni dummy változóvá (új változó: D3_Bp). Az új változó 1 értéket vesz fel, ha valaki Budapesten lakik, egyébként 0-t, tehát ebben az esetben a nem Budapesten lakók alkotják a referenciacsoporthoz.
- D4: Nominális skálán mért változó, szükséges átkódolni összesen négy dummy változóvá (új változók: D4_sajat, D4_kollegium, D4_bejar, D4_egyebd). Az új változók akkor vesznek fel 1 értéket, ha az adott egyén saját vagy szülei, rokonai tulajdonában álló budapesti lakásban él (D4_sajat), kollégiumban lakik (D4_kollegium), naponta bejár Budapestre (D4_bejar) vagy egyéb választ adott (D4_egyebd). A referenciacsoporthoz, vagyis akikhez viszonyítani fogjuk az eredményeket, a Budapesten albérletben lakók alkotják.
- D5: Nominális skálán mért változó, szükséges átkódolni összesen négy dummy változóvá (új változók: D5_resz, D5_alkalmi, D5_nem, D5_egyebd). Az új változók akkor vesznek fel 1 értéket, ha az adott egyén részmunkaidőben dolgozik (D5_resz), alkalmi munkát vállal (D5_alkalmi), nem dolgozik (D5_nem) vagy egyéb választ adott (D5_egyebd). A refe-

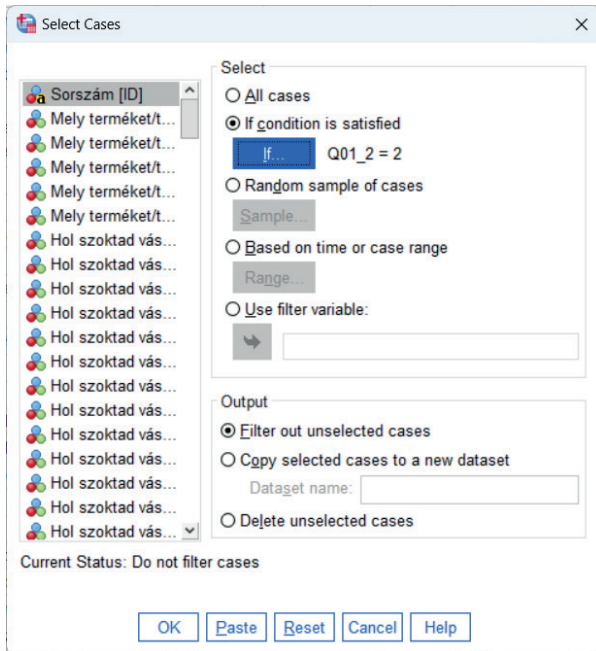
renciacsoportot, vagyis akikhez viszonyítani fogjuk az eredményeket, a tanulmányaik mellett főállásban dolgozók alkotják.

- D7: Arányskálán mért változó, átalakítást nem igényel.
- Q18_02, Q18_04, Q18_05, Q18_07, Q18_15, Q18_16: Likert-skálás kérdések, amelyek statisztikai értelemben véve ordinális skálán vannak mérve, azonban marketingkutatósi szempontból metrikus változóknak szoktuk tekinteni ezeket, így jelen esetben további átalakítást nem igényelnek.
- D6_2, D6_4: Arányskálán mért változók, átalakítást nem igényelnek.
- Qo8_2: Arányskálán mért változó, átalakítást nem igényel.

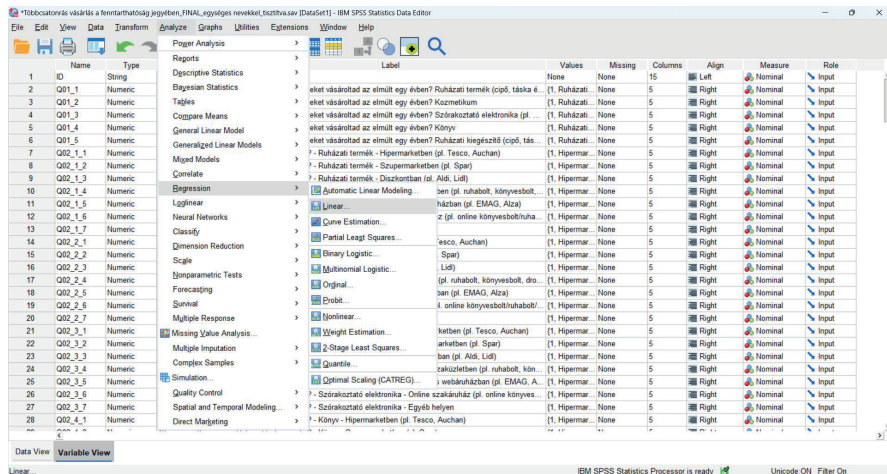
A szükséges új dummy változók létrehozását a **Transform / Create Dummy Variables** paranccsal végezzük el, majd a létrejött dummy változókat átnevezzük a fentiek alapján.



Ezt követően szükséges leszűrni az adatbázist azokra a válaszadókra, akik az elmúlt egy évben vásároltak kozmetikumot. Ehhez a **Data / Select Cases** parancsot használjuk.

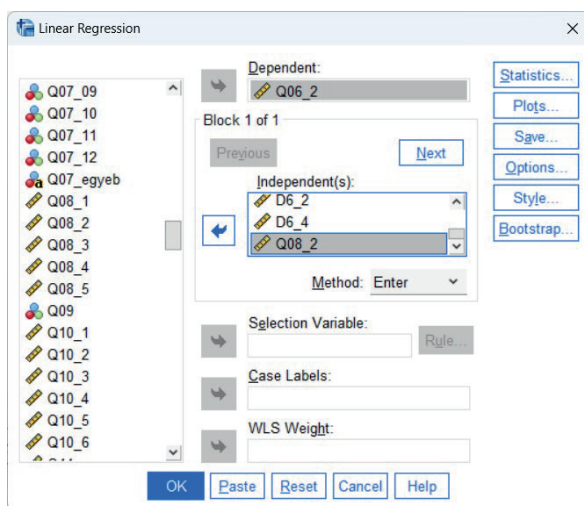


A szűrés elvégzése után a lineáris regresszió parancsot kell kiválasztani SPSS-ben: **Analyze / Regression / Linear**.



A felugró ablakban ki kell jelölni a függő (*Dependent*), illetve a magyarázó (*Independent(s)*) változókat. Jelen esetben a függő változó a Q06_2, míg a magyarázó változók: D1_ferfi, D3_Bp, D4_sajat, D4_kollegium, D4_bejar, D4_egyebd, D5_

resz, D5_alkalmi, D5_nem, D5_egyebd, D7, Q18_02, Q18_04, Q18_05, Q18_07, Q18_15, Q18_16, D6_2, D6_4 és Q08_2.



Lekért adatok a menüpontokból:

Ezek pontos kiválasztása esetről esetre változó lehet, de általában az alábbiakat érdemes kiválasztani.

Statistics

- ✓ *Estimates*: regressziós együtthatók, paraméterek egyedi tesztelése – *t*-próbák
- ✓ *Model fit*: modell globális tesztelése – *F*-próba
- ✓ *Descriptives*: változók leíró statisztikái
- ✓ *Part and partial correlation*: multikollinearitás vizsgálatához

Plots

- ✓ *X*: ZPRED (standardizált előrejelzés), *Y*: ZRESID (standardizált hibtag) (homoszkedaszticitás szemrevételezés vizsgálatához)
- ✓ *Standardized residual plots*: Histogram (hibtagok normális eloszlásának vizsgálatához)

Ezután lefuttatva a regressziót, megkapjuk az eredményeket több táblázatba szétbontva, ábrákkal kiegészítve. Az eredmények értelmezése előtt azonban fontos az előfeltevések közül azok vizsgálata, amelyeké megtehető.

A regresszió torzítatlansága az első négy feltevés teljesülése esetén biztosított. Ezek teljesüléséről eltérő mértékben és módokon tudunk meggyőződni.

1. **feltevés (linearitás)**: ez a feltevés grafikusan, a függő és a magyarázó változók ábrázolásával elvileg vizsgálható, azonban ehhez egyrészt nagyon

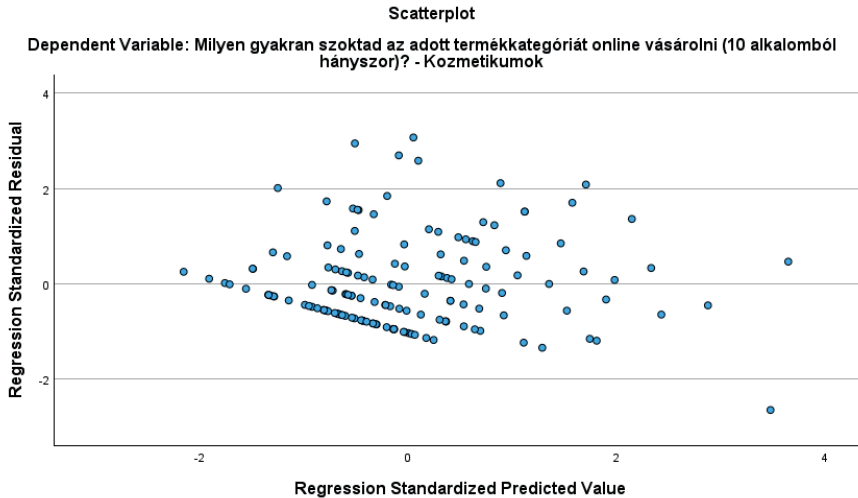
sok ábrát kellene szemrevételezéssel áttekinteni, illetve valójában a feltételes (a többi magyarázó változóra kontrollált) kapcsolatoknak kell lineárisnak lenniük, az ábrák viszont a nem feltételes kapcsolatokat mutatják csak. A szemrevételezés pedig nem tekinthető elfogadott statisztikai módszernek. Így a legtöbb esetben ez egy logikai átgondolást igényel, annak átgondolását, hogy mennyire reális azt feltételezni, hogy az adott magyarázó változó lineáris módon hat a függő változóra.

2. **feltevés (véletlen minta):** a regressziós elemzés pillanatában a minta már adottság, erre a feltételre a mintavétel során kell kiemelt figyelmet fordítanunk (csakúgy, mint a minta reprezentativitásának kérdésére).
3. **feltevés (tökéletes multikollinearitás hiánya):** ez könnyen vizsgálható az SPSS által kiadott korrelációs mátrix (*Correlations*) alapján. Tökéletes multikollinearitás esetén egyébként le sem fut a regresszió, tehát ennek a feltevésnek a megszegése azonnal láthatóvá válik. Pontosabban, amennyiben két magyarázó változó tökéletesen összefügg egymással, akkor az SPSS az egyiket automatikusan eltávolítja a modelltől, és így közli az eredményeket. Ugyanakkor a **magyarázó változók közötti** nagyon jelentős (0,9 vagy inkább a fölötti, illetve -0,9 vagy inkább annál kisebb) korrelációk sem kedvezők. Ilyen esetekben érdemes lehet átgondolni, hogy mindkét érintett változóra szükség van-e, össze lehet-e esetleg vonni a két változót. Jelen esetben abszolút értelemben a legnagyobb korreláció a Q18_15 és a Q18_16 változók között van, értéke 0,816. Ez még bőven elfogadhatónak tűnik, így nincs szükség egyéb átalakításokra.
4. **feltevés (exogenitás):** ez a feltevés a regresszió validitásának, torzítatlanságának kulcsa, viszont teljesülése empirikusan nem vizsgálható az adatok alapján. Általánosságban elmondható, hogy nagyon ritkán teljesül ez a feltevés, hiszen például jelen esetben is a kozmetikumok online vásárlására vélhetően olyan tényezők is hatnak, amelyek nem részei az adatbázisunknak, tehát nem tudjuk megjeleníteni a regresszióban. Ilyen változó lehet például a kedvenc márka elérhetősége a közeli üzletekben vagy egyedi kozmetikumok vásárlása, amelyek sokszor csak online érhetők el a gyártó webshopján keresztül.

A többi feltevés teljesülése már nem szükséges ahhoz, hogy torzítatlan legyen a regressziós becslésünk, azonban érdemes megvizsgálni a teljesülésüket, hogy biztosak lehessünk abban, hogy a lehető legjobb lineáris becslőfüggvényt használjuk (5. feltevés), illetve, hogy a később elvégzendő tesztek eredményei megfelelők (6. feltevés).

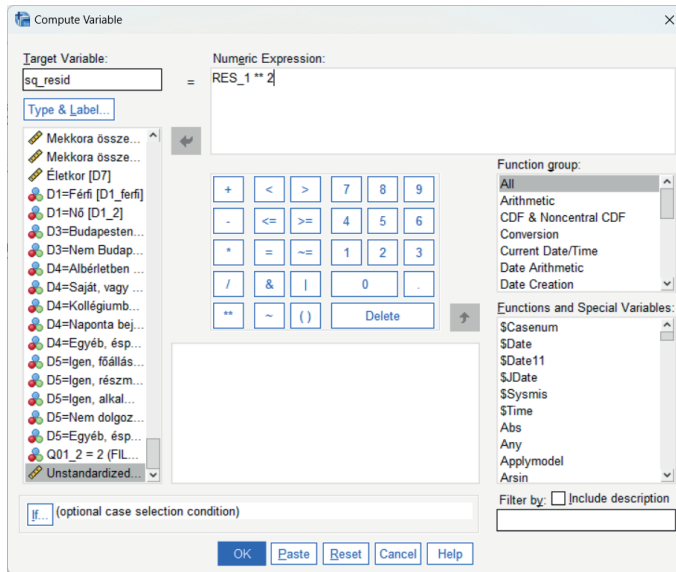
5. **feltevés (homoszkedaszticitás):** ez a feltevés a regressziós hibatagok szórásának állandóságára vonatkozik, teljesülése esetén a magyarázó

változók értékétől nem függ a hibatarag szórása. A homoszkedaszticitás tesztelésére léteznek különféle statisztikai tesztek (pl. Breusch–Pagan-teszt), azonban vizuálisan is vizsgálható, hogy a hibataragok szórása mennyire tűnik állandónak a prediktált függő változó értékének függvényében. Az alábbi ábrán jól látszik, hogy minél nagyobb a függő változó becstt értéké, annál nagyobb sávbán szóródnak a hibataragok, ami heteroszkedaszticitásra utal.



A homoszkedaszticitás formális teszteléshez többféle teszt is létezik, most a Breusch–Pagan-teszt SPSS-ben való megvalósítását mutatjuk be. E teszt lényege, hogy ha a négyzetes hibatarag (amivel a hibatarag szórását becsttjük, hiszen várható értéké 0) értékét befolyásolják a modellben szereplő magyarázó változók értéké, az heteroszkedaszticitásra utal. Amennyiben a négyzetes hibatarag értéké nem magyarázható a magyarázó változók által, akkor a modell homoszkedasztikus. A teszt sajnos nem érhető el közvetlenül SPSS-ben. A manuális megvalósítás lépései az alábbiak.

- I. A regresszió lefutttatásakor le kell mentenünk a hibataragokat, amelyhez a korábban bemutatott **Linear Regression** felugróablakban a **Save** menüpontot választva a **Residuals** alatt kell bejelölni az **Unstandardized** lehetőséget, vagyis a regresszió nem standardizált hibataragjait kell lementenünk.
- II. A lementett hibataragok négyzetes értékét kell előállítanunk, amihez a **Transform / Compute Variable** parancsot használjuk, az új változót most „sq_resid” néven nevezzük el.



- III. Újra lefuttatjuk a regressziót ugyanazokkal a magyarázó változókkal, mint eredetileg, de a függő változót cseréljük a II. pontban előállított négyzetes hibataragra (sq_resid).
- IV. A III. pontban lefuttatott regresszió globális F -tesztjét szükséges megvizsgálni. Amennyiben ez alapján a regresszió szignifikáns, vagyis a p -érték döntési határtól függően 5% vagy 1% alatt van, akkor az eredeti regresszió hibatarjai heteroszkedasztikusak, hiszen a magyarázó változók érdemben hatnak a négyzetes hibatar értékére. Egyébként, ha tehát a modellt elutasítjuk, akkor a hibatarok homoszkedasztikusak. Jelen esetben a hibatar négyzetét magyarázó regresszió inszignifikáns még 10%-os szignifikanciaszinten is ($p = 0,461$), így az eredeti regresszió hibatarjai homoszkedasztikusak, vagyis a feltevésnek megfelelők.

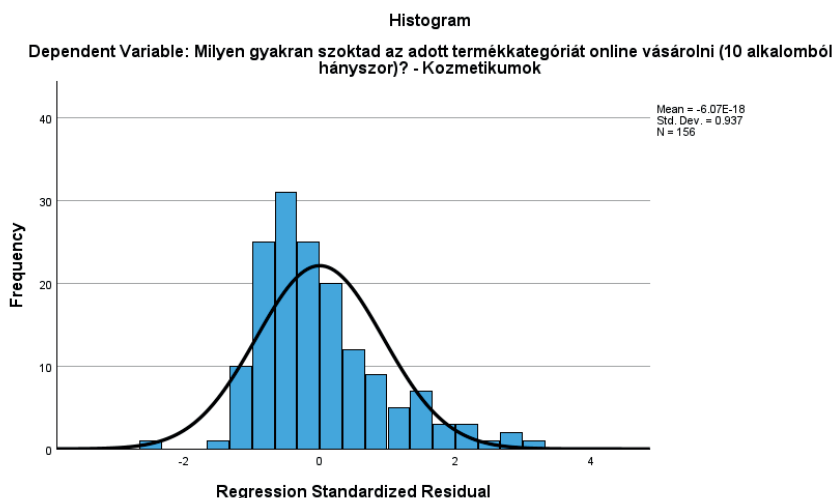
ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1006.543	19	52.976	1.004	.461 ^b
	Residual	7177.129	136	52.773		
	Total	8183.672	155			

Az ábra és a formális statisztikai teszt tehát eltérő eredményekre vezetett, ami felhívja a figyelmet az ábrák alapján történő vizuális követ-

keztetéslevonások problematikájára, szubjektívására. Javasoljuk tehát minden esetben a formális statisztikai teszt elvégzését, megbízható következtetések levonására ugyanis csak az ad biztos alapot.

- feltevés (reziduumok normális eloszlása):** a hibatagok normális eloszlását szintén vizsgálhatjuk grafikusan és formális statisztikai tesztek segítségével is. Az alábbi hisztogram alapján támadhatnak kételyeink a hibatagok normális eloszlását illetően, amelyet sajnos igazolnak a normalitástesztek (Kolmogorov – Smirnov és Shapiro – Wilk) is, hiszen ezek p -értéke 0,001-nél kisebb, vagyis a normális eloszlás nullhipotézisét minden szokásos szignifikanciaszinten elutasítjuk. E tesztek lefuttatásához az **Analyze / Descriptive Statistics / Explore** parancsot kell lefuttatni a korábban (a homoszkedaszticitás tesztelésénél) lementett hibatagokra, majd a **Plots** menüpontnál bejelölni a *Normality plots with tests* opciót.



Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
RES_1 Unstandardized Residual	.106	156	<.001	.928	156	<.001

a. Lilliefors Significance Correction

Ugyanakkor a centrális határeloszlás tétele miatt nagy (jellemzően 100 fölötti elemszámú) minták esetén a hibatagok normalitását adott-ságnak szoktuk venni, így valójában a 6. feltevést nem szoktuk vizsgálni.

A feltevések vizsgálata után térjünk át a modell eredményeinek értékelésére, ezen belül is elsőként a modell egészének tesztelésére (globális F -teszt). Az alábbi ábra mutatja az F -teszt eredményeit, amely alapján a modell minden szokásos szignifikanciaszinten szignifikáns, hiszen a p -érték 0,001-nél is kisebb.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	261.644	19	13.771	2.905	<.001 ^b
	Residual	644.580	136	4.740		
	Total	906.224	155			

Az R^2 mutató értéke 0,289, vagyis a modell a függő változóban lévő variancia 28,9%-át képes megmagyarázni, pontosabban, az egyes kitöltők online kozmetikumvásárlási gyakoriságában lévő különbségek közel 30%-át vagyunk képesek megmagyarázni a modellben szereplő 19 magyarázó változóval, míg 70%-át más tényezők befolyásolják.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.537 ^a	.289	.189	2.17705

Miután meggyőződünk a modell alkalmazhatóságáról, azután érdemes megnézni a becsült paramétereket. Elsőként a paraméterek szignifikanciáját szükséges ellenőrizni a t -teszt alapján.

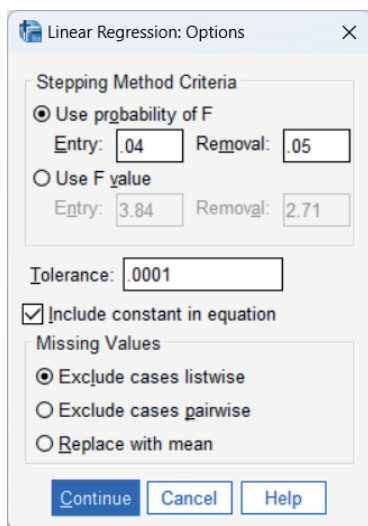
Coefficients^a

Model		Unstandardized Coefficients		Standardized	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	-.456	4.338		-.105	.916			
	D1_ferfi D1=Férfi	.110	.445	.022	.246	.806	.053	.021	.018
	D4_sajat D4=Saját, vagy szülőkrőlkonok tulajdonában lévő ingatlanban lakom	.167	.512	.030	.327	.744	-.012	.028	.024
	D4_kollegium D4=Kollégiumban lakom	.433	.504	.074	.860	.391	.025	.074	.062
	D4_bejar D4=Naponta bejárok Budapestre	.885	.544	.145	1.627	.106	.121	.138	.118
	D4_egyebd D4=Egyéb, éspedig:....	2.368	1.248	.173	1.897	.060	.103	.161	.137
	D5_resz D5=Igen, részmunkaidőben	1.060	1.405	.171	.754	.452	-.040	.065	.055
	D5_alkalmi D5=Igen, alkalmi munkákat vállalok	.554	1.428	.098	.388	.699	-.074	.033	.028
	D5_nem D5=Nem dolgozom	1.308	1.415	.270	.924	.357	.121	.079	.067
	D5_egyebd D5=Egyéb, éspedig	1.532	2.868	.051	.534	.594	.059	.046	.039
	D7 Életkor	-.006	.179	-.003	-.035	.972	.019	-.003	-.003
	Q18_02 Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Kosmetikumok esetén minden esetben megnézem a termékek összetételét.	.206	.105	.171	1.958	.052	.207	.166	.142
	Q18_04 Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Csomagolásmentesen vásárlók.	.246	.135	.154	1.819	.071	.086	.154	.132
	Q18_05 Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Nem vásárlók olyan kozmetikumokat, amelyeket állatokon teszteltek.	-.101	.111	-.088	-.912	.363	.005	-.078	-.066
	Q18_07 Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Ha van választási lehetőség, mindig azt a terméket választom, amely a legkisebb mértékben járul hozzá a környezeti ártalmakhoz.	-.044	.135	-.028	-.325	.746	.026	-.028	-.023
	Q18_15 Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Hajlandó vagyok többet fizetni a környezetbarát termékekért akkor is, ha van olcsóbb alternatíva.	-.247	.214	-.156	-1.154	.251	.050	-.098	-.083
	Q18_16 Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Hajlandó vagyok többet fizetni a társadalmilag felelős termékekért akkor is, ha van olcsóbb alternatíva.	.197	.218	.120	.904	.367	.117	.077	.065
	D6_2 Mekkora összeget költesz átlagosan havonta az alábbi tételekre? - Étkezés, élelmiszerek (Fűthő)	-1.867E-6	.000	-.017	-.182	.855	-.025	-.016	-.013
	D6_4 Mekkora összeget költesz átlagosan havonta az alábbi tételekre? - Szórakozás - (Fűthő)	1.089E-5	.000	.065	.673	.502	.003	.058	.049
	Q08_2 Hány alkalommal vásároltál online az elmúlt 1 hónap során? - Kosmetikum	1.401	.261	.414	5.365	<.001	.422	.418	.388

a. Dependent Variable: Q08_2 Milyen gyakran szoktad az adott termék kategóriát online vásárolni (10 alkalomból hányszor)? - Kosmetikumok

A regressziós outputtábla alapján a modellben szereplő magyarázó változók legtöbbje nem szignifikáns a legtöbbször alkalmazott 5%-os vagy 1%-os szignifikanciaszinten, mindössze egyetlen magyarázó lett szignifikáns, a Qo8_2.

A nem szignifikáns magyarázó változókat ilyenkor érdemes eltávolítani a modelltől. Azonban a magyarázó változók közötti kapcsolatok miatt elképzelhető, hogy egy magyarázó változó kivétele után a többi magyarázó változó hibátagja, így pedig szignifikanciája is módosulni fog. Ezért egyesével érdemes eltávolítani az inszignifikáns magyarázó változókat, a legkevésbé szignifikáns (legnagyobb p -értékkel rendelkező) magyarázó változóval kezdve. Ez az eljárás manuálisan is megvalósítható, de SPSS-ben a **Linear Regression** felugróablakban a *Method* résznél a *Backward* opciót választva ezt automatikusan is elvégzi a program. Ilyenkor érdemes még az *Options* menüpont alatt a *Stepping Method Criteria*-nál beállítani, hogy milyen szignifikanciaszinten lévő változókat szedjen még ki a program a modelltől. Ez általában 5% szokott lenni, tehát a *Removal* résznél 0,05-öt érdemes írni. Ezt viszont csak akkor fogadja el a program, ha az *Entry* résznél ennél alacsonyabb érték szerepel. Mivel az *Entry* a mostani elemzésnél nem játszik szerepet, így ide bármilyen 0 és 0,05 közötti szám megfelelő lesz (az alábbi példában 0,04 került beírásra).



A program 19 lépésben léptette ki a modellben felesleges magyarázó változókat, és végül csak egyetlen változó maradt meg, a Qo8_2. Jelen esetben tehát a *backward* elimináció eredményeképpen is ugyanazt a modellt kaptuk meg, mint amit akkor kaptunk volna, ha az összes inszignifikáns magyarázó változót azonnal eltávolítjuk a modelltől, de ez nem feltétlenül lesz igaz minden esetben.

Model	Coefficients ^a								
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			
	B	Std. Error	Beta			Zero-order	Partial	Part	
19	(Constant)	1.743	.196		8.907	<.001			
	Q08_2 Hány alkalommal vásároltál online az elmúlt 1 hónap során? - Kosmetikum	1.427	.247	.422	5.780	<.001	.422	.422	.422

a. Dependent Variable: Q06_2 Milyen gyakran szoktad az adott termékkategóriát online vásárolni (10 alkalomból hányszor)? - Kosmetikumok

A végső modell magyarázóereje már „mindössze” 17,8%, azonban ez önmagában nem probléma. Az R^2 mutató egy új változó modellbe való bevétele esetén csak nőhet, míg egy változó modellből való kivétele esetén minden esetben csökken. Önmagában tehát az az információ, hogy egy változó betétele/eltávolítása növelte/csökkentette a modell magyarázóerejét, még nem elégséges indok a bővebb modell választása mellett. Ennek oka, hogy még egy elméleti szempontból teljesen irreleváns változó is kis mértékben növelheti az R^2 mutatót, a kérdés tehát az, hogy az adott változó egyénileg szignifikáns-e a modellben.

A végső, immár mindössze egyetlen magyarázó változót tartalmazó modell a globális F -teszt alapján szintén érdemi magyarázó erővel bír ($p < 0,001$).

A modellben szereplő egyetlen magyarázó változóhoz tartozó becült paramétert úgy lehet értelmezni, hogy ha valaki az elmúlt egy hónapban minden más változatlansága mellett eggyel több alkalommal vásárolt kozmetikumot online, akkor várhatóan 10 kozmetikumvásárlásából 1,427 alkalommal többször fog online kozmetikumot vásárolni. Tehát akik gyakrabban vásároltak kozmetikumot online a kérdőív felvételét megelőző hónapban, azok arányaiban is többször vásárolnak kozmetikumot online, mint azok, akik ritkábban vásároltak kozmetikumokat online a kérdőív felvételét megelőző hónapban. Ez logikus, hiszen arra utal, hogy aki a közelmúltban vásárolt online kozmetikumot, az jobban preferálja az online csatornát, mint az, aki nem vagy ritkábban vásárolt kozmetikumot online a közelmúltban.

A konstans jelen esetben értelmezhető (nem mindig az), tehát érdemes erre is kitérni még. A konstans azt fejezi ki, hogy ha mindegyik magyarázó változó értéke zérus, akkor mennyi a függő változó várható értéke. Jelen esetben ez azt jelenti, hogy ha valaki nem vásárolt az elmúlt egy hónapban kozmetikumot online, akkor ő 10 kozmetikumvásárlásából 1,743 alkalommal vásárol kozmetikumot online. A minta 75%-a nem vásárolt a kérdőív felvételét megelőző hónapban kozmetikumot online, tehát ezek a fogyasztók alapvetően a fizikai boltokat részesítik előnyben kozmetikum vásárlásakor, esetükben az online kozmetikumvásárlás meglehetősen ritkának tekinthető, de előfordul, 10-ből egy-két esetben. (A minta csak azokat tartalmazza, akik a kérdőív lekérdezését megelőző egy évben vásároltak kozmetikumot.)

A kozmetikumok online vásárlási intenzitásánál tehát nagyon fontos tényezőnek bizonyult a közelmúltbeli konkrét online vásárlások száma, lényegében a

bevonódás (involvement). A termék kategóriát gyakrabban online vásárlók arányosan is többször vásárolnak online.

Emellett pedig rengeteg egyéb tényező határozza meg, hogy ki milyen arányban vásárol kozmetikumokat online, hiszen a variancia több mint 80%-át nem tudtuk megmagyarázni. Ez önmagában nem probléma, azonban utalhat arra, hogy a modell endogenitási torzítással terhelt, hiszen kimaradhattak olyan magyarázó változók is a modellből, amelyek egyszerre hatnak az online vásárlási arányra és a közelmúltbeli online vásárlások gyakoriságára, például a vásárlási kedv, a vásárlás, mint tevékenység szeretete. Amennyiben ez negatívan függ össze azzal, hogy ki milyen sokat vásárol online, illetve vásárlásai mekkora részét végzi online, akkor felfelé torzítja a becsült paramétert, tehát a valós értéknél nagyobb értéket kaphattunk. Ezt a kérdést azonban a jelen adatbázis alapján nem tudjuk eldönteni, ehhez elsősorban bővebb adatbázis, illetve magasabb szintű regressziós módszertan szükségeltetik, ami nem része jelen könyvnek.

A modell alapján viszont érdekes eredmény az is, hogy milyen tényezők nem befolyásolják azt, hogy ki milyen gyakran vásárol kozmetikumokat online. Sem a vizsgált demográfiai tényezők (nem, életkor, lakhely, munkaerőpiaci státusz, lakhatás, jövedelem), sem a kozmetikumokkal kapcsolatos attitűdök nem hatnak érdemben a függő változóra.

Végül, példaként értelmezzünk egy dummy változót is. Mivel ilyen változó végül nem lett szignifikáns a végső modellben, ezért a kiinduló modell D_4 kollegium változójának becsült paraméterét értelmezzük gyakorlásképpen. Ennek értéke 0,433 (p -értéke 0,391, tehát még 10%-os szignifikanciaszinten sem szignifikáns). A pontbecslést úgy lehet értelmezni, hogy ha valaki minden más változatlanúsága mellett kollégiumban lakik, akkor ő 10 kozmetikumvásárlásából várhatóan 0,433 alkalommal többször vásárol online, mint az, aki Budapesten albérletben lakik (utóbbiak alkották a referenciacsoportot). A kollégisták nagyjából 24 vásárlásonként vásárolnak eggyel többször online kozmetikumot, mint a Budapesten albérletben lakók. Tehát a kollégisták valamivel gyakrabban vásároltak kozmetikumokat online, bár még egyszer fontos megjegyezni, hogy az eltérés statisztikailag nem szignifikáns.

Szakirodalmi példa

A lineáris regresszió – ahogy azt korábban is említettük – nagyon elterjedt akadémiai kutatások során. A fejezet zárásaként röviden bemutatunk és értelmezzünk egy érdekes kutatást, *Melissa S. Kearney és Phillip B. Levine „Media Influences on Social Outcomes: The Impact of MTV’s 16 and Pregnant on Teen Childbearing”* című cikkét, amely az *American Economic Review*-ben jelent meg 2015-ben (105. évfolyam, 12. szám, 3597–3632. oldal).¹

¹ <https://www.doi.org/10.1257/aer.20140012>

A cikk azt vizsgálja, hogy az MTV-n futó „16 and Pregnant” című showműsor hogyan hatott a fiatalkori gyermekvállalásra. Lényegében a kutatási kérdés az volt, hogy ha a fiatalok látják, milyen nehézségekkel jár egy tinédzserkori terhesség, akkor jobban odafigyelnek-e rá, hogy elkerüljék azt. A kutatás során azt használták ki a szerzők, hogy a televízióműsor nézettsége az USA-n belül eltérő volt, így várható hatása is eltérő kellett, hogy legyen. Amennyiben tényleg csökkenti a fiatalkori terhességet a műsor megtekintése, akkor ennek a csökkenésnek nagyobbak kell lennie azokon a területeken, ahol többen nézték a műsort. A műsor nézettségét a 12 és 24 évesek közötti Nielsen ratinggel mérték, ami azt mutatja meg, hogy a célközönség (tehát a 12–24 évesek) hány százaléka látta a műsort.

A cikkben többféle lineáris regressziós modellt is megbecsültek a szerzők, ezek közül az alábbi táblázat mutatja a fő eredményeket. A fejezetben tárgyalt lineáris regresszió az OLS oszlopban található, a többi oszlopban már összetettebb (instrumentális változós) regressziós modellek eredményei láthatók, amelyekkel a cikk szerzőinek célja éppen a potenciális endogenitási torzítás kiszűrése volt, de ezekkel jelen könyvben már nem foglalkozunk. Akit érdekelnek ezek a modellek, azoknak javasoljuk a fejezet elején felsorolt további módszertani irodalmak tanulmányozását.

TABLE 1—ESTIMATES OF THE IMPACT OF 16 AND PREGNANT RATINGS ON TEEN BIRTH RATES

	OLS (1)	First stage (2)	IV (3)	Reduced form (4)
Dependent variable:	ln(birth rate)	16 and Pregnant ratings	ln(birth rate)	ln(birth rate)
16 and Pregnant ratings	−1.020* (0.552)		−2.368** (0.942)	
MTV ratings 2008–2009		1.511*** (0.204)		−3.581** (1.512)
Unemployment rate	−1.440*** (0.401)	−0.001 (0.026)	−1.487*** (0.375)	−1.485*** (0.432)
F-statistic on omitted instrument		48.1		

Notes: The birth data used for this analysis represents quarterly birth rates by DMA for conceptions leading to live births between 2005 and 2010. The sample size in each model is 4,919 (205 DMAs, 24 quarters, and one observation was dropped because there were no teen births). Coefficients and standard errors (reported in parentheses) in birth rate regressions are multiplied by 100. Each model also includes the percentage of a DMA’s female teen that is Hispanic and non-Hispanic black along with quarter and DMA × season fixed effects. Regressions are weighted by the relevant sample sizes for each outcome. Reported standard errors are clustered at the DMA level.

A regressziós eredmények azt mutatják, hogy ha a „16 and Pregnant” című műsort az adott földrajzi területen élő 12–24 éves célcsoport 1%-pontnyival nagyobb aránya nézte meg minden más változatlansága mellett, akkor várhatóan kb. 1%-kal csökkent a 15–19 éves korosztályon belül a születési arány. A hatás 10%-on szignifikáns (de 5%-on nem), ezt fejezi ki a becslést követően lát-

ható egy csillag (*). Az endogenitási torzítást kezelő becslés (3. oszlop) több mint kétszer ekkora hatást mutat, ami ráadásul már 5%-on is szingifikáns.

A szerzők eredményei tehát igazolják, hogy amennyiben egy adott területen többen nézték a „16 and Pregnant” című műsort, akkor ott jobban csökkent a fiatalok terhesség, ami kedvező (egyben meglehetősen érdekes) eredmény. Ugyanakkor ebből még nem feltétlenül következik, hogy a csökkenés oka a showműsor megtekintése volt.

A szerzők az ok-okozati kapcsolat igazolása érdekében megvizsgálták a Google Trends és a Twitter adatai alapján, hogy a fiatalok mennyire érdeklődtek a különböző fogamzásgátlási módszerek és az abortusz iránt, és tapasztalható volt-e növekvő érdeklődés abban az időszakban, amikor a „16 and Pregnant” műsoron volt. Az ezt vizsgáló regressziós eredmények nagyrészt igazolták a feltevéseket, ami alátámasztja, hogy valóban sikerült ok-okozati kapcsolatot identifikálniuk a szerzőknek a showműsor nézettsége és a fiatalok terhessége visszaesése között. A két változó közötti kapcsolatot a fogamzásgátlás és az abortusz növekvő ismertsége eredményezte, amit a showműsor váltott ki.

A cikk bemutatása több ok miatt is tanulságos. Egyrészt, képet kaphatunk arról, hogy milyen változatos kutatási kérdések megválaszolására alkalmas a lineáris regresszió, illetve az arra épülő szofisztikáltabb regressziós módszerek.

Másrészt, látható, hogy a legegyszerűbb lineáris regressziós modell sok esetben torzított, és ilyenkor nem elégedhetünk meg ennek az eredményeivel, szofisztikáltabb módszertanra van sokszor (bár nem mindig) szükség. Ez elővigyázatosságra int minket a regressziós eredmények értelmezésekor, azok kész tényként való kezelésével szemben.

Harmadrészt, felhívja a figyelmet arra, hogy mennyire nehéz és összetett dolog ok-okozati kapcsolatot becsülni. Ehhez nem elegendő, ha a regresszióban szereplő magyarázó változó hat a függő változóra, ez pusztán egy korrelációs együttmozgást mutat, ok-okozati kapcsolatokhoz ennél többre van szükség. Ezt is érdemes észben tartani egy-egy regressziós eredmény értelmezésekor.

Gyakorló feladat

Egy induló ruházati webshop árazási menedzsereként az a feladata, hogy meghatározza a kiszállítási díj mértékét. Vizsgálja meg, hogy milyen tényezőktől függ, hogy egy átlagos méretű csomag kiszállításáért mekkora díjat tartanak méltányosnak a vásárlók, ezek alapján – a modell korlátait is figyelembe véve – pedig tegyen javaslatot a webshop által alkalmazandó kiszállítási díj mértékére.

09 LOGISZTIKUS REGRESSZIÓ

Az előző fejezetben tárgyaltuk a regresszió alapjait, és már akkor említésre került a módszer egyik előnyeként, hogy nem csak metrikus, hanem kategorikus változókat is be lehet vonni az elemzésbe. Az előző fejezetben azonban csak arra az esetre tértünk ki, amikor a magyarázó változók között szerepel kategorikus változó. Ennek értelmezését is megadtuk a gyakorlati példa során.

Ebben a fejezet azt mutatjuk be, hogyan kell alkalmazni a regressziós eszköztárat olyan esetekben, amikor a függő változó nem metrikus. A nem metrikus függő változók lehetnek nominális és ordinális szinten mértek, illetve felvehetnek kettő vagy több lehetséges értéket is. Magyarázó változóként szerepeltetve egy kategorikus változót a mérési szint (nominális vagy ordinális), illetve a felvett kategóriák száma nem jelent érdemi problémát, ahogy ezt az előző fejezet gyakorlati példája is illusztrálta. Azonban amikor a függő változó lesz kategorikus, akkor már más a helyzet. Jelen könyv keretei között csak azt az esetet tárgyaljuk, amikor a függő változó kétféle értéket vesz fel (ún. dummy változó). Kettőnél több értéket felvevő kategorikus változó is szerepeltethető függő változóként egy regresszióban, erre szolgálnak – többek között – az ordinális (rendezett) logit, a nested logit és a multinomiális logit becslőfüggvények, amelyekről viszont ebben a könyvben nem ejtünk szót.

Kétértékű kategorikus változók esetén a konkrét számszaki értékek nem bírnak jelentőséggel, hiszen a két értéket csak arra használjuk, hogy elkülönítsük egymástól a különböző kimeneteket. Éppen ezért ebben a fejezetben végig azt feltételezzük, hogy a függő változóként szereplő kategorikus változó 0 és 1 kódolással rendelkezik. Amennyiben ez nem teljesül, akkor érdemes a változót ennek megfelelően átkódolni (SPSS-ben a **Transform / Recode into Different Variables** paranccsal).

Lineáris valószínűségi modell

Attól függetlenül, hogy a függő változó két diszkrét értéket vesz (vehet) csak fel, az előző fejezetben bemutatott lineáris regresszió lefuttatható, matematikailag ugyanis semmi sem gátolja meg azt, hogy a becsült paramétereket megkapjuk. Ha átgondoljuk a regresszió előző fejezetben ismerttetett hat előfeltevését, azok közül egyik sem követeli meg, hogy a függő változó kettőnél több értéket vegyen fel.

Azt az esetet, amikor a standard lineáris regressziós eszköztárat használjuk kétértékű függő változóval, lineáris valószínűségi modellnek nevezzük. Ennek

oka, hogy a végeredmény lényegében egy valószínűséget ad meg arra vonatkozóan, hogy a kettő közül vélhetően melyik kimenetel fog megvalósulni. Amennyiben a valószínűség alacsony, akkor a 0-val kódolt kimenet bekövetkezése a valószínűbb, míg magas valószínűség esetén az 1-gyel kódolt kimenet bekövetkezése az esélyesebb.

A lineáris valószínűségi modell torzítatlanságát ugyanaz a négy feltétel biztosítja, amit a lineáris regressziónál bemutatunk. Viszont a konstrukciójából következően a modell heteroszkedasztikus lesz, tehát mindenképpen korrigálni kell a standard hibákat (ezt a feltételt nem is szükséges ellenőrizni). Ezt leszámítva azonban más eltérés nem tapasztalható, az előző fejezetben leírtak tehát irányadók a lineáris valószínűségi modell esetén is.

A lineáris valószínűségi modell által becsült paraméterek mindig azt mutatják meg, hogy minden más változatlansága mellett az adott magyarázó változó értékének egy egységgel történő növekedése várhatóan hogyan befolyásolja (hány százalékponttal növeli meg) az 1-gyel kódolt kimenet bekövetkezési valószínűségét. Éppen ez a lineáris valószínűségi modell legnagyobb előnye, hogy feltevései és értelmezése is egyszerű, lényegében a standard lineáris regressziós modellel azonos.

Azonban problematikus, hogy a becslési konstrukcióból következően semmi sem gátolja meg, hogy 0-nál kisebb vagy 1-nél nagyobb becsült valószínűségeket kapjunk. Emellett a modell teljesen lineáris, vagyis egy magyarázó változó növekedésének vagy csökkenésének hatása ugyanakkora függetlenül attól, hogy mi az 1-gyel kódolt kimenet éppen aktuális bekövetkezési valószínűsége. Ez pedig sokszor nem helytálló, ahogy a következő példa is mutatja.

Vizsgáljuk meg egy termék (mondjuk növényi tejszín) vásárlási hajlandóságát, vagyis azt, hogy mitől függ, hogy egy adott fogyasztó megvásárolja-e az adott terméket vagy sem. Mivel a növényi alapú tejszín nagyjából 75%-kal alacsonyabb környezeti lábnyommal rendelkezik, mint a hagyományos tehéntejből készült tejszín, így a környezetvédelmi attitűd jelentősen befolyásolja a vásárlási hajlandóságot. Vizsgáljuk most kizárólag ennek a magyarázó változónak a hatását, a regressziós feltevésekkel összhangban, minden más változatlansága mellett.

A környezetvédelmi attitűd alacsony szintjéről elmozdulva a vásárlási hajlandóság jelentősen növekszik, azonban, egy szint után a kapcsolat már gyengül, hiszen, ha a vásárlási hajlandóság közelíti a 100%-ot, akkor a környezetvédelmi attitűd további növekedésének szerepe már egyre kevésbé képes tovább növelni a vásárlási hajlandóságot, egyszerűen abból kiindulva, hogy annak van egy felső korlátja (amikor a fogyasztó már biztosan megvásárolja az adott terméket, tehát az értéke 1). Ugyanez a logika igaz a másik irányba is. Ez éppen az a nem-linearitás, ami abból következik, hogy a felső (vagy az alsó) határhoz köze-

Ílve az egyes magyarázó változók parciális hatásai kényyszerűen egyre kisebbek kell, hogy legyenek. Ezt pedig a lineáris valószínűségi modell nem képes kezelni.

Ez a fajta problematika egyébként előkerül a standard lineáris regressziós modellben is. Erre jó példa, ha azt nézzük, hogy az előző példában említett növényi tejszínből hány darabot vásárol a fogyasztó egy hónap alatt. Erre a jövedelemnek, mint magyarázó változónak nyilván van hatása. Alacsony jövedelmek esetén a jövedelem növekedése meg tud jelenni többletfogyasztásban, bizonyos jövedelmi szint fölött azonban ez már nem igaz, egyszerűen azért, mert a fogyasztó elérte a telítődési pontját, adott volumennél többet nem szeretne vásárolni, vásárlását már nem a pénzügyi lehetőségei korlátozták a jövedelemnövekedést megelőzően sem. Ezt a problémát sokszor a változók logaritmizálásával kezelik, ami megszünteti a változók közötti lineáris kapcsolatot, mégis a logaritmizált változó modellben való szerepeltetésével továbbra is egy lineáris modellkeretben maradunk. Hasonló megoldás lehetséges a lineáris valószínűségi modell esetén is.

Logisztikus regresszió

A lineáris valószínűségi modell előző alfejezetben bemutatott problémáira a logisztikus regresszió adja meg (legalább részben) a megoldást. Ez meglehetősen hasonlóan működik, mint a lineáris regresszió, csak egy, a logisztikus eloszlásfüggvénnyel történő transzformációt iktatunk be, pontosabban:

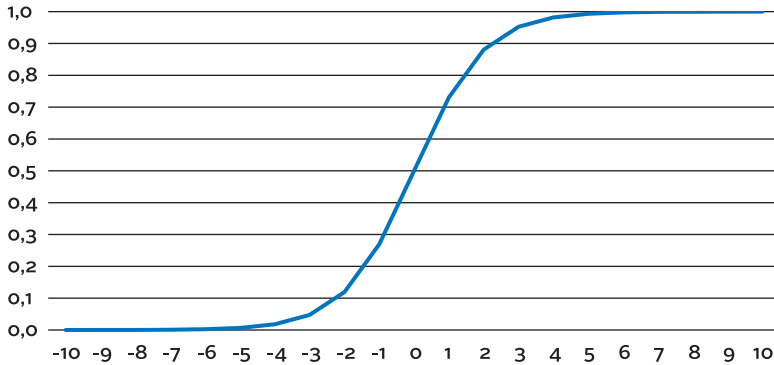
$$y = \Pr(y = 1|x) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u), \quad (9.1)$$

ahol $G(\cdot)$ a logisztikus eloszlásfüggvény:

$$G(z) = \frac{\exp(z)}{1 + \exp(z)}.$$

Mint minden eloszlásfüggvény, a logisztikus eloszlásfüggvény is 0 és 1 közötti értékeket vehet fel, z -ben szigorúan monoton növekvő, alakját a 9.1. ábra mutatja.

9.1. ábra: A logisztikus eloszlásfüggvény



Itt jegyezzük meg, hogy a G helyett igazából bármilyen más eloszlásfüggvény is alkalmazható lehet, gyakori még a standard normális eloszlásfüggvény használata, amit probit modellnek hívunk.

A 9.1. ábra alapján jól látható, hogy logisztikus regresszió esetén a becült valószínűség minden esetben 0 és 1 közé esik, illetve a magyarázó változóban bekövetkező változások okozta parciális valószínűségváltozások valójában függenek a kiinduló valószínűségtől. A magyarázó változó egységnyi növekedésének hatása eltérő lesz a függő változóra. Ha már eleve nagyon magas az 1-essel kódolt kimenet valószínűsége, akkor kisebb lesz a parciális növekedés, mintha egy alacsonyabb kiinduló valószínűségi szintről indultunk volna. Ez alapvetően kedvező tulajdonság, és a valóság adekvát leképezését adja, ahogyan azt az előző alfejezetben bemutatott rövid példa is illusztrálja.

Hátrány ugyanakkor, hogy a becült paraméterek nem értelmezhetők olyan közvetlenül és egyszerűen, mint a standard lineáris regresszió vagy a lineáris valószínűségi modell esetén. A becült paraméterek által jelzett hatás nagysága ugyanis attól függ, hogy az adott megfigyelés éppen hol helyezkedik el, mennyire van közel vagy távol a két vizsgált kimenethez (0-hoz vagy 1-hez). Erre még később részletesen visszatérünk.

A regresszióelemzésről szóló fejezet logikáját követve elsőnek a logisztikus regresszió esetén is a **modell egészének szignifikanciáját** vizsgáljuk meg. Erre szolgál a *log-likelihood* teszt, ami a globális F -teszthez hasonlóan azt a nullhipotézist teszteli, hogy a modell egyik magyarázó változója sem bír érdemi magyarázó erővel.

A modell globális tesztelésének hipotézisei:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1: \exists \beta_i \neq 0 (i=1, 2, \dots, k)$$

A lineáris regressziós modelltől eltérően a logisztikus regressziós modellt általában a legnagyobb esélyesség (*maximum likelihood, ML*) módszerrel becslik meg. Ez a *likelihood* függvény maximalizálását jelenti a magyarázó változók paraméterei segítségével. A becsült modell *likelihood* függvényének értékét pedig össze lehet vetni az üres modell, tehát a mindössze a konstansot tartalmazó modell *likelihood* függvényének értékével. Ez a *log-likelihood* teszt, amelynek tesztstatisztikája *khi-négyzet* eloszlást követ, ezért *khi-négyzet* tesztként is szoktak rá hivatkozni.

Amennyiben a *log-likelihood* teszt *p*-értéke magas (jellemzően 5% fölötti), akkor elfogadjuk a nullhipotézist, vagyis a modellnek nincs érdemi magyarázó ereje, főleges foglalkozni vele. A nullhipotézis elutasítása utal arra, hogy a modellel érdemes tovább dolgozni.

A modell teljesítményét emellett vizsgálhatjuk az előrejelzések pontossága alapján is, vagyis az alapján, hogy a modell hányszor állapította meg helyesen, hogy az adott megfigyelés esetén a függő változó értéke 0 vagy 1 lesz. Ezt hagyományosan egy klasszifikációs tábla segítségével lehet vizsgálni (9.1. táblázat) amelynek főátlójában vannak a helyesen klasszifikált megfigyelések (A és D), mellékátlójában pedig a helytelenül klasszifikáltak (B és C).

9.1. táblázat: Példa klasszifikációs táblára

Valós kimenet	Modell által becsült kimenet	
	O-való kódolt kimenet	1-gyel kódolt kimenet
O-való kódolt kimenet	(A)	(B)
1-gyel kódolt kimenet	(C)	(D)

A modell teljesítményét az mutatja meg, hogy a megfigyelések hány százalékát klasszifikálta helyesen, vagyis a 9.1. táblázat betűit használva (A) és (D) együttesen az összes megfigyelés mekkora hányadát teszi ki.

Mivel azonban a logisztikus regressziós modell nem kimeneteket, hanem valószínűségeket becsül csak, ezért szükséges egy határpont (*cut value*) megválasztása, ami alatti becsült valószínűség esetén a 0-val kódolt, míg az a fölötti becsült valószínűség esetén az 1-gyel kódolt kimenet bekövetkezését várjuk.

Ezt a határpontot a legtöbb esetben 50%-nál szoktuk meghatározni, de ettől természetesen el lehet térni.

A modell által létrehozott klasszifikációs táblát, és leginkább a helyesen klasszifikált megfigyelések arányát lehet viszonyítani ahhoz a kiinduló állapothoz, amikor minden megfigyelést a valószínűbb kimenethez sorolnánk. A modellünknek valójában ennél kell sokkal jobban működnie ahhoz, hogy elfogadható legyen.

Az egyes **magyarázó változók egyedi tesztelése** a lineáris regressziós modellnél bemutatottakkal analóg módon történik, azokat a változókat érdemes megtartani a modellben, amelyek legalább 5%-os szignifikanciaszinten szignifikánsnak bizonyulnak.

A **modell magyarázóerejét** szintén a modell által megmagyarázott variancia százalékában fejezzük ki, azonban ez logisztikus regressziók esetén nem triviális, hiszen a függő változó 0 vagy 1 értéket vehet csak fel, a regresszió eredményeként kapott értékek pedig 0 és 1 között szóródnak. Ezért többféle ún. pszeudó R^2 mutató létezik, ezek közül a leginkább elfogadott és legtöbbször használt a *Nagelkerke* R^2 , ami szintén százalékosan értelmezhető, tehát azt mutatja meg, hogy a modell a függő változó varianciájának hány százalékát képes megmagyarázni, mennyivel képes csökkenteni a függő változóban rejlő bizonytalanságot.

A **becsült paraméterek értelmezése** azonban összetettebb, mint lineáris regresszió esetén volt. Ahogy arról már korábban említést tettünk, a logisztikus transzformáció következtében a becsült paraméterek nem közvetlenül a függő változóra vonatkozó hatást mutatják. A becsült paraméterek előjele az egyetlen, ami ránézésre is értelmezhető. Pozitív előjel esetén az adott magyarázó változó növekedése növeli az 1-essel kódolt kimenet bekövetkezési valószínűségét, míg negatív előjel esetén az adott magyarázó változó növekedése csökkenti azt.

A logisztikus regresszió speciális abból a szempontból, hogy a becsült paramétereknek van konkrét jelentésük. Ez más hasonló modellek (pl. probit) esetében már nem igaz. A becsült paraméterek exponenciálisa ($\exp(\beta_i)$) az *odds* százalékos változását mutatja a magyarázó változó értékének egységnyi növekedésekor. Az *odds* a bekövetkezés és a nem-bekövetkezés valószínűségének hányadosa, matematikailag $p/(1-p)$, ahol p a függő változó 1-essel kódolt kimenetének bekövetkezési valószínűsége. Minél nagyobb tehát egy adott becsült paraméter exponenciálisa, a hozzá tartozó magyarázó változó értékének növekedése annál jobban növeli az *odds*ot.

Az *odds* értelmezése ugyanakkor nem feltétlenül triviális. Összességében igaz, hogy az *odds* növekedése az 1-essel jelölt kimenet bekövetkezési valószínűségének növekedését jelenti, de a kapcsolat nem lineáris. Sokszor pedig szükség van ennél intuitívabb magyarázatra is. Mivel azonban az adott magyarázó változó parciális hatása, vagyis függő változóra kifejtett hatása minden egyes megfi-

gyelés esetén eltérő, így a gyakorlatban az átlagos parciális hatást (*average partial effect, APE*) szokták meghatározni és értelmezni. Ennek értelmezése már analóg a lineáris regresszió esetén írottakkal, tehát az *odds*-nál sokkal kézzelfoghatóbb és egyszerűbb.

Az átlagos parciális hatás meghatározásához minden egyes megfigyelés esetében kiszámoljuk az adott magyarázó változó értéke 1-gyel való növekedésének hatását a kimenetre (függő változóra), majd ezeket átlagoljuk. Logisztikus regresszió esetén az i -edik magyarázó változó átlagos parciális hatását az alábbi képlet adja meg:

$$APE_i = \beta_i \frac{\sum g(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u)}{n}, \quad (9.2)$$

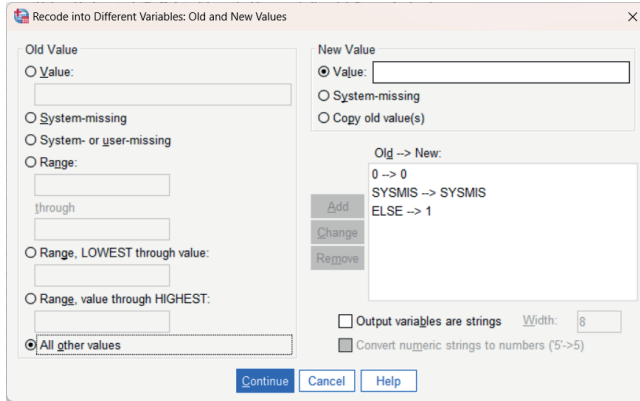
ahol $g(\cdot)$ a logisztikus sűrűségfüggvény, n pedig a mintaelemszám.

Az APE (illetve pontosabban annak százalékra felszorozott értéke) tehát úgy értelmezhető, hogy ha az adott magyarázó változó értéke minden más változatlansága mellett egy egységgel növekszik, akkor várhatóan mennyivel (hány százalékponttal) változik a függő változó 1-essel jelölt kimenetének bekövetkezési valószínűsége.

Gyakorlati példa

E rövid elméleti felvezetés és áttekintés után nézzük meg egy konkrét példán keresztül, hogy a lineáris valószínűségi modell és a logisztikus regresszió hogyan alkalmazható a gyakorlatban.

A gyakorlati példa során nézzük meg ugyanazt a kutatási kérdést, mint amit az előző fejezetben a regresszióelemzés során vizsgáltunk, annyi különbséggel, hogy most azt elemezzük, hogy mitől függ, hogy valaki vásárol-e kozmetikumot online vagy sem. Ezúttal tehát a Qo6_2 (*Milyen gyakran szoktad az adott termék kategóriát online vásárolni (10 alkalomból hányszor)? – Kozmetikumok*) változót átkódoljuk egy dummy változóvá, ami 1-et vesz fel, ha 10 alkalomból legalább egyszer online vásárolja a kozmetikumot az adott fogyasztó és 0-t, ha sosem vásárol kozmetikumot online. Az új változót a **Transform / Recode into Different Variables** paranccsal készítjük el, és Qo6_2_dummy néven mentjük el.



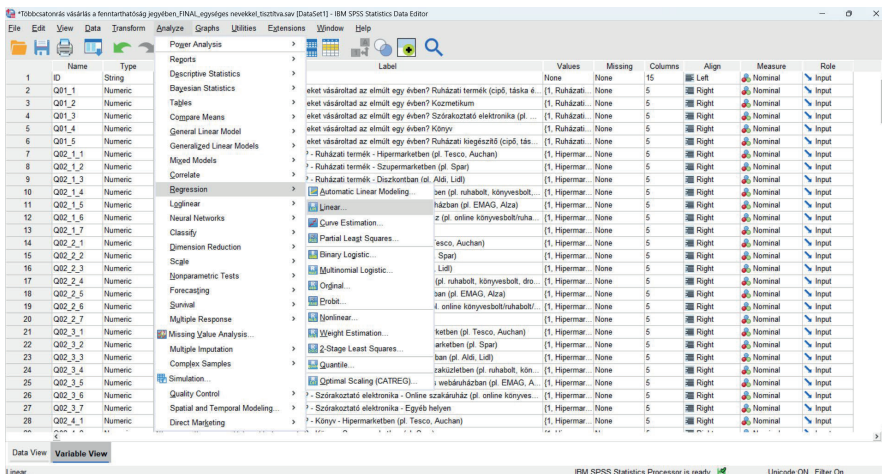
A modellek megbecslése előtt érdemes megnézni az új változó leíró statisztikáját az **Analyze / Descriptive Statistics / Frequencies** paranccsal. Ez alapján a minta 34,2%-a egyáltalán nem vásárol kozmetikumokat online, míg 65,8%-a valamilyen gyakorisággal igen.

Q06_2_dummy Vásárol-e kozmetikumokat online

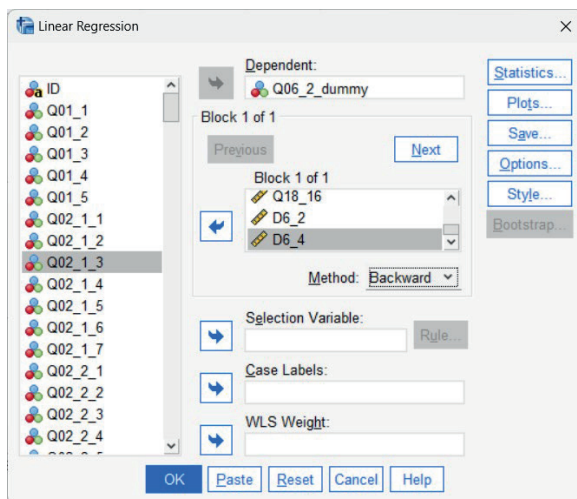
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00 Nem	88	34.2	34.2	34.2
	1.00 Igen	169	65.8	65.8	100.0
Total		257	100.0	100.0	

Az adatbázisban lévő változók közül elsősorban a demográfiai változókat (D1, D3, D4, D5, D7) vesszük figyelembe, emellett pedig néhány Likert-skálás kérdésre adott választ (Q18_02, Q18_04, Q18_05, Q18_07, Q18_15, Q18_16), valamint anyagi lehetőségeket mérő válaszokat (D6_2, D6_4). Mivel kizárólag azok érdekelnek minket, akik vásárolnak (az elmúlt egy évben vásároltak) kozmetikumokat, ezért ez bekerül szűrőfeltételnek (Q01_2 = 2). A változók bemutatása, illetve dummy változókká történő átalakítása az előző fejezet gyakorlati példájában került ismertetésre, most ugyanazokat a változókat fogjuk használni.

A lineáris valószínűségi modell becsléséhez a lineáris regresszió parancsot kell kiválasztani SPSS-ben: **Analyze / Regression / Linear**.



A felugró ablakban ki kell jelölni a függő (*Dependent*), illetve a magyarázó (*Independent(s)*) változókat. Jelen esetben a függő változó a Q06_2_dummy, míg a magyarázó változók: D1_ferfi, D3_Bp, D4_saját, D4_kollegium, D4_bejar, D4_egyebd, D5_resz, D5_alkalmi, D5_nem, D5_egyebd, D7, Q18_o2, Q18_o4, Q18_o5, Q18_o7, Q18_i5, Q18_i6, D6_2 és D6_4. A *Method* esetén érdemes rögtön kiválasztani a *Backward*-ot.



Lekért adatok a menüpontokból:

Statistics

- ✓ *Estimates*: regressziós együtthatók, paraméterek egyedi tesztelése – *t*-próbák
- ✓ *Model fit*: modell globális tesztelése – *F*-próba
- ✓ *Descriptives*: változók leíró statisztikái
- ✓ *Part and partial correlation*: multikollinearitás vizsgálatához

Options

- ✓ *Stepping Method Criteria / Use probability of F / Entry*: 0.04, *Removal*: 0.05

A modell lefuttatása után látható, hogy 14 lépésben jutottunk el a végső, már csak 5%-on szignifikáns magyarázó változókat tartalmazó modellig.

		Coefficients ^a							
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
14	(Constant)	.198	.100		1.994	.048			
	D4_egyebd D4=Egyéb, éspedig....	.442	.205	.164	2.155	.033	.128	.172	.160
	D5_alkalmi D5=Igen, alkalmi munkákat vállalok	.241	.104	.216	2.307	.022	.045	.183	.172
	D5_nem D5=Nem dolgozom	.229	.091	.242	2.520	.013	.130	.200	.187
	Q18_Q2 Mennyire jellemzők rád az alábbi állítások általában, amikor vásárolsz? - Kosmetikumok esetén minden esetben megnézem a termékek összetételét.	.069	.018	.295	3.902	<.001	.309	.301	.290

a. Dependent Variable: Q06_2_dummy Vásárol-e kozmetikumokat online

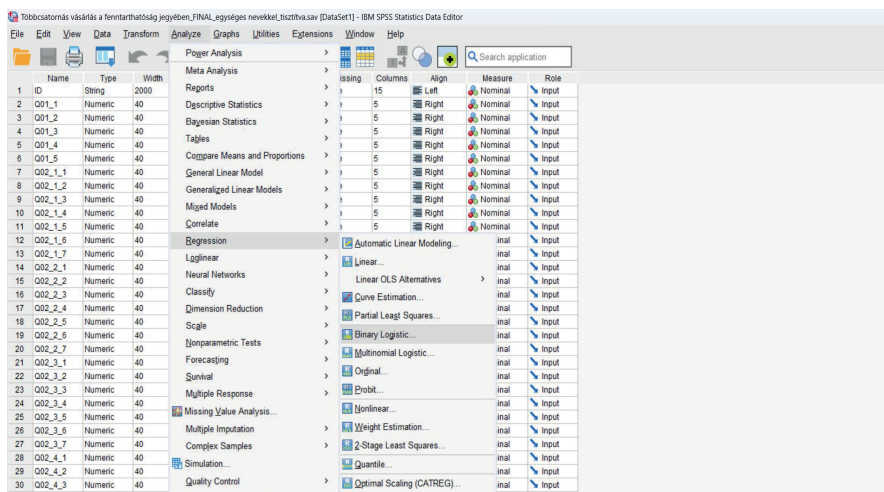
A végső modell egésze minden szokásos szignifikanciaszinten szignifikáns ($p < 0,001$), és az R^2 mutató értéke 0,153, vagyis a modell magyarázóereje 15,3%. A modellben négy magyarázó változó maradt, amelyeket a következőképpen lehet értelmezni:

- D4_egyebd: Amennyiben valaki alapvetően nem Budapesten lakik (az egyéb válaszok szöveges kifejtései erre utaltak), akkor ő minden más változatlansága mellett a Budapesten albérletben lakókhöz képest 44,2%-pontosan nagyobb valószínűséggel vásárol kozmetikumokat online (is).
- D5_alkalmi: Amennyiben valaki tanulmányai mellett alkalmi munkákat végez, akkor ő minden más változatlansága mellett a tanulmányaikkal mellett főállásban dolgozókhöz képest 24,1%-pontosan nagyobb valószínűséggel vásárol kozmetikumokat online (is).
- D5_nem: Amennyiben valaki tanulmányai mellett nem dolgozik, akkor ő minden más változatlansága mellett a tanulmányaikkal mellett főállásban dolgozókhöz képest 22,9%-pontosan nagyobb valószínűséggel vásárol kozmetikumokat online (is).

- Q18_o2: Amennyiben valakire az 1–7 Likert-skálán mérve 1-gyel nagyobb mértékben jellemző, hogy megnézi a kozmetikumok összetételét, akkor ő kb. 6,9%-pontos valószínűséggel vásárol kozmetikumokat online (is).

Az online (is) történő kozmetikumvásárlást tehát alapvetően azok választják, akik nem Budapesten élnek, nem végeznek tanulmányaik mellett rendszeres munkát, és szívesebben nézik meg a kozmetikumok összetételét.

Ezeket az eredményeket vessük most össze a logisztikus regresszió eredményeivel. Ehhez az **Analyze / Regression / Binary Logistic** parancsot kell választani.



A megjelenő felugróablakban kell megadni a modellben szereplő változókat. A függő változó (*Dependent*) ebben az esetben is az újonnan létrehozott Qo6_2_dummy, míg a magyarázó változók (*Covariates*) szintén a D1_ferfi, D3_Bp, D4_sajat, D4_kollegium, D4_bejar, D4_egyebd, D5_resz, D5_alkalmi, D5_nem, D5_egyebd, D7, Q18_o2, Q18_o4, Q18_o5, Q18_o7, Q18_o15, Q18_o16, D6_2 és D6_4. A Method esetén érdemes rögtön kiválasztani a *Backward: LR-t*, vagyis a *likelihood ratio* alapján történő *backward* eliminációt.

Lekért adatok a menüpontokból:

Categorical: Itt lehet kiválasztani azokat a változókat, amelyek kategorikusak. A logisztikus regresszió esetén nem lett volna szükség a dummy változók legenerálására, a program itt automatikusan elkészítette volna. Ettől független a kategorikus függő változókat itt ki lehet jelölni.

Options:

- ✓ *Probability for Stepwise: Entry:* 0.04, *Removal:* 0.05 (ezzel azt határozzuk meg, hogy *backward* elimináció során az 5%-on nem szignifikáns változókat akarjuk eltávolítani a modelltől)
- ✓ *Classification cutoff:* 0.5 (ezzel határozzuk meg, hogy a klasszifikációs tábla elkészítése során 50%-nál nagyobb becslt valószínűség esetén klasszifikálja 1-es kimenetűnek az adott megfigyelést a program)

A modell lefuttatása utáni output eltér a regresszió esetén tapasztalttól. Az első három táblázat a kategorikus változók számozását mutatja be, utána pedig a „*Block 0: Beginning Block*” című rész az ún. üres modellt, vagyis azt a modellt, amiben semmilyen magyarázó változó sem szerepel. Ez szolgál referenciamodellként, a logisztikus regresszió célja alapvetően egy ennél jobb modell építése. Az ebben szereplő klasszifikációs tábla azt mutatja, hogy ha minden esetben azt „tippeli” a modell, hogy online is vásárol kozmetikumot az adott kitöltő, akkor az 66,5%-os sikerrátával rendelkezik, hiszen a modellben a többség vásárol kozmetikumot online (ahogy azt a fentebbi leíró statisztikai táblázatban láttuk). A fenti leíró statisztikai tábla és az itt látható klasszifikációs tábla elemszáma eltér, ez azért van, mert néhány megfigyelés kiesett a mintából hiányos adatok miatt.

Classification Table^{a,b}

Observed		Predicted		Percentage Correct	
		Q06_2_dummy Vásárol-e kozmetikumokat online .00 Nem	1.00 Igen		
Step 0	Q06_2_dummy Vásárol-e kozmetikumokat online	.00 Nem	0	53	.0
		1.00 Igen	0	105	100.0
Overall Percentage					66.5

a. Constant is included in the model.

b. The cut value is .500

A regressziós output következő, „*Block 1: Method = Backward Stepwise (Likelihood Ratio)*” című blokkjában szerepel a magyarázó változókat is tartalmazó modell eredménye. Először is azt láthatjuk, hogy a modell szignifikánsan jobb-e, mint az előző blokkban bemutatott üres modell, vagyis volt-e értelme a modellépítésnek. Ez hasonló, mint a lineáris regresszió esetén a modell globális tesztelése, csak itt az ún. *omnibusz* tesztet használjuk, ami *khi-négyzet* alapú. A tesztek közül alapvetően a *Model* sor alatti az érdekes, ez hasonlítja össze az üres modellt és az általunk épített modellt. Jelen esetben a modellünk minden szokásos szignifikanciaszinten szignifikáns, hiszen a *p*-érték 0,001 alatti.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 15 ^a	Step	-2.952	1	.086
	Block	26.643	4	<.001
	Model	26.643	4	<.001

a. A negative Chi-squares value indicates that the Chi-squares value has decreased from the previous step.

A modell magyarázóereje a Nagelkerke R^2 mutató alapján 0,215, ami közepesnek tekinthető. Ez magasabb érték, mint amit a lineáris valószínűségi modellnél láttunk, bár a két modell ilyen szempontú összevetése az eltérő mutatószámok miatt nem lehetséges.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
15	174.954 ^a	.155	.215

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Szintén a modell teljesítményét méri a klasszifikációs táblában a helyesen prediktált megfigyelések aránya. Ez az üres modellben látott 66,5%-hoz képest 75,3%-ra nőtt, tehát közel 10 százalékponttal lett magasabb. Látható, hogy a modell nem teljesített túl jól azok körében, akik nem vásárolnak kozmetikumokat online. A klasszifikációs táblát – ahogy arra a lábjegyzete is utal – úgy állította elő a program, hogy 50%-nál nagyobb becslt valószínűség esetén klasszifikálta úgy az adott kitöltőt, hogy vásárol kozmetikumokat online.

Classification Table^a

Observed		Predicted		Percentage Correct	
		Q06_2_dummy Vásárol-e kozmetikumokat online .00 Nem	1.00 Igen		
Step 15	Q06_2_dummy Vásárol-e kozmetikumokat online	.00 Nem	22	31	41.5
		1.00 Igen	8	97	92.4
Overall Percentage					75.3

a. The cut value is .500

A modellben maradt magyarázó változók végül pontosan megegyeznek azokkal, amelyek a lineáris valószínűségi modell alapján érdemi módon befolyásolják azt, hogy valaki vásárol-e online kozmetikumokat: Budapesten kívüli lak-

hely, egyetemi tanulmányok melletti alkalmi munkavállalás vagy egyáltalán nem történő munkavállalás, illetve a termékek összetételének vizsgálata, átnézése. Mindegyik magyarázó változó hatása pozitív, tehát növelik a kozmetikumvásárlás valószínűségét.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 15 ^a						
D4=Egyéb, éspedig:....(1)	21.025	17358.999	.000	1	.999	1351825555.7
D5=Igen, alkalmi munkákat vállalok(1)	1.159	.540	4.603	1	.032	3.187
D5=Nem dolgozom(1)	1.074	.455	5.578	1	.018	2.927
Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Kozmetikumok esetén minden esetben megnézem a termékek összetételét.	.348	.097	12.994	1	<.001	1.416
Constant	-1.499	.515	8.492	1	.004	.223

a. Variable(s) entered on step 1: D1=Férfi, D4=Saját, vagy szülő(k)rokonok tulajdonában levő ingatlanban lakom, D4=Kollégiumban lakom, D4=Naponta bejárok Budapestre, D4=Egyéb, éspedig:.... D5=Igen, részmunkaidőben, D5=Igen, alkalmi munkákat vállalok, D5=Nem dolgozom, D5=Egyéb, éspedig, Életkor, Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Kozmetikumok esetén minden esetben megnézem a termékek összetételét, Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Csomagolásmentesen vásárolok, Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Nem vásárolok olyan kozmetikumokat, amelyeket állatokon teszteltek, Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Ha van választási lehetőség, mindig azt a terméket választom, amely a legkisebb mértékben járul hozzá a környezeti ártalmakhoz, Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Hajlandó vagyok többet fizetni a környezetbarát termékekért akkor is, ha van olcsóbb alternatíva, Mennyire jellemzőek rád az alábbi állítások általában, amikor vásárolsz? - Hajlandó vagyok többet fizetni a társadalmilag felelős termékekért akkor is, ha van olcsóbb alternatíva, Mekkora összeget költesz átlagosan havonta az alábbi tételekre? - Étkezés, élelmiszerek (Fűhó), Mekkora összeget költesz átlagosan havonta az alábbi tételekre? - Szórakozás - (Fűhó).

A becslt paraméterek exponenciálisa ($Exp(B)$) mutatja az *oddsra* kifejtett közvetlen hatást. A D4_egyebd változó esetén nagyon kevés volt az azt választó kitöltők száma, ez okozhat irreálisan nagy becslt hatásokat, ezért ennek az értelmezésétől eltekintünk. A többi becslt paraméter értelmezése az alábbi:

- Amennyiben valaki minden más változatlansága mellett egyetemi tanulmányai mellett alkalmi munkát vállal, úgy kb. 3,187-szer magasabb annak valószínűsége, hogy vásárol kozmetikumokat online, mint azé, hogy nem, azokhoz képest, akik egyetemi tanulmányaik mellett főállásban dolgoznak.
- Amennyiben valaki minden más változatlansága mellett egyetemi tanulmányai mellett nem vállal munkát, úgy kb. 2,927-szer magasabb annak valószínűsége, hogy vásárol kozmetikumokat online, mint azé, hogy nem, azokhoz képest, akik egyetemi tanulmányaik mellett főállásban dolgoznak.
- Amennyiben valakire az 1–7 Likert skálán eggyel nagyobb mértékben jellemző, hogy megnézi a kozmetikumok összetételét, esetében kb. 1,426-szor magasabb annak a valószínűsége, hogy vásárol kozmetikumokat online, mint azé, hogy nem.

Az *odds*ra kifejtett hatások kevésbé intuitívak, azonban adnak egyfajta értelmezést. Az APE sokkal intuitívabb, ráadásul értéke összevethető a lineáris valószínűségi modellben látott értékekkel, azonban az SPSS jelenleg ezt nem tudja közvetlenül kiszámítani, manuálisan kell elvégezni a számításokat, amennyiben valaki erre igényt tart. Ilyen esetekben egyébként a kutatók jellemzően más programcsomagok használatára térnek át, pl. R, Stata vagy EViews.

A becsült modellek alapján tehát az látható, hogy sikerült néhány olyan demográfiai és attitűdbeli tényezőt azonosítani, amelyek hozzájárulnak ahhoz, hogy valaki online (is) vásároljon kozmetikumokat. Ezzel együtt a modellek magyarázóereje, illetve a helyesen prediktált megfigyelések aránya nem túlságosan magas, tehát érdemes még továbbfejleszteni a modellt, ehhez viszont egy bővebb, jobban erre a kérdésre fókuszáló adatgyűjtésre van szükség. Ez az a probléma, amit a regressziós elemzés során sokszor említettünk már, hogy a megfelelő regressziós modell felépítésekor sokszor ütközünk adatkorlátokba, amelyek nehezítik az elemzést. Ezt logisztikus regresszió esetén is észben kell tartani.

Szakirodalmi példa

A logisztikus regresszió szintén gyakran alkalmazott módszertan akadémiai kutatások során. A fejezet végén *Elliot Fishman, Simon Washington, Narelle Haworth és Angela Watson „Factors influencing bike share membership: An analysis of Melbourne and Brisbane”* című cikkét tekintjük át röviden, amely a Transportation Research Part A-ben jelent meg 2015-ben (71. évfolyam, 17–30. oldal).²

A kutatás célja annak feltárása volt, hogy milyen tényezők befolyásolják azt, hogy valaki felhasználója lesz-e egy közösségi kerékpárrendszernek. A szerzők két ausztrál nagyvárost vizsgáltak, Melbourne-t and Brisbane-t. Az adatokat egy online kérdőíves felmérésen keresztül szerezték, amelyet a rendszert használók és nem használók is kitöltöttek.

A szerzők a kérdőíves felmérésből származó változókat felhasználva építettek fel logisztikus regressziós modelljüket, aminek függő változója az volt, hogy az adott egyén használja-e a közösségi kerékpárrendszert. A nem szignifikáns változókat eltávolítva a modelltől, az alábbi táblázat mutatja a végső eredményeket. A magyarázó változók referenciacsoportjait mutatja a táblázat második oszlopa. A becsült paraméterek értékét adja meg a negyedik oszlop, míg ennek szignifikanciáját az ötödik oszlop. Az *odds ratio* (6. oszlop) nem más, mint a becsült paraméterek exponenciálisa, ez mutatja tehát az adott magyarázó változó *odds*ra gyakorolt hatását.

2 <https://www.doi.org/10.1016/j.tra.2014.10.021>

Table 4
Logistic regression analysis.

Variables included in the model	Level/referent	z-Statistic	Coefficient	Sig.	Odds ratio	Odds ratio 95%CI
Impact of mandatory helmet legislation on riding (0 = does not reduce riding, 1 = reduces riding)	Does not reduce riding	5.85	3.20	<0.001	24.5	8.4–71.6
Riding activity in the past month (0 = no riding, 1 = riding)	No riding	32.33	1.75	0.020	5.8	1.3–25.3
Convenience as an encouraging factor for private bike riding (1 = not at all, 5 = a lot)	Not at all	2.29	0.65	0.022	1.9 [#]	1.09–3.33
Age (0 = 35 years and over, 1 = 18 = 34)	35 years and over	2.0	1.2	0.047	3.3	1.02–10.83
Income (1 = \$10,400 or less, 10 = \$104,000 or more)	<\$10,400	2.51	0.27	0.012	1.3 [#]	1.06–1.61
Work within 250 m of docking station (0 = no, 1 = yes)	No	7.18	3.40	<0.001	29.9	11.81–75.49
Constant		-8.76	-13.86	<0.001	9.56e-07	4.30e-08–0.0000212

[#] This value is cumulative, meaning that for each higher increment, the odds of being a bike share member increases by the Odds Ratio (OR).

Az eredmények alapján látható, hogy a kötelező bukósisakviselésre való negatív reakció, a megelőző hónapi aktív kerékpározás, a kerékpározás során a kényelem minél nagyobb fontossága, a 18 és 34 év közötti életkor, a magasabb jövedelem és a munkahelyhez közeli dokkoló állomás szignifikánsan növelte a közösségi kerékpárrendszer használatát. Ezek közül is a munkahelyhez közeli dokkolóállomás hatása volt a legnagyobb, amennyiben a munkahely 250m-es körzetében volt dokkolóállomás, az közel 30-szorosára növelte az *odds*ot, tehát a közösségi kerékpárrendszer használata és nem használata valószínűségének hányadosát.

A konkrét használati valószínűségre vonatkozó hatások bemutatásánál a szerzők nem az APE-t számolták ki vagy a klasszifikációs táblát készítették el, hanem azt vizsgálták meg, hogy egy-egy magyarázó változó értékeinek módosítása hogyan befolyásolja a használat valószínűségét, ha a többi magyarázó változó értéke éppen átlagos vagy egy meghatározott értéket vesz fel. Az alábbi grafikonon az látszik, hogy a kerékpározáskori kényelem fontossága növeli a közösségi kerékpárrendszerek használatának valószínűségét, illetve ez a valószínűség sokkal nagyobb, ha van a munkahely közelében dokkolóállomás, ha valaki kerékpározott a megelőző hónapban, illetve, ha magas az egyén jövedelme. (Az egyes színes vonalaknál a nem jelölt változók éppen az átlagos értéküket veszik fel.)

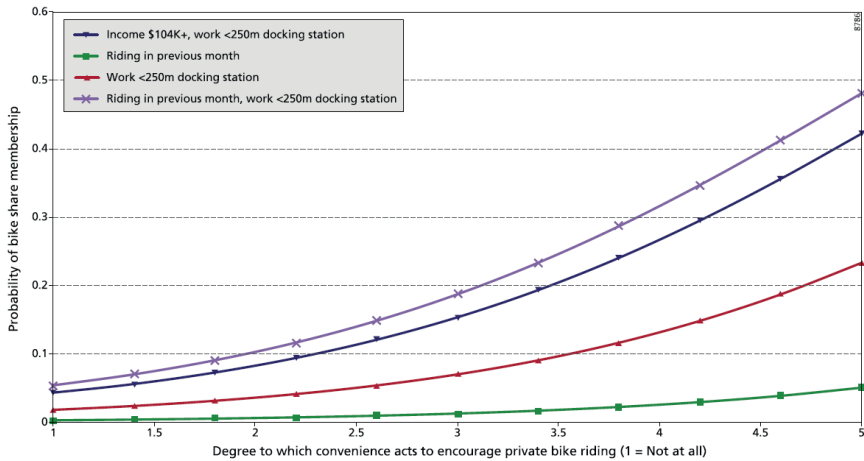


Fig. 8. Probability of bike share membership under different scenarios, with convenience extended through its range of means.

Bár összességében a becsült valószínűségek alacsonyak, a fenti ábra alapján jól láthatók az eltérések, tehát az egyes magyarázó változók hatásai. A logisztikus regresszió eredményeinek bemutatására tehát többféle lehetőség áll rendelkezésre, érdemes mindig azt választani, ami az adott kutatási kérdés szempontjából a leginkább releváns, vagy az adott környezetben a legjobban bemutatható, vizualizálható.

Gyakorló feladat

Egy induló ruházati webshop marketingmenedzsereként az a feladata, hogy meghatározza az üzlet célcsoportját. Ehhez első körben azt szeretné megvizsgálni, hogy milyen demográfiai és attitűd tényezők határozzák meg azt, hogy valaki vásárol-e ruhákat online. Végezze el az elemzéseket, és mutassa be a javasolt célcsoportot.

10 FAKTORELEMZÉS

Elméleti bevezetés

Eddig olyan módszerekkel foglalkoztunk, amelyek két vagy több változó közötti összefüggést vizsgáltak, oksági kapcsolat feltételezésével vagy anélkül. A változók közötti kapcsolat vizsgálatát megelőzi az a lépés, hogy mely változókat vonjunk be a vizsgálatba, melyek között tételezhetjük fel az esetleges összefüggést. Abban az esetben, ha a függő változó meghatározásához viszonylag kevés független változó elegendő és ezek könnyen meghatározhatók, akkor az összefüggésvizsgálat módszerei általában jól alkalmazhatók. A társadalomtudományokban és a marketingkutatásban is gyakran nem ez a helyzet. Az emberi magatartás vagy szűkebben tekintve a fogyasztói magatartás vizsgálatánál általában nagyszámú befolyásoló tényezőt kell figyelembe venni, amelyek gyakran egyenrangúak egymással olyan értelemben, hogy nem tudunk oksági kapcsolatot feltételezni vagy azt sem tudjuk, melyiket indokolt belevonni a vizsgálatba.

Könyvünkben az első olyan módszer, amely csak többváltozósaként működik, vagyis nincs kétváltozós opciója, a faktorelemzés vagy faktoranalízis. A faktorelemzés két nagy módszert jelenthet, a strukturafeltáró, exploratív módszert és a strukturát ellenőrző, konfirmatórikus módszert. Ez utóbbi a strukturális modellek módszerei közé sorolható, amelyben egy elméleti modell alapján felírt struktúra vizsgálata, tesztelése történik a konfirmatórikus faktorelemzési módszerrel, ezen módszer tárgyalása meghaladná a könyv kereteit.

Az exploratív faktorelemzés célja a kiinduló változók közötti strukturák feltárása.

A faktorelemzés olyan kölcsönös összefüggésen alapuló módszer, ahol nagyszámú kiinduló változót vonunk be az elemzésbe, amelyek között nincs függő vagy független változó. A faktorelemzést a következő célokkal végezzük:

- **Magyarázó tényezők, faktorok azonosítására**, amelyek az adott változók közötti korrelációt magyarázzák.
- **Kevesebb számú, korrelálatlan változó azonosítására**, amelyek az adott, korrelált változókat helyettesítik további többváltozós elemzésekben.
- Néhány kiemelkedően fontos **változó azonosítására**, amelyek később többváltozós elemzésekhez használhatók.

A faktorelemzés **látens**, nem közvetlenül mért változókat eredményez, melyet a **manifeszt** – megfigyelt, mért – változók határoznak meg. A fogyaszt-

tói magatartás kutatásakor gyakran kutatunk olyan fogalmakat, amelyeket nem tudunk közvetlenül, direkt módon megkérdezni hanem ezeket a fogalmakat a megkérdezésben „körülírjuk”. Ezeket a fogalmakat látens változónak nevezhetjük, amelyeket közvetlenül nem tekintünk mérhetőnek, mérésük ún. indikátorokkal, manifeszt változókkal történik. Ezek az indikátorok olyan állítások, amelyekkel leírunk adott magatartás típusokat, olyan módon, amit már meg lehet kérdezni, értékelteni lehet közvetlen módon. Ezeket az állításhalmazokat több-tételes skálának is nevezzük (multi item scale), amelyben a skála elemeit elméleti megfontolások, fogyasztói magatartás elméleti modellek alapján állíthatjuk össze. A szakirodalomban egy-egy modellt tesztelésére, empirikus kutatására gyakran találhatunk olyan validált skálákat, amelyeket már igazoltak, publikáltak, amelyek alkalmazásakor azonban a saját mérésünkre vonatkozó konzisztencia vizsgálatot el kell végeznünk.

A validált skála másik előállítási módja a saját skála fejlesztés, amelynek tárgyalásától eltekintünk már csak azért is, mert a skálafejlesztést a kutatások érett fázisában javasolja a szakirodalom.

Az exploratív faktorelemzés eredményeként létrehozott faktorok mesterséges, látens, **standardizált változók**, melyek jellemzői, hogy átlaguk=0 és szórásuk=1.

A faktorelemzés **feltétele**, hogy legalább háromszor annyi válaszadónk legyen, mint bevont változónk. A kiinduló változóinknak metrikusnak kell lennie és közöttük elvárt a viszonylag magas korreláció (multikollinearitás).

Ez teszi lehetővé a faktorelemzés fő céljának az elérését: csökkenteni a kiinduló változók számát az új, látens változók előállításával úgy, hogy közben a kiinduló változókban levő információtartalomtól a lehető legkevesebbet veszítsük el, vagyis az információtartalom nagy részét megtartsuk a változók számának csökkentés ellenére. A faktorelemzés során ezt a két, látszólag ellentmondó célt kell szem előtt tartani.

A faktorelemzés tipikus eljárását a következő lépések jellemzik:

1. A kiinduló változók körének meghatározása és a módszer által megkívánt feltételek ellenőrzése
2. A faktorelemzés módszerének kiválasztása, a faktorsúlyok meghatározása
3. A közös faktorok számának és a magyarázott varianciarányának a meghatározása
4. A faktorrotáció kiválasztása és a közös faktorok értelmezése
5. A faktorértékek meghatározása

A kiinduló metrikus változók közötti korreláció megvizsgálására előállíthatjuk a korrelációs mátrixot, amelyben mind az együtthatók nagyságát, mind a szignifikanciaszintet vizsgálnunk kell. A szakirodalom ajánl olyan mutatókat és pró-

bákat, amelyek a korrelációs mátrix elemeiből kerülnek kiszámításra és a javasolt küszöbértékek alapján könnyen eldönthető a faktorelemzésre való megfelelés.

A szakirodalom által a legalkalmasabbnak tartott döntési kritérium a Kaiser-Meyer-Olkin (KMO) kritérium, amely a korrelációs mátrixból kerül kiszámításra. Ez az érték nagyon megkönnyíti a döntést a kiinduló adatok alkalmasságáról, hiszen egy számított érték alapján eldönthető az alkalmasság. A mutató értéke 0 és 1 közé eshet, az alkalmassághoz legalább 0,5-nek kell lennie, minél magasabb, annál alkalmasabb a kiinduló adatbázis.

Hasonlóan megkönnyíti a döntést a Bartlett-próba, amelynek szignifikancia szintjének a könyvben tárgyalt többi próbához hasonlóan kisebbnek vagy egyenlőnek kell lennie 0,05-nél. Ez az esetek nagy részében teljesül, emiatt is ez a gyengébb kritérium, ezért is javasolt a KMO érték alapján történő döntés.

A faktorelemzés matematikai alapja szerint minden egyes kiinduló változó felírható új változók, azaz a faktorok lineáris kombinációjaként, ahol a faktorok közös és egyedi faktorokra oszthatók:

$$X_i = A_{i1} F_1 + A_{i2} F_2 + A_{i3} F_3 + \dots + A_{im} F_m + V_i U_i$$

ahol:

X_i = i-edik standardizált változó

A_{ij} = az i-edik változó j-edik közös faktorra vonatkozó többszörös standardizált parciális regressziós együtthatója

F = a közös faktor

V_i = az i-edik változó j-edik közös faktorra vonatkozó többszörös standardizált parciális regressziós együtthatója

U_i = az i-edik változó egyedi faktora

m = a közös faktorok száma

Az egyedi faktorok (U_i) korrelálatlanok egymással és a közös faktorokkal (F).

A közös faktorok is kifejezhetők a megfigyelt, kiindulási változók lineáris kombinációjaként.

$$F_i = W_{i1} X_1 + W_{i2} X_2 + W_{i3} X_3 + \dots + W_{ik} X_k$$

ahol:

F_i = az i-ik faktor becslése

W_i = súly vagy faktorérték együtthatója

k = a változók száma

A faktorelemzés során a következő legfontosabb fogalmakkal foglalkozunk:

- ✓ **Kommunalitás:** A variancia azon hányada, amelyet egy változó magyaráz. Ez egyben a közös faktorok által magyarázott variancia aránya is. *Minimum elvárt érték 0,25, amennyiben nem teljesül, az adott változót törölni kell a faktorelemzésből.*
- ✓ **Sajátérték:** A variancia azon hányada, amelyet egy-egy faktor magyaráz, vagyis azt mutatja, hogy egy faktor mennyit magyaráz az összes változó varianciájából.
- ✓ **Faktorsúly:** A változók és a faktorok közötti korrelációs együtthatók, ezek a lineáris kombináció paraméterei. A cél, hogy abszolút értékben minél nagyobb legyen. *Minimum elvárt értéke 0,4, amennyiben nem teljesül, az adott változót törölni kell a faktorelemzésből.*
- ✓ **Faktorsúly-ábra:** Az eredeti változók ábrázolása, ahol a faktorokat mint tengelyeket használjuk.
- ✓ **A faktor mátrix:** Az előállított faktoroknak az egyes megkérdezettekre vonatkozóan becsült értékei.
- ✓ **Kaiser-Meyer-Olkin (KMO)mutató:** A korrelációs mátrix elemeiből számítódik és azt méri, hogy az adatbázis alkalmas-e a faktorelemzésre. Ha az érték kisebb, mint 0,5, akkor az adatbázis alkalmatlan a faktorelemzésre. *Értékhatárok: $KMO \geq 0,9$ – kiváló, $KMO \geq 0,8$ – nagyon jó, $KMO \geq 0,7$ – megfelelő, $KMO \geq 0,6$ – közepes, $KMO \geq 0,5$ – gyenge, $KMO \leq 0,5$ – elfogadhatatlan*
- ✓ **Bartlett-féle gömbölyűségi teszt:** A faktorelemzés elvégezhetőségének próbája. *Ho: a megfigyelt változók korrelációs mátrixa egységmátrix, vagyis a változók páronként korrelálatlanok*
- ✓ **Varianciahányad:** a varianciának az a hányada, amelyet egy-egy faktor magyaráz.
- ✓ **Reziduum:** A megfigyelt, eredeti korrelációs mátrixban található korrelációs együtthatók és a reprodukált, azaz a faktormátrixból becsült korrelációs együtthatók közötti különbségek.
- ✓ **Sajátérték-ábra:** A sajátértékek ábrázolása az előállított faktorok sorszámanak függvényében.
- ✓ **Rotáció:** a faktorok tengelyeit elforgatjuk úgy, hogy könnyebben értelmezhetőek legyenek a faktorok.

A faktorelemzés elvégzése

A faktorelemzés több módszerrel is elvégezhető, a két leggyakrabban használt módszer a főkomponens elemzés és a főtengety elemzési módszer.

A faktorelemzés elvileg két különböző cél elérésére használható:

1. Nagyszámú korrelált változó redukálása a lehető legkisebb számú tényezőre.
2. A változók közötti összefüggések mögött álló *okok* (tényezők) *feltárása* és előállítása.

Az 1. célkitűzéshez a *legalkalmasabb a főkomponens-elemzés* (PCA). Itt olyan kis számú faktorokat (főkomponenseket) keresnek, amelyek a változókban lévő információ maximális varianciáját (információját) képesek reprodukálni. Ehhez természetesen kompromisszumot kell kötni a faktorok számának minimalizálása és az információveszteség minimalizálása között.

A 2. célkitűzéshez a *közös faktorelemzés az alkalmas és gyakran használt módszer*. A faktorokat a megfigyelt változók és korrelációik *okaként* értelmezik. Ebben az esetben feltételezik, hogy a faktorok *nem magyarázzák a változók teljes varianciáját*. Ezért a korrelációs mátrix nem reprodukálható teljes mértékben a faktorsúlyokkal.

A gyakorlatban való alkalmazási gyakoriság alapján mi a továbbiakban a főkomponens elemzést mutatjuk be, a közös faktorelemzés kivitelezésekor az SPSS-ben ugyanazokat a parancsokat és lépéseket kell választani, mint a főkomponens elemzéskor, csak az opciók térnek el. A módszer elméletének mélyebb megismerésére a Malhotra, Simon (2009) és a Backhaus et al. (2018) szakkönyveket ajánljuk. A faktorelemzés különböző módszerei esetén ugyanazon lépések alkalmazása miatt is sok statisztikai programban, az SPSS-ben is a főkomponens elemzés szerepel előre beállított (default) extrakciós eljárásként.

A főkomponens-elemzés

A faktorelemzésben a kiinduló változók gyakran standardizálásra kerülnek, erre akkor van szükség, ha eltérő metrikájúak a kiinduló változók. Egy változó varianciája a változóban lévő információ mértékegysége. Ha egy változó varianciája 0, akkor nem tartalmazhat semmilyen információt. Ellenkező esetben a következő érvényes: a standardizálás után minden változó varianciája 1.

A főkomponens elemzés első lépésben előállítja az elméleti alapmodell alapján a kiinduló változókat az új változók, a faktorok lineáris kombinációjaként, ahol az együttthatók a faktorsúlyok. Az első lépésben a teljes faktorsúly mátrixban annyi faktor szerepel, ahány kiinduló változó és ekkor ezek a kombinációk a kiinduló változók varianciáját 100%-ban tartalmazzák. Ebből a teljes faktorsúly mátrixból kerül kiválasztásra a redukált faktorsúly mátrix, amely a közös faktorok kiválasztását jelenti. Ehhez a faktorsúlynégyzet mátrix előállítása szükséges, amelyben a sor- és oszlopösszegek, a peremértékek alapján törthet a kiválasztás.

A faktorelemzés központi kérdése, hogy a kiinduló változóinkat hány új látens változóba, azaz hány faktorba kívánjuk összetömöríteni. Itt nincs egyetlen követendő szabály, azonban vannak iránymutatások, melyek segítenek a kiválasztásban.

1. **A Priori meghatározás.** Előfordulhat, hogy előzetes kutatásokból tudjuk, hány faktort akarunk előállítani.
2. **A saját értéken alapuló meghatározás.** Ebben az esetben azokat a faktorokat választjuk ki közös faktorként, amelyek sajátértéke nagyobb, mint 1.0 (**Kaiser kritérium**). A **sajátérték** a faktorhoz kapcsolódó variancia nagyságát fejezi ki.
3. **A magyarázott varianciarányadon alapuló meghatározás.** Ebben az esetben úgy határozzuk meg a faktorok számát, hogy a magyarázott kumulált variancia hányad elérjen egy adott szintet, ezt legalább 60%-ban határozzuk meg.
4. **A sajátértékábrán (scree-plot) alapuló meghatározás.** Ahol az ábrán törés látható, ott van az ugrás a nagy sajátértékű faktorok és a kisebbek között.
5. **A kétfelé osztás megbízhatóságán alapuló meghatározás.** Ebben az esetben a mintát kétfelé osztjuk és mindkét felén elvégezzük a faktorelemzést. Csak azokat a faktorokat tartjuk meg, amelyekben a faktorsúlyok között magas az egyezés a két almintában.
6. **Szignifikancia-vizsgálaton alapuló meghatározás.** A sajátértékek szignifikanciáját határozzuk meg, és csak a statisztikailag szignifikáns faktorokat tartjuk meg. Ennek az a hátránya, hogy nagy minta esetén (nagyobb, mint 200) sok olyan faktor is szignifikáns lesz, amely egyébként nem jelentős.

A leggyakrabban alkalmazott kiválasztási módszer a sajátérték és a Kaiser kritérium alapján történik (ez a default beállítás). A Kaiser kritérium azt jelenti, hogy az 1-nél nagyobb sajátértékű faktorokat, komponenseket válasszuk ki közös faktornak. Ezt a kiválasztást mindig össze kell kapcsolni a magyarázott varianciarányad figyelembe vételével, így tulajdonképpen a két kritériumot együttesen vesszük figyelembe. A magyarázott varianciarányadnál ökölszabályként azt alkalmazhatjuk, hogy semmiképpen ne legyen 50% alatti, de a szakirodalmi ajánlás inkább a 60%-os küszöbérték. A Kaiser kritérium és a magyarázott varianciarányad együttes figyelembe vételekor gyakran előfordul, hogy el kell térnünk az 1-nél nagyobb sajátértékű faktorok kiválasztásától, különösen abban az esetben, ha az 1-nél kisebb sajátértékűek közül a legnagyobb sajátérték(ek) 1-nél alig kisebbek. Ebben az esetben nagyon is érdemes megnézni a magyarázott varianciarányadot abban az esetben, ha 1-nél kisebb, de közel álló nagyságú faktort is

bevonunk a közös faktorok közé. Ajánlható, hogy a közös faktorszám növelést a magyarázott varianciarányad növekedéssel való összevetésben vizsgáljuk, valamint vizsgáljuk meg az egyes esetekben a faktorok értelmezhetőségét is.

Látható, hogy ezek a döntési pontok is alátámasztják azt a kritikát, hogy a faktorelemzés sok kutatói szabadságot vagy önkényességet tesz lehetővé. Az igaz, hogy a kutató döntési helyzetekbe kerül, de a statisztikai mutatószámok és próbák figyelembe vétele pontosan a döntés megalapozottságát segítik, ami az önkényességet nagyban korlátozza. A módszer célja és alkalmazásának eredménye az elemző számára több előnyt jelentenek, mint amennyi bizonytalansággal a közös faktorok kiválasztása járhat. Minden esetre érdemes többféle faktorszámot és faktorelemzési módszert is alkalmazva eljárni mind a módszer, mind a közös faktorok számának kiválasztásakor.

A következő lépés maga a főkomponens elemzés elvégzése. Ebben az esetben megkapjuk a kiinduló változók és a közös faktorok közötti kapcsolatot leíró egyenletekből az együtthatókat, vagyis a faktorsúlyokat. A faktorsúlyok azt jelentik, hogy az eredeti változó és a faktor között milyen szoros az összefüggés, vagyis az eredeti változó mekkora súllyal szerepel az adott faktor előállításában. Az első faktorsúly mátrixban a főkomponens elemzés esetében a módszer célja és algoritmusa következtében az első faktor domináns szerepű lesz, vagyis a legtöbb kiinduló változó ehhez az első faktorhoz fog tartozni. Részben ennek az egyoldalú helyzetnek a megváltoztatása céljából, részben azokra az esetekre, amikor egy változó besorolása a faktorokhoz nem látszik egyértelműnek, javasolja a módszertan a faktorok rotálását. A faktorok rotálásakor, forgatásakor a kommunalítások, vagyis a teljes magyarázott varianciarányad nem változik, csak az egyes faktorokhoz tartozó varianciarányad.

A rotációra kétféle esetet különböztethetünk meg:

- a leggyakrabban használt forgatási módszer a *merőleges transzformáció*, amikor úgy történik a forgatás, hogy a faktorok derékszöget zárnak be egymással. A leggyakrabban alkalmazott ortogonális forgatási módszer a VARI-MAX transzformáció, amely csökkenti az egy faktorhoz tartozó változók számát, így a változók elosztása jobban segítheti az értelmezést. Az ortogonális forgatás eredményeként a faktorok páronként korrelálatlanok.
- a forgatás *ferdeszögű* is lehet, ekkor a faktorok tetszőleges szöveget zárhatnak be egymással. Ezt a forgatást akkor javasolják, amikor feltételezhető a faktorok közötti korreláció, ilyenkor ugyanis a keletkező faktorok nem korrelálatlanok.

Az új változókat, amelyeket mi hoztunk létre, csak a kiinduló változók segítségével értelmezhetjük. Megkérdésen alapuló adatunk ugyanis a kiinduló változóinkra van. Az értelmezés a faktorsúly mátrix alapján történik, amikor is a

faktorok és az eredeti változók egymáshoz rendelését a magas faktorsúlyok esetében tehetjük meg. A módszertani javaslat a küszöbértékre a 0,5 -es faktorsúly, abszolút értékben. Az abszolút értékben 0,5-nél magasabb, de negatív előjelű faktorsúly értelmezése: a változó magas súllyal részt vesz a faktor előállításában, de az eredeti változóhoz képest a faktorban az az állítás ellenkező előjellel, vagyis ellentétes irányú értelmezéssel szerepel.

Az elemzés befejező lépése a faktor (főkomponens) értékek kiszámítása, amelyet a módszer felkínál számunkra. Ebben az esetben kiszámítjuk az összes megkérdezettre vonatkozóan az új, mesterséges változóink értékét, amelyek standardizált értékek, a várható értékük 0 és a szórásuk 1. Így a faktorértékek 0 körüliek vagy pozitívak vagy negatívak lehetnek. Amennyiben a faktorokat lementjük az adatbázisunkba, a megkérdezettjeink magatartását elemezhetjük az új, látens fogalmak és az őket jellemző faktorértékek alapján. Az értelmezésnél figyeljünk arra, hogy az eredeti változóknál milyen volt az értékelő skála értelmezése. Amennyiben például egy 1-től 7-ig tartó skála esetében a 7-es képviselte a teljes egyetértést vagy a legnagyobb fontosságot, akkor a faktorértékek közül a pozitívak az átlaghoz (0) képest nagyobb vagy sokkal nagyobb egyetértést, fontosságot jelentenek, illetve a negatív értékek fordítva, az átlaghoz képest alacsonyabb vagy sokkal alacsonyabb egyetértést vagy fontosságot jelentenek. Figyeljünk erre, különösen nemzetközi kutatások esetében nem egyértelmű, hogy az értékelő skála pontjainak mi a jelentése.

Milyen vizsgálatokat végezhetünk a faktorértékekkel, a látens változóinkkal azon kívül, hogy minden egyes válaszadóra megismerjük az értéküket? Mivel ezek új, metrikus változók az adatbázisunkban, elvégezhetünk minden olyan összefüggésvizsgálatot, amelyet egyébként a többi változónkra végezhetünk. Ez lehetővé teszi azt a típusú szegmentációt, amikor egyes előre meghatározott csoportokra vonatkozóan (például demográfiai vagy más változó szerinti csoportokra) megvizsgáljuk, hogy a faktorok várható értéke az egyes csoportokban szignifikánsan különbözik-e egymástól.

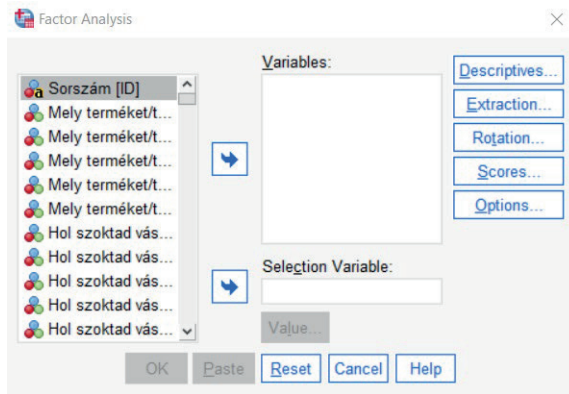
Ez a szegmentációnak az előre meghatározott csoportokra való egyik lehetséges vizsgálati módszere.

Gyakorlati példa

A válaszadók vásárlási döntéseit befolyásoló leíró változók (Q5_01-től Q5_13-ig) esetében végezzünk faktorelemzést, és válasszuk ki a legjobb megoldást. A rotálatlan vagy a Varimax rotált megoldást alkalmazzuk inkább?

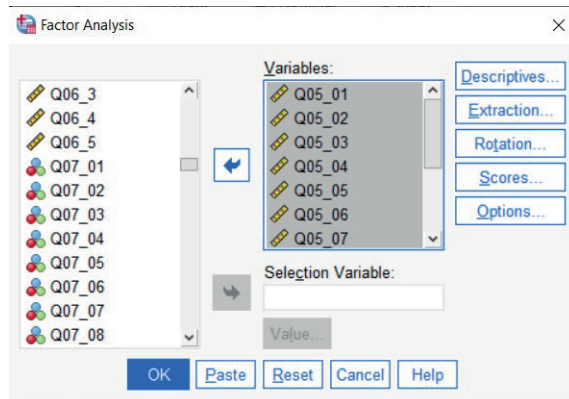
MEGOLDÁS

Elérés útvonala: Analyze/ Dimension Reduction/ Factor



Változók bevitele: Variables: Q5_01-től Q5_13-ig

Szűrés: Q3=1 a ruházati termékek vásárlása (213 megkérdezett)



Lekért adatok:

Descriptives

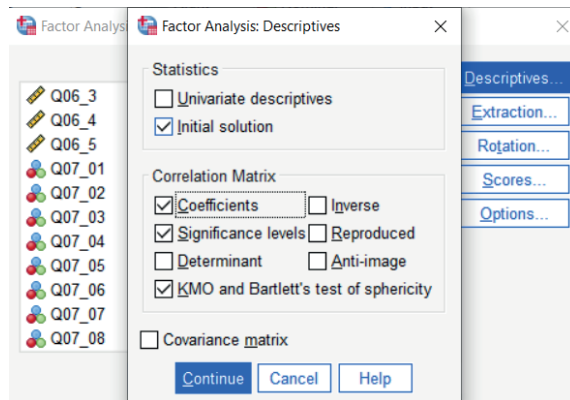
Statistics:

- ✓ Initial solution – kiinduló adatok a kommunalitásnál megjelenjenek (értéke 1)

Correlation Matrix:

- ✓ Coefficiens – elemzésbe bevont változók közötti korrelációt mutatja

- ✓ Significance levels – elemzésbe bevont változók közötti korreláció szignifikanciaszintjét mutatja
- ✓ KMO and Bartlett's test – a faktorelemzés előfeltételeként szolgáló mutatók és próbák eredményét adja



Az utoljára vásárolt kategóriában (Q03) 213 megkérdezett jelölte meg a ruházatot, a döntési magatartást erre a kategóriára vonatkozóan vizsgáljuk. A szűréshez a **Data / Select cases / If condition satisfied** menüpontot használjuk, ahol Q03=1 beállítást kell válasszunk.

Kérdőívünkben több olyan kérdést is találunk, amelyek a vásárlói magatartást többtétéles skálával írják le, erre vonatkozó állításokat tartalmaznak. Az 5. számú kérdésben a vásárlási döntést befolyásoló tényezőkre vonatkozó 13 állítás található, amelyeket 7 fokozatú skálán értékelnek a megkérdezettek. A metrikus változó feltétele teljesül, a változók közötti korrelációt több módon is vizsgálhatjuk. Elsősorban a korrelációs mátrixot célszerű felírni, ezt tartalmazza a 10.1. és 10.2. táblázat.

10.1. táblázat A Q5_01-től Q5_13 kiinduló változók korrelációs mátrixa

	Q5_01	Q5_02	Q5_03	Q5_04	Q5_05	Q5_06	Q5_07	Q5_08	Q5_09	Q5_10	Q5_11	Q5_12	Q5_13
Q5_01	1												
Q5_02	0,4	1											
Q5_03	0,53	0,635	1										
Q5_04	0,382	0,381	0,427	1									
Q5_05	0,262	0,179	0,124	0,139	1								
Q5_06	0,181	0,172	0,145	0,105	0,436	1							
Q5_07	0,135	0,045	0,109	0,209	0,334	0,302	1						
Q5_08	0,294	0,215	0,202	0,238	0,137	0,308	0,153	1					
Q5_09	0,205	0,126	0,144	0,192	0,226	0,272	0,219	0,642	1				

	Q5_01	Q5_02	Q5_03	Q5_04	Q5_05	Q5_06	Q5_07	Q5_08	Q5_09	Q5_10	Q5_11	Q5_12	Q5_13
Q5_10	0,243	0,265	0,191	0,222	0,506	0,535	0,307	0,434	0,529	1			
Q5_11	0,134	0,152	0,037	0,115	0,402	0,337	0,279	0,127	0,167	0,371	1		
Q5_12	0,204	0,296	0,264	0,137	0,386	0,392	0,227	0,306	0,43	0,584	0,452	1	
Q5_13	0,201	0,214	0,15	0,164	0,183	0,264	0,136	0,2	0,259	0,366	0,511	0,542	1

10.2. táblázat A korrelációs mátrixból számolt KMO mutató és Bartlett teszt

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0,801
Bartlett's Test of Sphericity	Approx. Chi-Square	951,497
	df	78
	Sig.	<,001

A KMO mutató értéke 0,6-nál magasabb, a Bartlett teszt szignifikáns, ezek alapján megállapíthatjuk, hogy a kiinduló változók adatbázisa megfelelő a faktor-elemzés elvégzésére.

Extraction

Method – a faktor-elemzés során használni kívánt módszer kiválasztására szolgál. Alapbeállításként a főkomponens elemzés van beállítva, amely az adatok teljes varianciáját veszi figyelembe. A főkomponens elemzés akkor javasolható, ha a fő cél az, hogy meghatározzuk azon faktorok legkisebb számát, amelyek a legtöbb varianciát magyarázzák, és amely faktorok alkalmazhatók későbbi többváltozós elemzésekben. Ezeket a faktorokat főkomponenseknek nevezzük. □ *Principal components*

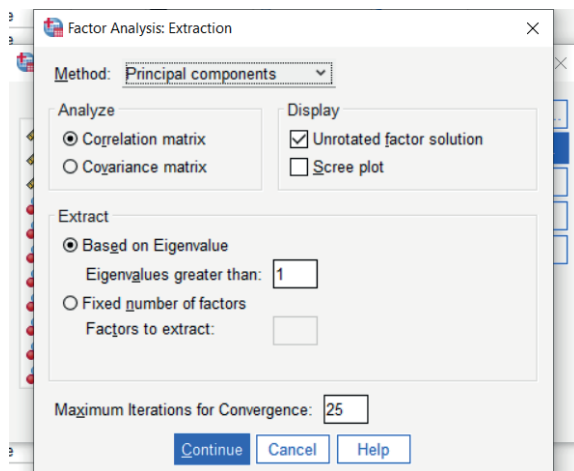
Analyze

- ✓ *Correlation matrix* – az elemzést a korrelációs mátrixból kiindulva végezze el a program.
- ✓ *Display*
 - rotáltat is mutassa be a program.
 - *Scree plot* – sajátérték ábra jelenjen meg.

Extract – Ebben a pontban tudjuk megjelölni, hogy mi alapján szeretnénk a megoldást megjeleníteni. Mint láthattuk, több szempont alapján is meghatározható az ideális faktorszám. Az SPSS-ben automatikusan be tudjuk állítani, hogy a kívánt sajátérték alapján – jelen esetben nagyobb mint 1 (Kaiser kritérium) – alakítsa ki a faktorokat a program. Amennyiben más módszer alapján szeretnénk a kívánt faktorszámot kiválasztani és elemezni – a priori, magyarázott variancia -,

a „Fixed number of factors” pontban kell megadnunk, hogy az eredmények tükrében, milyen faktorszámú megoldással kívánunk dolgozni.

- ✓ *Eigenvalues greater than: 1* (az első futtatásnál jellemzően ezzel a beállítással kezdünk, és ha szükséges, akkor később áttérünk az alábbi, fix számú faktort kínáló megoldásra)
VAGY
- ✓ *Fixed number of factors (pl. a priori meghatározott, magyarázott variancia)*



A közös faktorok és a faktorok által magyarázott variancia kiszámítása a faktorsúlynégyzet mátrix peremértékei alapján történik. Az oszlopösszegek a sajátértékek, a táblázatban a Kaiser kritérium alapján kiválasztott közös faktorok szerepelnek. A közös faktorok által magyarázott varianciát a kommunalitások tartalmazzák.

10.3. táblázat Redukált faktorsúlynégyzet mátrix

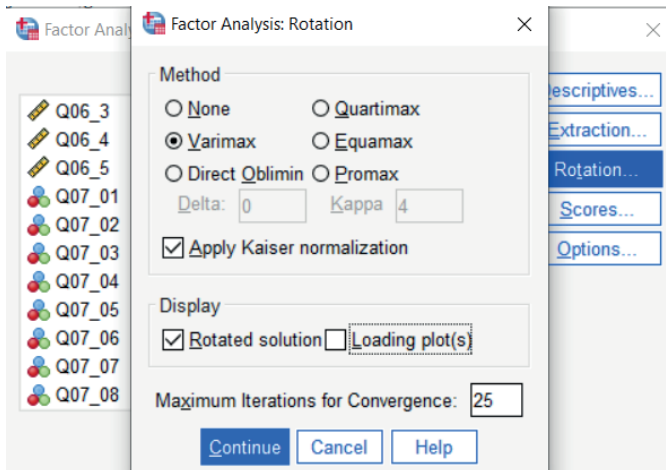
	1	2	3	4	kommunalitások (sorösszegek)
Q5_10	0,605284	0,062001	0,019321	0,003364	0,68997
Q5_12	0,534361	0,049284	0,008836	0,099856	0,692337
Q5_6	0,375769	0,082944	0,001089	0,059536	0,519338
Q5_5	0,345744	0,066049	0,066564	0,166464	0,644821
Q5_13	0,326041	0,036864	0,069696	0,312481	0,745082
Q5_11	0,305809	0,141376	0,195364	0,029241	0,67179
Q5_1	0,278784	0,265225	0,002116	0,011664	0,557789

	1	2	3	4	kommunalitások (sorösszegek)
Q5_3	0,242064	0,495616	0,016641	0,000049	0,75437
Q5_2	0,269361	0,342225	0,04	0,017424	0,66901
Q5_4	0,205209	0,256036	0,000289	0,026896	0,48843
Q5_8	0,331776	0,001681	0,443556	0,005929	0,782942
Q5_9	0,3721	0,018496	0,400689	0,009025	0,80031
Q5_7	0,191844	0,042025	0,009409	0,331776	0,575054
sajátértékek (oszlopösszegek)	4,384146	1,859822	1,27357	1,073705	

A redukált faktorsúlymátrix

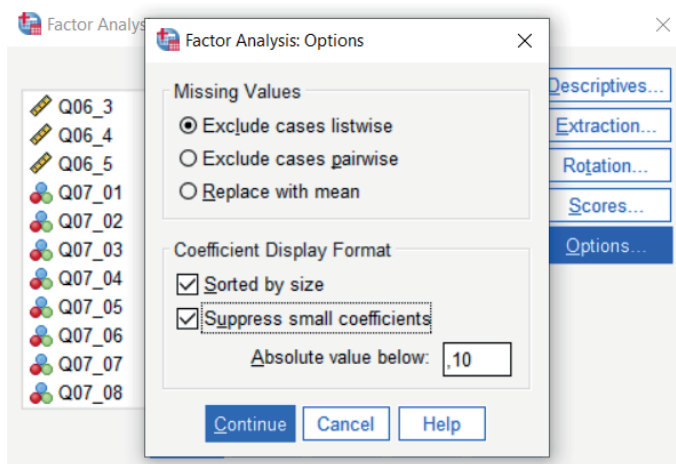
Rotation – itt állítható be, hogy rotált vagy nem rotált megoldással szeretnénk dolgozni. Amennyiben csak a nem rotált megoldás érdekel minket a „None” pontot jelöljük meg. Amennyiben bármilyen rotált megoldás érdekes lehet, válasszuk ki a kívánt módszert. Mivel az Extraction menüpontban jelöltük a „Display – Unrotated factor solution” pontot, ezért ebben az esetben is látjuk a nem rotált megoldást is. Amennyiben végső döntésünk során a rotálatlan megoldást szeretnénk elmenteni, akkor itt mindig a mentés előtt a „None” részt válasszuk ki!

✓ *Varimax*



Options

- ✓ *Sorted by size* – a változók faktorsúly szerinti sorba rendezése. Érdeemes mindig választani a könnyebb értelmezhetőség miatt.
- ✓ *Suppress small coefficients* – a kis faktorsúlyokat figyelmen kívül hagyjuk, a küszöbérték szabályozható, a default 0,10



Scores

- ✓ *Save as variable* – Az első futtatáskor még NEM választjuk ki a mentést. Akkor kell csak kiválasztani, ha már kiválasztottuk az ideális megoldást, és szeretnénk a létrehozott faktorokat elmenteni későbbi elemzések céljából. Ezek a változók a Variable View-ban megjelennek, ahol a Label cellában tudjuk őket elnevezni.

ÉRTÉKELÉS

1. Korrelációs tábla

Mivel multikollinearitást keresünk a változócsoportban, a korrelációs táblában minél több változó pár között szeretnénk látni szignifikáns, 0,3-at meghaladó korrelációt.

2. Faktorelemzés feltétele

KMO értéke 0,6 feletti, tehát az elemzés elvégezhető, mivel a főkomponens jól illeszkedik az adatokhoz. A **KMO** pontos értéke 0,801, tehát megfelelő illeszkedésünk van, a változók alkalmasok faktorelemzésre.

Bartlett-féle gömbölyűségi teszthez tartozó szignifikanciaszint 0,000, vagyis a Hot, mely szerint a megfigyelt változók korrelációs mátrixa egységmátrix, vagyis a változók páronként korrelálatlanok, elvethetjük.

Ezeknek az eredményeknek a tükrében az adatbázisunk alkalmas a faktor-elemzésre.

Az ideális faktorszám kiválasztása

Sorba vesszük, hogy a különböző kritériumok hány faktoros megoldást eredményeznek:

Kaiser kritériumot a „Total Variance Explained” táblában tudjuk ellenőrizni. Az „Eigenvalues total” oszlopban találhatóak a sajátértékek, annak alapján, hogy a faktorokhoz tartozó sajátérték magasabb legyen, mint 1, az ideális megoldásnak a 4 faktoros megoldás mutatkozik.

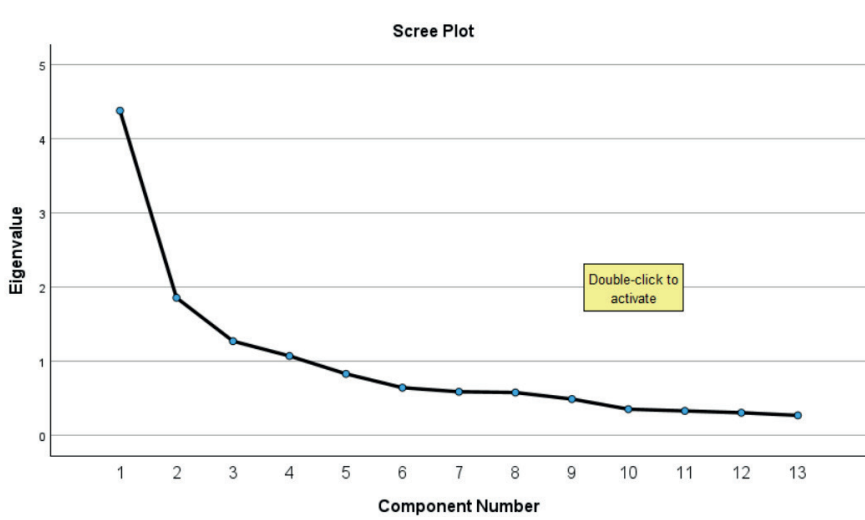
A magyarázott varianciahányad alapján (a 4 faktor által magyarázott teljes varianciahányad a kumulált oszlopokban található), vagyis, hogy a faktorok által magyarázott variancia legalább 60% legyen, az ideális faktorszám ugyanúgy 4 faktor.

Component	Total Variance Explained								
	Total	Initial Eigenvalues		Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
		% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4,383	33,713	33,713	4,383	33,713	33,713	2,439	18,762	18,762
2	1,859	14,299	48,012	1,859	14,299	48,012	2,135	16,426	35,188
3	1,274	9,799	57,811	1,274	9,799	57,811	2,021	15,545	50,732
4	1,073	8,257	66,068	1,073	8,257	66,068	1,994	15,336	66,068
5	,832	6,403	72,471						
6	,646	4,968	77,439						
7	,591	4,546	81,985						
8	,581	4,467	86,452						
9	,492	3,788	90,240						
10	,356	2,737	92,977						
11	,333	2,559	95,536						
12	,308	2,368	97,905						
13	,272	2,095	100,000						

Extraction Method: Principal Component Analysis.

Előfordulhat azonban, hogy egy magasabb magyarázott varianciahányad elérése a cél, példánkban 5 faktossal már 72, 471% a magyarázott variancia, itt azonban elengedjük a Kaiser kritériumot, ugyanis az 5. faktor sajátértéke 0,832. Ebben az esetben az extraction menüpontban a fixen előállítandó faktorszám opciót választjuk a sajátérték kritérium helyett.

A **sajátérték ábra (scree plot)** a faktorok sorrendjében (x tengely) ábrázolja a sajátértékeket (y tengely). Az itt használatos „**könyökkritérium**” alapján azt a faktorszámot kell választani, ahol az ábra meredeksége hirtelen csökken, ellaposodik, azaz van egy töréspont. Ez alapján, az ideális megoldásnak a 3 faktoros megoldás mutatkozik, mert ez után van egy erősebb, meredekebb törés.



Mivel az Extraction menü Extract pontjában a „Based in Eigenvalue”-t jelöltük meg, ezért az SPSS azt a megoldást mutatja most számunkra, amikor a faktorok sajátértéke magasabb, mint 1 (Kaiser kritérium), tehát a 4 faktoros megoldást. Nézzük meg a rotálatlan faktor mátrix és rotált faktormátrix esetében a faktorok és faktorsúlyok alakulását, valamint a kommunalitások értékét ebben az esetben.

Vegyük először a rotálatlan faktor mátrixot szemügyre (10.4. táblázat). A táblázat soraiban az elemzésbe bevont változók (kérdések) találhatóak, a Component részben megjelenő oszlopok pedig a kiválasztott megoldás esetében mutatja a faktorszámokat, jelen esetben 4 faktoros megoldással találkozunk. A táblázat belsejében található számok a faktorsúlyok. Ezek mutatják, hogy az adott változó az adott faktorhoz, milyen erősen „tartozik”, vagyis az adott változó és faktor korrelációs együtthatója ez az érték. Minél magasabb a faktorsúly, annál erősebb kapcsolat áll fenn a változó és az adott sorszámú faktor között.

Első lépésben vizsgáljuk meg minden egyes változó esetében, hogy abszolút értékben melyik sorszámú faktorhoz tartozik a legerősebb faktorsúly. Ezt célszerű valamilyen színnel külön jelölni (ehhez excelbe praktikus átvinni a táblázatot).

A faktorsúlyokat megvizsgálva megállapítható, hogy a legmagasabb faktorsúly egyetlen esetben sem alacsonyabb, mint 0,4 (ez a minimálisan elvárt faktorsúly). Az első 9 változó esetében az 1-es faktorhoz tartozó faktorsúly a legmagasabb, a következő 3 változó esetében a 2-hez, 2-2 esetében a 3 és 4-eshez, míg az 5-ös faktorhoz egyetlen egy változó sem tartozik. Mivel van egy faktorunk, amelyhez egyetlen egy változó sem tartozik a legerősebb súllyal, ezért azt a megoldást nem használhatjuk a későbbi elemzések során.

Vizsgáljuk meg a **Varimax rotált** megoldást is. a 'Rotated Component Matrix' táblában (10.5. táblázat).

Ebben az esetben az *egyes faktorokhoz tartozó változók száma is kiegyensúlyozottabb* képet mutat, ugyancsak nincs 0,4 alatti faktorsúlyunk, és minden egyes faktorhoz tartozik legalább egy változó. A kommunalitások is rendben vannak a megoldás esetében, mivel *egyetlen egy esetben sem találkozunk 0,25 alatti kommunalitással* (10.6. táblázat).

Amennyiben találkoznánk abszolút értékben 0,4 alatti legnagyobb faktorsúllyal vagy 0,25 alatti kommunalitással, akkor azokat a változókat ki kellene zárunk az elemzésből, és úgy újra lefuttatni a faktorelemzést.

A kiinduló főkomponens mátrix: több faktor nem azonosítható egyértelműen, a változók többsége az első faktornál rendelkezik magas súllyal.

10.4. táblázat Component Matrix^a

	Component			
	1	2	3	4
A vásárlással jól járok.	,778	-,249	-,139	,058
A vásárlás után úgy érzem, hogy okos döntést hoztam.	,731	-,222	,094	-,316
A vásárlásaim tökéletesen illeszkedjenek az igényeimhez.	,613	-,288	,033	,244
Azokat a termékeket tudjam megvenni, amikre igazán vágyom.	,588	-,257	,258	,408
Büszke legyek magamra, hogy jól vásároltam.	,571	-,192	,264	-,559
A vásárlás után jól érzem magam.	,553	-,376	,442	-,171
Vásárlásaimat a lehető legkényelmesebben intézzem.	,528	,515	,046	,108
A vásárlás érdekében ne kelljen extra erőfeszítéseket tennem.	,492	,704	,129	-,007
Vásárlásaim ne okozzanak túl sok gondot, nyugót számomra.	,519	,585	,200	-,132
A vásárlásaimat képes legyek minél gyorsabban megvalósítani.	,453	,506	,017	,164
Olcsóbban tudjam a termékeket megvásárolni, mint máshol.	,576	,041	-,666	-,077
Jó áron tudjam a termékeket megvenni.	,610	-,136	-,633	-,095
Jó minőségű termékeket tudjak megvásárolni.	,438	-,205	,097	,576

Extraction Method: Principal Component Analysis.
a. 4 components extracted.

10.5. táblázat Rotated Component Matrix^a

	Component			
	1	2	3	4
A vásárlás érdekében ne kelljen extra erőfeszítéseket tennem.	,863	,078	,008	,055
Vásárlásaim ne okozzanak túl sok gondot, nyugót számomra.	,780	,245	-,014	,034
Vásárlásaimat a lehető legkényelmesebben intézzem.	,711	,053	,165	,148
A vásárlásaimat képes legyek minél gyorsabban megvalósítani.	,665	-,032	,166	,131
Büszke legyek magamra, hogy jól vásároltam.	,128	,843	-,007	,135
A vásárlás után jól érzem magam.	,010	,723	,382	-,056
A vásárlás után úgy érzem, hogy okos döntést hoztam.	,164	,708	,231	,331
Azokat a termékeket tudjam megvenni, amikre igazán vágyom.	,129	,248	,752	,035
Jó minőségű termékeket tudjak megvásárolni.	,083	-,027	,750	,076
A vásárlásaim tökéletesen illeszkedjenek az igényeimhez.	,070	,281	,604	,265
Jó áron tudjam a termékeket megvenni.	,072	,160	,149	,864
Olcsóbban tudjam a termékeket megvásárolni, mint máshol.	,202	,056	,073	,856
A vásárlással jól járok.	,146	,402	,496	,511

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 5 iterations.

10.6. táblázat Communalities

	Initial	Extraction
Vásárlásaimat a lehető legkényelmesebben intézzem.	1,000	,558
Vásárlásaim ne okozzanak túl sok gondot, nyugót számomra.	1,000	,669
A vásárlás érdekében ne kelljen extra erőfeszítéseket tennem.	1,000	,754
A vásárlásaimat képes legyek minél gyorsabban megvalósítani.	1,000	,488
Azokat a termékeket tudjam megvenni, amikre igazán vágyom.	1,000	,644
A vásárlásaim tökéletesen illeszkedjenek az igényeimhez.	1,000	,519
Jó minőségű termékeket tudjak megvásárolni.	1,000	,575

	Initial	Extraction
Olcóbban tudjam a termékeket megvásárolni, mint máshol.	1,000	,783
Jó áron tudjam a termékeket megvenni.	1,000	,800
A vásárlással jól járok.	1,000	,691
A vásárlás után jól érzem magam.	1,000	,672
A vásárlás után úgy érzem, hogy okos döntést hoztam.	1,000	,692
Büszke legyek magamra, hogy jól vásároltam.	1,000	,745

Extraction Method: Principal Component Analysis.

Mivel minden kritériumnak megfelelünk, ezért a megoldás ideálisnak tűnik, tehát nevezzük el a faktorokat. A faktorok nevét a hozzájuk tartozó változók alapján adjuk meg, mivel azoknak a változók jelentését tömörítik legnagyobb súllyal.

Amennyiben egy változó a faktorhoz negatív súllyal tartozik, akkor annak a változónak az ellentétes jelentését kell figyelembe vennünk.

Ezek alapján a faktorainkat a következőképpen nevezhetjük el:

10.7. táblázat

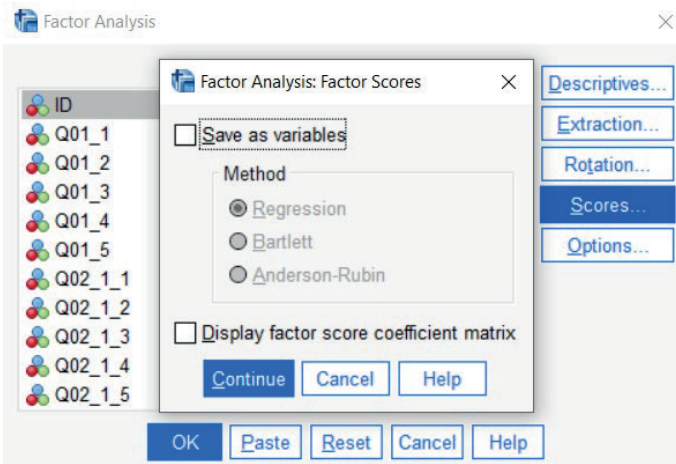
A kiinduló változók	Az új változók, a faktorok				
	1	2	3	4	
A vásárlás érdekében ne kelljen extra erőfeszítéseket tennem.	,863	,078	,008	,055	Kényelmes vásárlás fontossága
Vásárlásaim ne okozzanak túl sok gondot, nyugót számomra.	,780	,245	-,014	,034	
Vásárlásaimat a lehető legkényelmesebben intézzem.	,711	,053	,165	,148	
A vásárlásaimat képes legyek minél gyorsabban megvalósítani.	,665	-,032	,166	,131	
Büszke legyek magamra, hogy jól vásároltam.	,128	,843	-,007	,135	A jó döntés fontossága
A vásárlás után jól érzem magam.	,010	,723	,382	-,056	
A vásárlás után úgy érzem, hogy okos döntést hoztam.	,164	,708	,231	,331	
Azokat a termékeket tudjam megvenni, amikre igazán vágyom.	,129	,248	,752	,035	Megfelelő termék vásárlása
Jó minőségű termékeket tudjak megvásárolni.	,083	-,027	,750	,076	
A vásárlásaim tökéletesen illeszkedjenek az igényeimhez.	,070	,281	,604	,265	

A kiinduló változók	Az új változók, a faktorok				
	1	2	3	4	
Jó áron tudjam a termékeket megvenni.	,072	,160	,149	,864	Jó áron vásárlás
Olcsoóbban tudjam a termékeket megvásárolni, mint máshol.	,202	,056	,073	,856	
A vásárlással jól járnak.	,146	,402	,496	,511	

Amennyiben van olyan változó, amely a döntő faktorsúly mellett más faktorokban is nagy súllyal rendelkezik, mint jelen esetben a „A vásárlással jól járnak” változó, megfontolhatjuk ennek a változónak az elhagyását és újrafuttatjuk a faktoranalízist.

Amennyiben nehezen értelmezhető a megoldás (azaz az egy faktorba sorolt változók között tartalmi összefüggés nincs), vagy nehezen elnevezhetően érezzük őket, érdemes megnézni más megoldási opciókat is. Ezt úgy tehetjük meg, hogy újrafuttatjuk a faktoranalízist, de az 'extraction' menüben sajátérték/eigenvalues helyett azt választjuk, hogy „fix number of factors”, majd megadjuk a kívánt faktorszámot. Ez nem változtatja a KMO és a Bartlett eredményeit, mivel ugyanazt a változócsoportot elemezzük, ellenben előfordulhat, hogy könnyebben értelmezhető megoldásra jutunk.

Jelen példánkban a sajátértékek alapján is jól értelmezhető faktorokhoz jutottunk, ezért úgy döntünk, hogy ezzel a megoldással megyünk tovább. Amennyiben ezekkel a faktorokkal a jövőben szeretnénk dolgozni, el kell őket menteni, melyet az *Analyze/Dimension Reduction/Factor* menüponton belül a Scores menüvel tudunk megtenni, a 'Save as variables' opció segítségével.



A mentés eredményeként az adatbázisban megjelennek az új változók, ahol a Label cellába a korábbi elnevezéseket be tudjuk gépelni. Ettől kezdve ezekkel a *standardizált, metrikus változókkal* bármilyen elemzést el tudunk végezni.

A Name oszlopba az SPSS automatikusan ad egy kódot a faktoroknak. A FAC jelenti, hogy faktorelemzés eredményeként jelentek meg az új változók, az „_” utáni szám jelöli, hogy az adott adatbázison belül hányadik elmentett megoldásunk van. Azok a változók, ahol az „_” utáni szám megegyezik egy megoldáshoz tartoznak. Az „_” előtti szám mutatja, hogy az adott változók, az adott, elmentett faktorelemzés hányas számú faktora.

Pl. FAC1_2 – a másodsorra elmentett faktorelemzés első számú faktora

FAC2_2 – ugyanezen faktorelemzés – másodsorra elmentett – második számú faktora.

Érdekes az általunk adott fantázianeveket eltárolni az adatbázisban, ezeket a „label”-be kell írni a további elemzésekhez.

Ahogy a változó típusából is látjuk, az új változók, a faktorok mindig metrikus változók lesznek.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
186	FAC1_1	Numeric	11	5	Kényelmes vásárlás	None	None	13	Right	Scale	Input
187	FAC2_1	Numeric	11	5	Jó döntés	None	None	13	Right	Scale	Input
188	FAC3_1	Numeric	11	5	Megfelelő termék vásárlás	None	None	13	Right	Scale	Input
189	FAC4_1	Numeric	11	5	Jó áron való vásárlás	None	None	13	Right	Scale	Input

Az adatbázis Data View-jában láthatjuk, hogy az adott válaszadóhoz, az adott faktor esetében milyen érték tartozik. Azok a válaszadók, akik az elemzésben bevont változók közül legalább egyre nem választottak automatikusan hiányzó értéként jelennek meg.

	FAC1_1	FAC2_1	FAC3_1	FAC4_1
1	1,37363	-,03926	-2,11067	,51580
2	-,65454	,78505	-1,51406	,28796
3	-,09229	-,33323	-,62311	,85819
4	-1,17904	-1,35904	-1,93961	,08357
5	,14153	,64561	-,05677	1,24297
6	-,28783	,43166	-,27865	,80003
7	-,59183	-,43666	-2,14024	,83808
8	,25879	,82496	-,32423	,34728
9	-1,54904	-,23394	-,47316	,80043
10	,73724	1,12420	-1,24553	1,12865
11	-,09922	,79152	-1,02516	1,29508
12	-1,10098	,40039	-1,40070	,13445
13	,10906	-,05284	-2,53136	-,72235
14	1,57849	,46733	-1,48818	-,39488
15	1,13563	,52629	-1,14202	-,29553
16	-,20719	,20921	-1,96736	-,38574
17	-1,40744	,23536	-1,11076	,89248
18	,38557	,58789	-,73231	,81200
19	,66735	,14658	-,72115	-,32989
20	,40106	,11527	-3,13045	-,47472
21	-1,72707	-,33208	-1,22463	,98227
22	,48710	1,48867	-1,36795	,00584
23	-,17532	,78647	-,72016	,38356
24	,05129	,40008	-,16769	1,30612
25	-,68098	-,07242	-1,23192	1,81777

A faktorok standardizált változók, tehát a 0 az átlagos szintet jelöli, a faktorelemzésbe bevont változók esetében a skála két végpontja az „Egyáltalán nem értek egyet”, illetve a „Teljes mértékben egyet értek” voltak. Ezért, *ha egy válaszadónál az adott faktorhoz tartozó érték 0, akkor ő átlagos szinten ért egyet az adott faktoral, ha 0 feletti értéket látunk, akkor átlag feletti szinten ért egyet az adott tényezővel.* Minél magasabb értéket látunk, annál jobban ért egyet átlag felett az adott értékkel. *Negatív érték esetében a válaszadó az adott tényezővel átlag alatti szinten ért egyet.* Minél magasabb a szám abszolút értékben, annál inkább átlag alatti szinten ért egyet.

Például: az első válaszadó a kényelmes vásárlás (1,374) és a jó áron vásárlás (0,518) tényezőkkal átlag feletti szinten ért egyet (átlag feletti szinten jellemző saját viselkedésére az adott tulajdonság), míg a második faktorról, a jó döntés fontosságával (0,039) átlagos szinten, a termék megfelelőségével (-2,110) az átlagnál jóval alacsonyabb szinten ért egyet. Legkevésbé a minőségi, megfelelő termékre való törekvés jellemzi az adott válaszadót.

A faktorértékek értelmezése a kiinduló változók skála-értelmezésének felel meg, ahol a magasabb érték jelzi az egyetértést, az alacsony az elutasítást. Ellenkező értelmezésű kiinduló skálánál a faktorértékek értelmezése is ellentétesen történik.

Az Analyze/Reports /Case Summaries parancsot lefuttatva lekérhető a válaszadókhoz tartozó részletes faktorértékek.

Hasonlóan az első válaszadó jellemzéséhez az összes válaszadót jellemezhetnénk az új változók, a faktorok értékei alapján. Ezeket az értékeket úgy is értelmezhetjük, hogy „ezt mondták volna a válaszadók, ha a most létrehozott, látens változókat, fogalmakat kérdeztük volna meg”. Így minden egyes válaszadó esetében megvizsgálhatnánk a rá jellemző vásárlói magatartást, de ez nem célja az elemzéseknek. A kutatásnak, elemzésnek a célja a tipikus vásárlói magatartással rendelkező csoportok feltárása, amelyet vagy előre megadott csoportokra vonatkozóan vizsgálhatunk, vagy struktúra feltáró módszerrel tárjuk fel az adott magatartási változók szerinti csoportokat, szegmenseket.

Gyakorló feladat

A kérdőívünkben a Q12_1 – Q12_16 kérdések a kiszállítással kapcsolatos attitűdökre vonatkoznak. Vizsgáljuk meg faktorelemzéssel, milyen fő dimenziók jellemzik a megkérdezetteket! A faktorértékek felhasználásával vizsgáljuk meg, hogy a kiszállítással kapcsolatos dimenziók összefüggésben állnak-e a megkérdezettek nemével (D1) és azzal, hogy a megkérdezett találkozott-e vásárlása során a környezetbarát termék-jelleg feltüntetésével (Q19)! Írja le és indokolja az összefüggésvizsgálati módszer választását!

11 KLASZTERELEMZÉS

Elméleti bevezetés

A klaszterelemzés célja, hogy több változó figyelembevételével a válaszadókat ún. klaszterekbe, csoportokba soroljuk. A módszer során befelé minél homogénebb (azaz az egy klaszterbe tartozó megfigyelések a lehető leghasonlóbbak legyenek egymáshoz) és kifelé minél heterogénebb (azaz egyes klaszterek határozottan különüljenek el a többi klasztertől) csoportokat szeretnénk létrehozni. A marketingben gyakorta használják a módszert a piac szegmentálására, célpia-cok meghatározására.

A klaszterelemzés feltáró adatelemzési módszer, mivel a vizsgált objektumok csoportosítására vonatkozó javaslatokat eredményez, és így „új megállapításokat” hoz létre, illetve struktúrákat fedez fel az adathalmazokban.

A klaszterelemzés nehézsége és szépsége hasonló a faktorelemzéshez. Itt sincs egyetlen megoldási mód, amelyet követhetünk, azonban számos segítség áll a rendelkezésünkre, hogy meghozzuk a döntésünket a végső csoportosításról.

A felhasznált változók metrikusak és nem metrikusak is lehetnek, a módszer egyes lépéseiben eltérő elemeket találunk. Az elemzés során *nem teszünk különbséget függő és független változó között, nincs oksági kapcsolatra vonatkozó hipotézis.*

A klaszterelemzés feltétele, hogy a felhasznált változók között *ne legyen, vagy alacsony korreláció legyen.*

Amennyiben a kiinduló adatok korrelálatlansága nem teljesül, az elmélet többféle áthidalási lehetőséget javasol:

- 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.
- 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 használat azért nem terjedt el jobban, azért választják leggyakrabban az euklideszi távolságot, mert a számítógépes programok egy része nem tartalmazza a Mahalano-

bis-távolságot, az euklideszi vizont valamennyinek része. Megjegyezzük, hogy a marketingkutatói alkalmazásokban is leggyakrabban az euklideszi távolságfogalmat használják.

- Azon változók kizárása, amelyek egymással magas szinten korrelálnak. Amennyiben két ismérv 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, de a gyakorlati esetekben ilyen magas korreláció ritkán fordul elő.
- Arra is ügyelnie kell, hogy a kiinduló adatbázisban ne legyen olyan ismérv, 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. lépés: a hasonlóságok és távolságok kiszámítása

A klaszteranalízis első lépése a kiinduló változók alapján az objektumok (válaszadók) közötti hasonlóságok/távolságok kiszámítása. A hasonlóságok kiszámítására a **nem metrikus adatok** esetében kerül sor, ezt egy egyszerű, fiktív példával mutatjuk be.

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:

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

S – 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 fantázianévvvel) hasonlítunk össze a tulajdonságaik alapján (1 – az adott tulajdonság jellemző, 0 – az adott tulajdonság nem jellemző).

11.1. táblázat

Mosópor	Kiváló tisztító-hatású	Extra fehérítő-hatású	Nagy-szemcsés	Kék színű	Nagy kiserelésű	Folyékony kiserelésű	Óceán-illatú
„Tisztító”	1	0	1	0	1	1	0
„Fehérítő”	1	1	1	0	1	0	0

a – mindkét objektumnál jelen van = 3

b – csak a 2. objektumnál van jelen = 1

c – csak az 1. objektumnál van jelen = 1

d – egyiknél sincs jelen mindkét tulajdonság = 2

δ, λ – a lehetséges (konstans) súlyok

Forrás: Backhaus et al. [2018], és saját példa.

A hasonlósági mértékeket a 11.2. táblázatban foglaljuk össze

11.2. táblázat

A koeficiens 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

Metrikus kiindulási adatok esetében az objektumok hasonlósága helyett lehetőségünk van távolságot számítani. 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.

Ebből a képletből kapjuk speciális esetként, ha $r = 1$, akkor a city-block metrikát, **ha $r = 2$, akkor az euklideszi távolságfogalmat**. Az euklideszi távolság a leggyakrabban használt távolság metrika, az SPSS-ben is ez az alapbeállítás.

Az összevonási eljárás kiválasztása

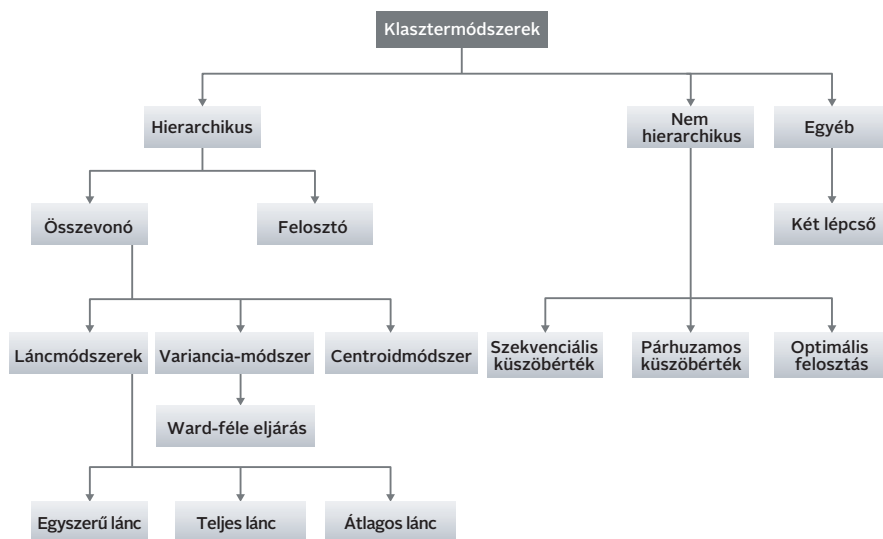
Az összevonási algoritmusokat több szempont szerint is csoportosíthatjuk, az egyik lehetőség, hogy egy vagy több változót veszünk figyelembe a csoportosítás során kiinduló változóként. Mivel a társadalomtudományi kutatásokban és a marketingkutatásban is a klaszterelemzés nagy előnyének éppen azt tekintjük, hogy egyidejűleg több változót is figyelembe vesz az összevonási eljárásnál és csoportképzésnél, ezért csak azokkal az eljárásokkal foglalkozunk, amelyek egyidejűleg több kiinduló változót vesznek figyelembe.

Egy másik csoportosítási szempont az összevonás során alkalmazott lépések lehet, ezek közül a marketingkutatási gyakorlatban való elterjedt alkalmazásuk miatt a hierarchikus és a particionáló eljárásokkal foglalkozunk részletesen.

A hierarchikus módszereken belül megkülönböztetünk agglomeratív-hierarchikus és osztó-hierarchikus módszereket. Az agglomeratív módszerek a legfinomabb felosztáson alapulnak, azaz minden objektum saját klasztert alkot, és egymás után egyesíti az objektumokat vagy klasztereket hasonlóságuk vagy távolságuk alapján. Az osztó módszerek éppen ellenkező lépéssel indul, amikor minden objektum egy klaszterben van, és több klasztert képeznek egymás után a legkülönbözőbb objektumok vagy klaszterek eltávolításával és új klaszterek kialakításával. A hierarchikus módszerekben közös, hogy egy objektum klaszterhez rendelése végleges, és utólagos „újrászortírozás” nem végezhető.

A nem hierarchikus vagy particionáló módszerek az objektumok adott csoportosításán alapulnak, és az objektumokat addig szervezik át a csoportok között, amíg *egy célfüggvény optimális értékét* el nem érik. Az úgynevezett „K-means” algoritmusoknál ez úgy történik, hogy egy bizonyos számú – a felhasználó által meghatározott – klaszter jön létre.

A II.1. ábra összefoglalóan tartalmazza az egyes eljárásokat:



Forrás: Malhotra, Simon (2009), 20.4. ábra

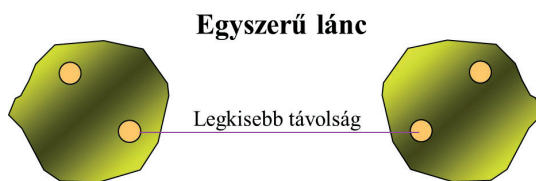
Elsősorban az alkalmazások gyakorisága miatt mi csak az összevonó módszerekkel foglalkozunk. A hierarchikus eljárásoknál kiinduló állapotban minden egyes elem egy önálló klasztert képez. Ezután minden egyes, a vizsgálatba bevont elemre kiszámítjuk az összes távolságot, majd azokat vonjuk össze, amelyeknek 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.

1. Lánc módszerek

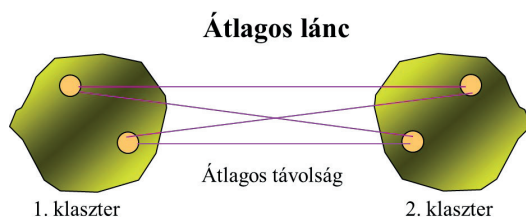
Az egyszerű lánc a minimális távolság, azaz a legközelebbi szomszéd elvén alapul (az utasításban: nearest neighbor): Minden lépésnél a két klaszter közötti távolság a két legközelebbi pontjuk közötti távolság és két klasztert akkor vonunk össze, ha a közöttük lévő távolság a legrövidebb.



A teljes lánc a legtávolabbi szomszéd elvén alapul (az utasításban: furthest neighbor): a két klaszter közötti távolságot a két legtávolabbi pontjuk közötti távolságként számolják, két klasztert itt is akkor vonunk össze, ha ez a távolság köztük a legrövidebb.



Az átlagos lánc esetében két klaszter távolságát az összes elem páronkénti távolságának átlagából számítjuk ki, ahol a pár egyik tagja az egyik klaszter eleme és két klaszter itt is a legrövidebb távolság alapján kerül összevonásra.

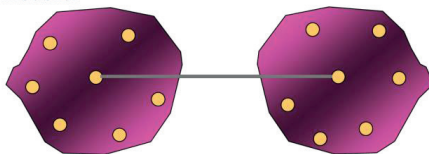


2. Variancia módszerek

Ezeket a módszereket csak metrikus adatok esetén alkalmazhatjuk, az átlag és a szórás számítása miatt.

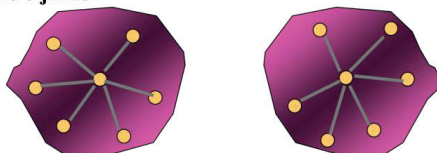
Centroid módszer: a klaszterek közötti távolságot az összes változó átlagaként számított centroidok közötti távolságként határozzák meg és a legkisebbet választják az összevonásra.

Centroid-módszer



A Ward-eljárásnál kiszámítják a klaszter szórását és azzal a klaszterrel vonják össze, amelyenél a szórás a legkisebb mértékben növekszik. A marketingkutatói szakirodalom elsősorban ezt az összevonási módszert ajánlja alkalmazni.

Ward-féle eljárás



A klaszterek számának meghatározása

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 szegmenskép-

zés szempontjából egyformán szélsőségesnek tekinthető. A kutatónak kell meghoznia a döntést arról a klaszterszámról, a döntést statisztikai mutatók segítik.

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. 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 vagy nagyon kevés 1-et meghaladó értékünk van. Amennyiben a kiinduló változóink faktorok, akkor az átlag és szórás számításával kapott eredményekből ezt az értéket azonnal le tudjuk olvasni a szórásból (standard deviation), mivel ebben az esetben 1-gyel kellene osztanunk a j -edik változó csoportban mért varianciáját.

A klaszterek összevonásának grafikus megjelenítése a dendrogram. A dendrogram azt mutatja, hogy az egyes objektumok hogyan *egyesülnek egymás után klaszterekké* (az ábrán alulról felfelé haladva). A döntés, hogy hány klasztert válasszunk, azaz hány klaszter (minél kisebb) biztosítja a csoportok optimális homogenitását és ugyanakkor a csoportok közötti megfelelő heterogenitást, az adott célfüggvénytől függ. Ez lehet például az egyes objektumoknak a klaszter középpontjától való távolsága, amelynek homogén klaszterek esetén természetesen a lehető legkisebbnek kell lennie. Mérlegelni érdemes, hogy melyik összevonás mennyire javítja vagy rontja a célfüggvényt.

A klaszterszám megállapítását segíti az értelmezhetőség is: a klaszterek értelmezése a klaszterképző változók t értéke alapján történik. 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 becsült mintabeli szórásával (a Student-féle t -eloszlás alkalmazásával). A t -értékek olyan normált értékeket jelentenek, amelyeknél a nulla körüli értékek az adott tulajdonság átlagos „képviselőt”, 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.

A klaszterek értelmezhetőségét segítheti 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 klaszterek számának meghatározásakor gyakran konfliktus lehet a célok között a „klasztermegoldás homogenitási követelménye” és a „klasztermegoldás kezelhetősége” között. Ezt a konfliktust logikai megfontolásokkal is fel lehet oldani, de ezeknek csak a kiválasztandó klaszterek számára kell vonatkozniuk, és nem a klaszterekben összefoglalt esetekre kell alapulniuk.

A végső klasztermegoldás kiválasztása a kutató alapvető ítéelőképességét és jártasságát igényli, és sokan túl szubjektívnek tartják. Bár nagyon alapos módszereket dolgoztak ki a klasztermegoldások értékelésére, még mindig a kutatóra van bízva annak eldöntése, hogy melyik klasztermegoldást tartja megfelelőnek az adott kérdés megválaszolására. Ezért alapvető fontosságú, hogy a kutatók a lehető legjobb objektivitásra törekedjenek, és érdemi megfontolások vezéreljék őket, különösen a tervezés és az értelmezés meghatározásakor.

Gyakorlati példa

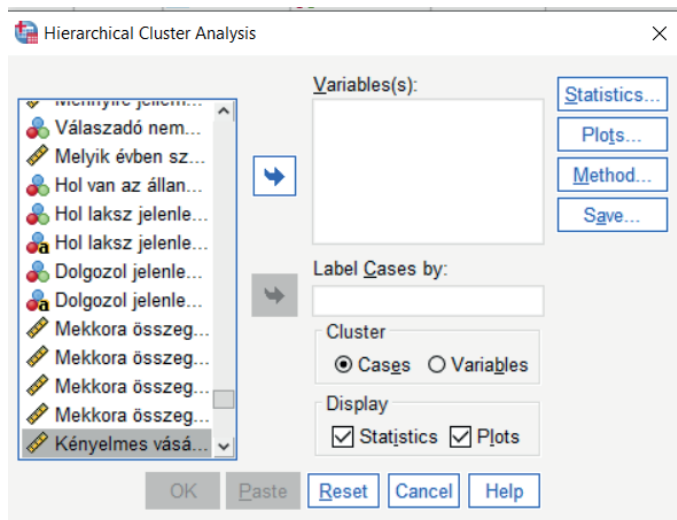
A megkérdezettek vásárlási döntési magatartását jellemző állításokból (13 állítás) hozzunk létre szűkebb, összegző dimenziókat, és ezek alapján vizsgáljuk meg, hogy milyen fogyasztói csoportok azonosíthatók a mintán.

MEGOLDÁS

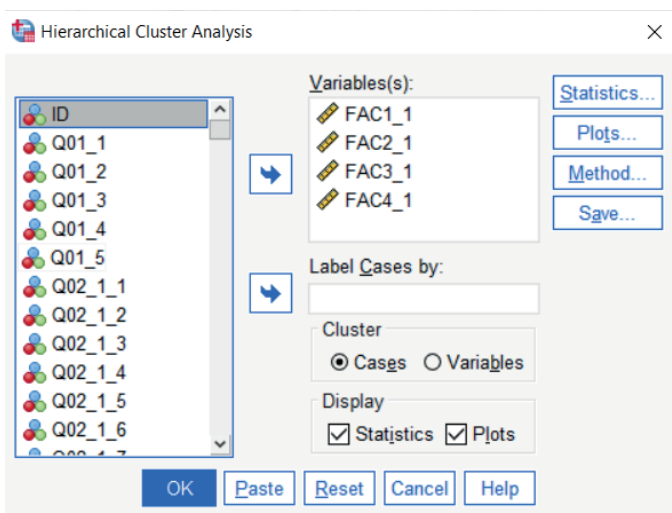
A faktorelemzésnél bemutatott példához hasonlóan ugyanazt a 13 állítást kiinduló változóként a főkomponens-elemzésbe bevonva, ebben az esetben szűrés nélkül, ugyanazt a 4 faktort kaptuk, amelyeket a szűréssel elvégzett vizsgálatban. Az így származtatott új változók a következők: Kényelmes vásárlás fontossága, a Jó döntés fontossága, Megfelelő termék, Jó áron való vásárlás faktor. Ezeket

az újonnan létrehozott korrelálatlan változókat használjuk fel a klaszterelemzés során.

Elérés útvonala: Analyze/Classify/Hierarchical Cluster



Változók bevitele: Variables: fac1_I, fac2_I, fac3_I, fac4_I



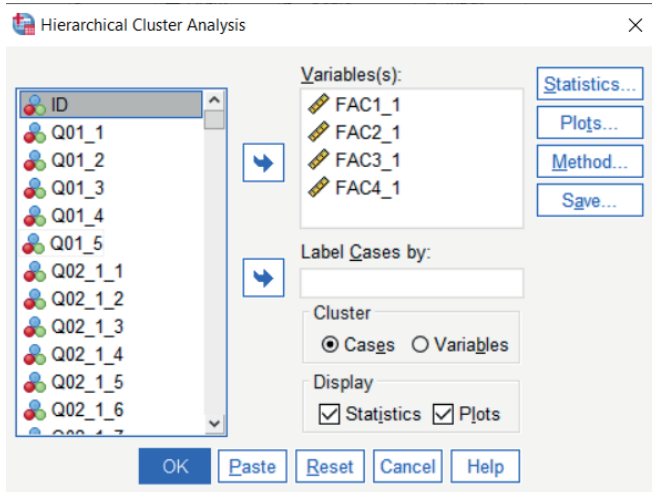
Lekért adatok:

Cluster

- ✓ cases → az válaszadókra akarunk csoportosítást látni

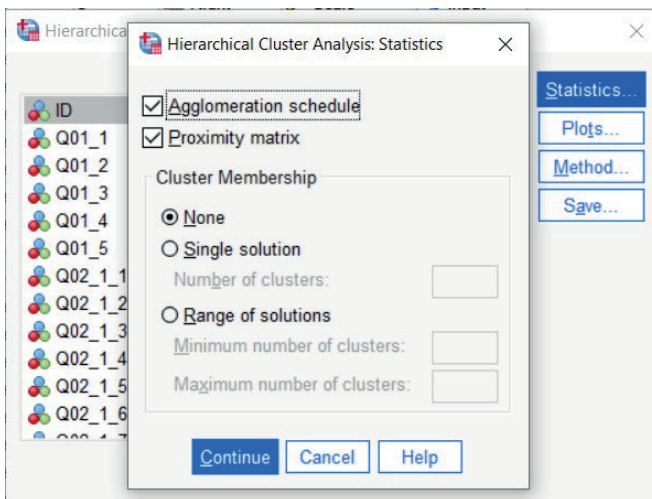
Display

- ✓ Statistics
- ✓ Plots



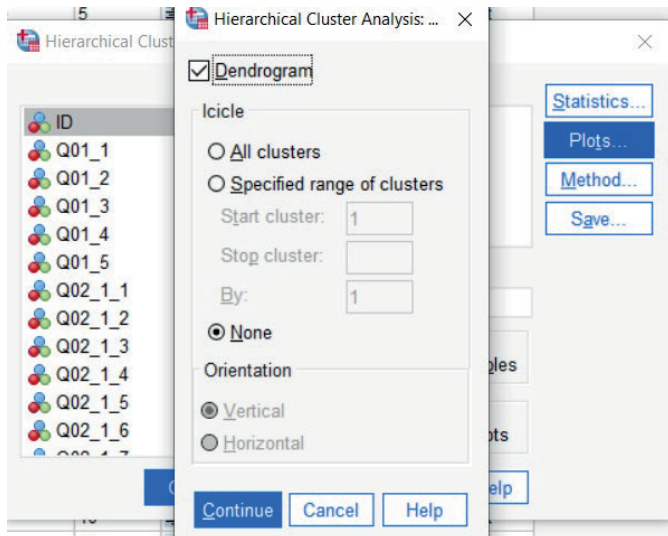
Statistics

- ✓ Agglomeration schedule → összevonási séma
- ✓ Proximity matrix → elemtávolság mátrix



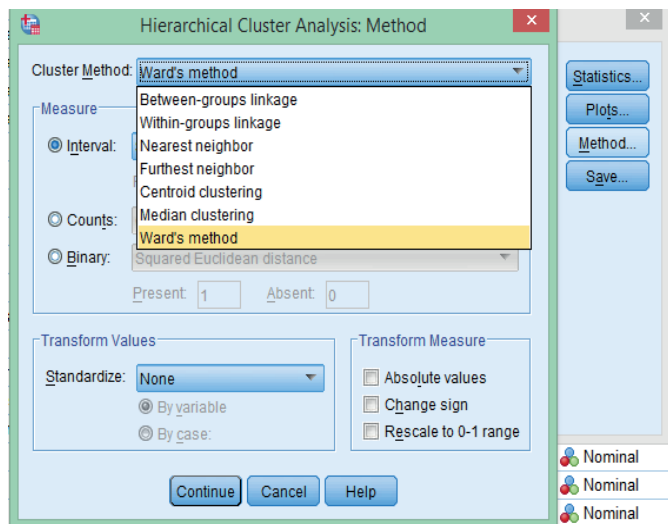
Plots

- ✓ Dendrogramm
- ✓ Icicle → jégcsap diagramm. Nem javasolt lekérni, mert lefagy a gép! (vertical/horizontal-tájolás)

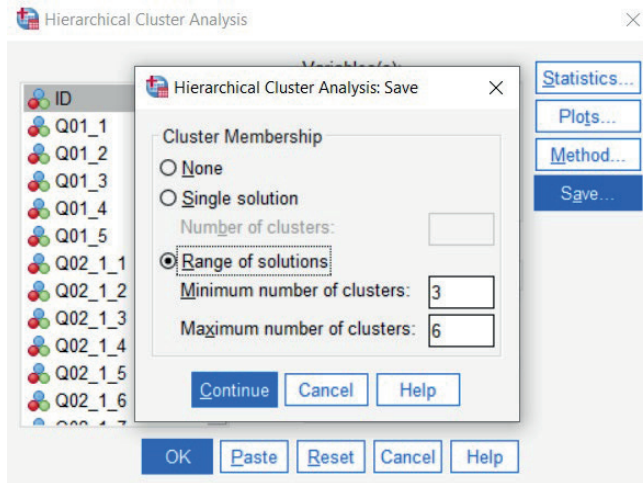


Methods – ideális megoldás megtalálása

- ✓ Ward's method



Save – mentés (csak jó eredmény esetén, tehát az első futtatásnál még nem érdemes választani, vagyis ekkor a „cluster membership ▶ none”-t választjuk.) Ha azonban már megfelelőnek látszik az eredmény, akkor érdemes vagy egy megoldást (single solution) vagy néhány megoldást menteni, hogy meg tudjuk őket vizsgálni.



ÉRTÉKELES

1. Proximity – távolság mátrix

a. Megmutatja, hogy a különböző válaszadók milyen távol helyezkednek el egymástól a választott távolságdefiníciót használva. Minél nagyobb a távolság a mátrixban két szám között, azok annál jobban különböznek egymástól a vizsgált változók mentén, azaz annál kevésbé valószínű, hogy végül azonos klaszterbe kerülnek majd. A diagonálisban 0 található, mert önmagukkal azonosak a megfigyelések.

2. Agglomeration schedule – összevonási séma

Megmutatja, hogy a klaszterelemzés különböző lépéseiben (stage) mely változók, vagy csoportok kerülnek összevonásra, milyen koefficiens (távolság) értékkel. Emellett információt kapunk arról, hogy az egyének/csoportok az elemzés során össze lettek-e már vonva bárkivel (Stage cluster first appears), valamint, hogy mely következő lépésben kerülnek újra összevonásra valakivel (Next stage).

A táblázatban található koefficiens értékek alapján lehetséges döntést hozni a klaszterek számáról. a Coefficients oszlopban a távolságok szerepelnek, ame-

lyek alapján a klaszterek összevonásra kerülnek. A koefficiens a legnagyobb értéket akkor veszi fel, amikor az összes elem egy nagy klaszterbe került (aminek az értelmezhetőség szempontjából persze nincs jelentősége, ez egy technikai lépés). A klaszterszám választásának egyik ajánlása, az 50%-os szabály szerint (hol éri el a maximum koefficiens érték a felét) jelen esetben a 6-7 klaszteres megoldás lenne célszerű (max. érték 1556 → fele 778 → ahhoz közeli érték a 792-es értéktől *hány csoportunk marad*).

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	195	196	0,000	0	0	2
2	193	195	0,000	0	1	3
3	192	193	0,000	0	2	4
4	188	192	0,000	0	3	5
...
382	2	3	697,022	370	376	387
383	4	93	739,419	373	352	384
384	4	275	792,705	383	374	387
385	1	134	846,850	378	361	386
386	1	74	974,055	385	381	388
387	1	4	1149,023	382	384	389
388	2	64	1332,270	386	379	389
389	1	2	1556,000	388	387	0

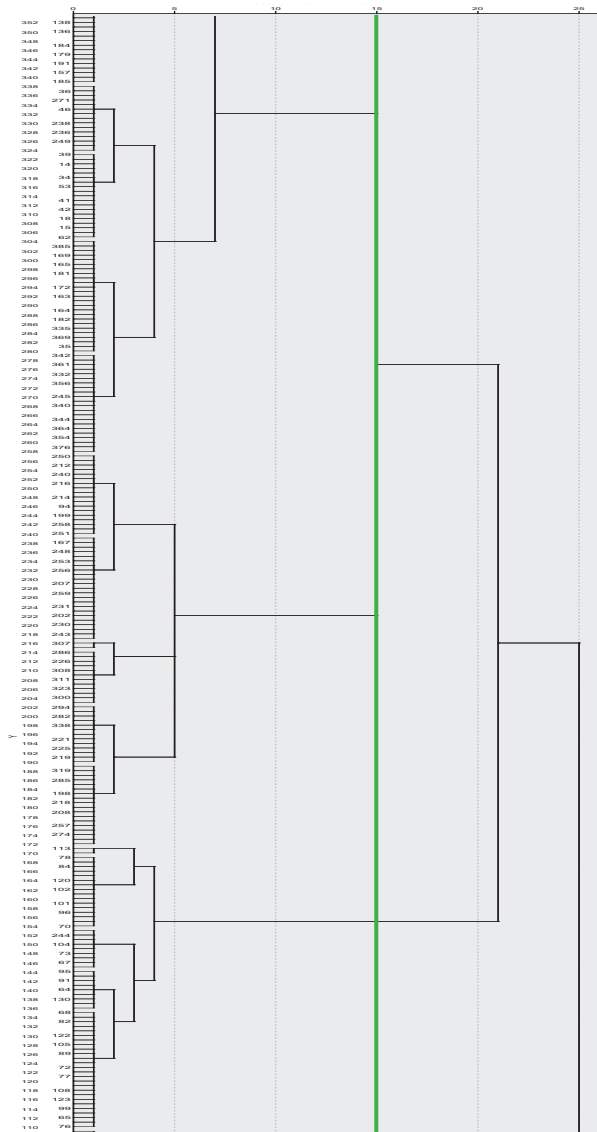
3. Dendrogram

A dendrogram is a válaszadók kapcsolódását mutatja a klaszterösszevonások transzformált távolsága alapján. Ez alapján is meg lehet határozni a klaszterszámot, ha a kutató egy saját maga által meghatározott transzformált érték mentén megnézi, hány nem „csatlakozott” csoport található még a dendrogramnál.

Például, ha az összevonásnál maximum 15 távolságot engedünk meg, akkor a 6 klaszteres megoldást kell választanunk. (Ezt praktikusán úgy kapjuk meg, hogy a 15-ös távolságnál (zöld vonal) megnézzük, hogy hány vízszintes vonalat metszünk.)

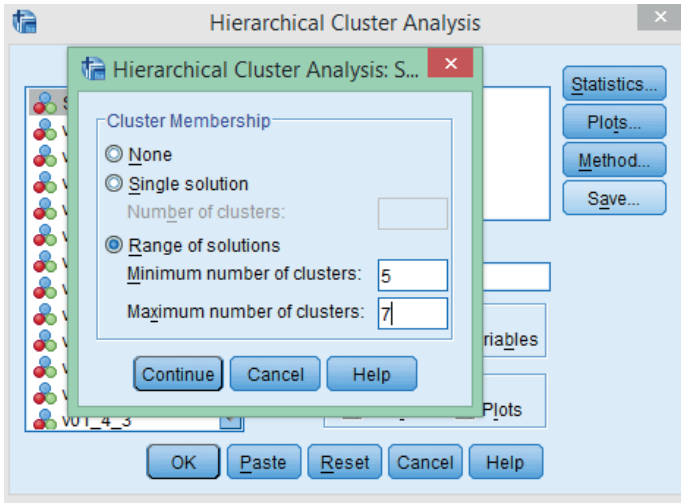
11.2. ábra A dendrogram részlete

(a teljes dendrogram a <https://www.youtube.com/@SPSSagyakorlatban> linken érhető el)



Az 50%-os elv alapján a 6-7 klaszteres megoldás tűnik ideálisnak, most menjünk ezzel tovább. A tapasztalatok alapján ilyenkor érdemes a $+/-1$ „elvet” követve elmenteni (tehát az 5, 6, 7 klaszteres megoldást is) és megvizsgálni ezeket a megoldási opciókat is.

Ahhoz, hogy meg tudjuk vizsgálni, ezek közül melyik klaszterszám lesz számunkra a megfelelő, mind a 3 megoldást el kell mentenünk az Analyze/Classify/Hierarchical Cluster menü Save pontjában.



A mentésnek köszönhetően, a faktorokhoz hasonlóan a klasztermegoldások az adatbázisban is megjelentek.

Mind az 5, 6 és 7 klaszteres megoldásra létrehozott az SPSS egy-egy változót, azaz összesen 3 új változónk keletkezett. A létrejövő változók nominális változók, amelyek azt mutatják meg, hogy az egyes megfigyelések hányadik klaszterbe tartoznak a besorolás szerint. A Data View-ban láthatjuk, hogy adott válaszadó, mely klaszterbe tartozik.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
223	CLU4_5	Numeric	8	0	Ward Method	None	None	10	Right	Nominal	Input
224	CLU3_5	Numeric	8	0	Ward Method	None	None	10	Right	Nominal	Input
225	CLU7_6	Numeric	8	0	Ward Method	None	None	10	Right	Nominal	Input
226	CLU6_6	Numeric	8	0	Ward Method	None	None	10	Right	Nominal	Input
227	CLU5_6	Numeric	8	0	Ward Method	None	None	10	Right	Nominal	Input

	4_4	CLU3_4	CLU5_5	CLU4_5	CLU3_5	CLU7_6	CLU6_6	CLU5_6
1	-	-	-	-	-	-	-	-
2	1	1	1	1	1	1	1	1
3	-	-	-	-	-	-	-	-
4	2	2	2	2	2	2	2	2
5	-	-	-	-	-	-	-	-
6	1	1	1	1	1	3	1	1
7	2	2	2	2	2	2	2	2
8	-	-	-	-	-	-	-	-
9	2	2	2	2	2	2	2	2
10	2	2	2	2	2	2	2	2

Ahhoz, hogy el tudjuk dönteni, melyik megoldás az ideális,

- meg kell vizsgálnunk a klaszterekbe tartozó válaszadók számát,
- klaszterek elemszámát megvizsgálva kerülnünk az extrém nagy vagy kicsi klaszterelemszámokat
- valamint a klaszterelemzésbe bevont változók átlagait a klasztereken belül és a hozzájuk tartozó szórásokat.

Célunk:

- hogy a szórások az adott klaszter adott változója esetében minél alacsonyabbak legyenek (Amennyiben faktorokkal klaszterezünk minél kevesebb esetben szeretnénk 1-nél magasabb szórást látni), ez a befelé való homogenitást mutatja.
- valamint, hogy minél több esetben legyen a klaszterek átlaga között szignifikáns különbség (ez pedig a csoportok közötti heterogenitást mutatja). Ehhez a varianciaelemzést tudjuk segítségül hívni a vizsgálni kívánt megoldások esetében.

Analyze/Compare Means / Means

dependent list: faktor változók

independent list: klaszter változók

Az eredményeinket megvizsgálva megállapítható, hogy minden esetben a csoportok átlagai között szignifikáns különbség van ($\text{sig}=0,000$), valamint a klaszterek elemszáma is hasonlóan alakul. Mivel az 6 klaszteres megoldás esetében csak egy esetben találkozunk 1-nél magasabb szórással, ezért ezzel a megoldással dolgozunk tovább. A klaszterek elemszáma a 6 klaszteres megoldás esetében rendre 81,51,58,62,86,52.

A klaszterek értelmezése a faktorokkal az átlagértékek alapján történik, például az első klaszter tagjainak kissé kevésbé fontos a kényelmes vásárlás (-0,11), míg a jó döntés jó áron értékes számukra(0,604 és 0,472) vagy az ötös klaszter tagjainak a legfontosabb a kényelmes vásárlás és a megfelelő termék (0,414 és 0,413), de a jó áron faktor a számukra a legérdektelenebb (-0,919).

11.3. táblázat Report

Ward	Method	Kényelmes vásárlás	Megfelelő termék	Jó áron vásárlás	Jó döntés
1	Mean	-0,11	0,179	0,472	0,604
	N	81	81	81	81
	Std. Deviation	0,509	0,699	0,483	0,395
2	Mean	-0,212	-1,746	-0,439	-0,001
	N	51	51	51	51
	Std. Deviation	0,816	0,765	1,054	0,883
3	Mean	-1,675	0,296	-0,079	0,067
	N	58	58	58	58
	Std. Deviation	0,619	0,896	1,227	0,912
4	Mean	0,325	-0,003	0,644	-1,403
	N	62	62	62	62
	Std. Deviation	0,616	0,845	0,585	0,946
5	Mean	0,413	0,414	-0,919	-0,007
	N	86	86	86	86
	Std. Deviation	0,571	0,663	0,835	0,813
6	Mean	1,173	0,419	0,535	0,671
	N	52	52	52	52
	Std. Deviation	0,242	0,360	0,367	0,264
Total	Mean	0,000	0,000	0,000	0,000
	N	390	390	390	390
	Std. Deviation	1,000	1,000	1,000	1,000

Az ANOVA tábla szignifikancia értékei mindenütt nagyon alacsonyak, vagyis mindegyik klaszterre teljesül az, hogy jól elkülönül a többitől.

11.4. táblázat Anova Table

		Sum of Squares	df	Mean Square	F	Sig.
Kényelmes vásárlás fontossága	Between Groups	258,976	5	51,795	152,967	<,001
	Within Groups	130,024	384	,339		
	Total	389,000	389			
Megfelelő termék fontossága	Between Groups	187,138	5	37,428	71,198	<,001
	Within Groups	201,862	384	,526		
	Total	389,000	389			
Jó áron vásárlás	Between Groups	141,673	5	28,335	43,993	<,001
	Within Groups	247,327	384	,644		
	Total	389,000	389			
Jó döntés érzése	Between Groups	175,507	5	35,101	63,135	<,001
	Within Groups	213,493	384	,556		
	Total	389,000	389			

Az ideális megoldás kiválasztása után a klaszterelemzésbe bevont változók átlagai alapján el kell neveznünk a klasztereket, majd jellemeznünk kell őket a klaszterelemzésbe be nem vont, de a kutatási kérdés szempontjából releváns változók alapján. Ehhez a kereszttáblát, valamint a varianciaelemzést tudjuk segítségül hívni.

11.5. táblázat A klaszterek értelmezése

	A döntésre figyelők	Nem törődők	A termékre figyelő hagyományos vásárlók	Árra figyelők	Megfelelő termékre figyelők	Kényelmesen vásárlók, akik mindenre figyelnek	
A klaszter-elemzésbe bevont változók	N(fő)	81	58	62	86	52	
	Kényelmes vásárlás fontossága	-0,11 -	-1,67 --	0,32 +	0,41 +	1,17 ++	Szig.
	Megfelelő termék fontossága	0,18 +	0,29 +	0,02 ~	0,41 +	0,42 +	Szig.
	Jó áron vásárlás	0,47 +	-0,08 ~	0,64 +	-0,92 --	0,53 +	Szig.
Klaszter-elemzésbe nem vont változók	Jó döntés érzése	0,60	0,07 ~	-1,40	-0,03 ~	0,67	Szig.
	A válaszadó neve	kissé felülreprezentáltak a nők	~	~	~	~	Szig.
	A legutóbb vásárolt termék	~	kissé inkább ruházat	inkább elektronikai termékek	inkább ruházat és könyv	~	Nem szig.
	Dolgozik-e az egyetemi tanulmányok mellett	~	inkább nem	~	inkább igen	~	Nem szig.
	Ha online keres, akkor online vásárol is	inkább igen	~	inkább igen	~	inkább igen	Szig.

ahol:

~ : átlagosan jellemző + : az átlagnál jobban jellemző - : az átlagnál kevésbé jellemző

A fenti táblázatban összefoglaltuk, hogy a vizsgált változók keresztábra elemzés vagy varianciaelemzés során szignifikáns kapcsolatot mutattak-e a klaszterváltozóval.

A klaszterelemzésbe bevont változók, jelen esetben a faktorok klasztereken belüli átlagai (ANOVA táblából kinyert érték) alapján fantázianeveket adunk az egyes klasztereknek. Ezeket a neveket az SPSS Variable view felületén is rögzíthetjük a „Values” oszlopban, hogy a későbbi elemzések során már ezek az elnevezések jelenjenek meg.

A klaszterelemzésbe be nem vont változók a klaszterek jellemzését segítik. Ez a klaszterelemzéssel való szegmentálás egyik legnagyobb előnye: ha eldöntöttük a szegmenseket és eltároltuk az adatbázisban, akkor az adatbázis összes többi változójával lehetséges a klaszterek jellemzése, természetesen a kérdések értelme szerint kiválasztva és a változó jellege szerinti összefüggésvizsgálati módszerrel vizsgálva.

Gyakorló feladat

Vizsgálja meg, hogy a faktorelemzés gyakorló feladatában kapott eredmény, a megkérdezetteknek a kiszállítással kapcsolatos attitűdjéből (Q12_1 – Q12_16 kérdések) számolt fő dimenziók alapján lehetséges-e szegmentálni a megkérdezetteket? Ha igen, jellemezze a szegmenseket a klaszterképző és más változók alapján is!

12 A CONJOINT ALAPJAI

Elméleti bevezetés

Ahhoz, hogy megértsük mi is a conjoint módszer, érdemes kitekinteni a csapatjátékok működésére. Teljesen mindegy, hogy milyen csapatjátékot veszünk alapul, hiszen minden csapatjátéknál a cél mindig a csapatunk győzelme az ellenféllel szemben. A csapat különböző posztokon játszó játékosokból áll és az edző a posztokon az aktuális ellenfél felállításának függvényében azokat a játékosait veti be, amely felállással a csapata a legnagyobb valószínűséggel győzni fog. A siker függ attól is, hogy az ellenfél mit lép, milyen csapatot küld a pályára, hiszen az edzőknek ehhez kell igazítaniuk a saját csapatuk felállítását. De lehetőség van arra is, hogy a meccs közben a különböző posztokon a játékosokat cserélgessék, vagyis a csapatot „finomhangolják” annak tükrében, hogy hogyan alakul az aktuális mérkőzés.

A conjoint esetében a csapat, aminek a győzelmét szeretnénk elérni, az egy termék- vagy szolgáltatáskonceptió. A cégünk ajánlata vagy ajánlatai versenyeznek a piacon elérhető más ajánlatokkal (ellenfelekkel). A conjoint esetében a posztok, amelyen különböző játékosok vethetők be, a termék- vagy szolgáltatásjellemezők, úgymint a márka, az ár, a kiszereleméret, a percdíj és így tovább. Ezeket a termék- vagy szolgáltatásjellemezőket szokás attribútumoknak vagy dimenzióknak is nevezni. A posztokon bevethető játékosok pedig a termék- vagy szolgáltatásjellemezők konkrét előfordulásai, úgymint a szénsavas üdítő márkáknál a Coca Cola, a Pepsi Cola, stb., a kiszereleméretnél az 1,5 liter, 2 liter, stb., az ár esetén pedig az 500 forint, a 600 forint, stb. A conjoint terminológiájában a termék- vagy szolgáltatásjellemezők konkrét előfordulásait szinteknek hívjuk.

A jelen esettanulmányban a 12.1. táblázat mutatja azt a dimenzió- és szintezettet, amelyet felhasználtunk a webshopok ajánlatainak a leképezésére.

12.1. táblázat: dimenziók és szintjeik

Dimenzió	Szint 1	Szint 2	Szint 3	Szint 4
Ár	Ingyenes	Ingyenes kiszállítás 15 000 forint felett	990 forint	Ingyenes hűségprogrammal
Szállítás teljesítésének ideje	Másnap	Harmadnap	Legalább 3 nap, de választott szállítási dátum	Legalább 5 nap, de választott szállítási dátum
Szállítás módja	Választott cím, hétköznapokon (9:00 és 18:00 között)	Választott cím, kétórás időintervallumban	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	Csomagautomata (0-24 órában)
Visszaküldési lehetőség	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	Ingyenesen csomagküldőszekrényben (0-24 órában)	990 Ft csomagküldőszekrényben (0-24)	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)

A mi esetünkben a csapat 4 posztból, azaz 4 dimenzióból állt, minden poszton 4 játékost vethettünk be a győzelem érdekében, azaz minden dimenzióknak 4 szintje volt. A kérdőívben egy-egy döntési feladatban mindig 2-2 koncepciót versenyeztettünk meg egymással, a válaszadó feladata pedig az volt, hogy a két webshop koncepció vagy ajánlat esetén mindig megmondja, hogy inkább melyik alternatívát preferálná egy 7 fokú Likert skála segítségével.

Az első úgynevezett döntési feladat, ahol a válaszadónak ki kellett választania a számára vonzóbb vagy jobban preferált webshopot, az 12.1. ábrán látható.

12.1. ábra: az első döntési feladat

Ár:	990 forint	Ingyenes				
Szállítás teljesítésének ideje:	Harmadnap	Legalább 5 nap, de választott szállítási dátum				
Szállítás módja:	Választott cím, kétórás időintervallumban	Választott cím, hétköznapokon (9:00 és 18:00 között)				
Visszaküldési lehetőség:	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)				
Egyértelműen a bal webshopot választanám		Egyértelműen a jobb webshopot választanám				
1	2	3	4	5	6	7

A második döntési feladat 12.2. ábra szerint alakul.

12.2. ábra: a második döntési feladat

Ár:	Ingyenes hűségprogrammal	990 forint				
Szállítás teljesítésének ideje:	Legalább 5 nap, de választott szállítási dátum	Legalább 3 nap, de választott szállítási dátum				
Szállítás módja:	Csomagautomata (0-24 órában)	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)				
Visszaküldési lehetőség:	990 Ft csomagküldő-szekrényben (0-24)	Ingyenesen csomagküldő-szekrényben (0-24 órában)				
Egyértelműen a bal webshopot választanám		Egyértelműen a jobb webshopot választanám				
1	2	3	4	5	6	7

És így tovább folytatódott az interjú, ugyanis minden válaszadónak 18 darab döntési feladatot kellett teljesítenie, azaz összesen 36 versengő csapattal (webshoppal) találkozott a kérdőív során.

A conjoint elsődlegesen arra fókuszál, hogy valamennyi dimenzió valamennyi szintjét „beárazza” vonzerő vagy *hasznosság* szempontjából, vagyis meghatározza azt, hogy milyen mértékben képesek az egyes szintek hozzájárulni ahhoz, hogy egy webshop kiszállítási ajánlata sikeres legyen. A conjoint terminológiájában a szintek olyan részhasznosság értékekkel bírnak, amelyek összege kiadja egy ajánlat összhassznossági értékét vagy összesített vonzerejét. A conjoint által becsült részhasznossági értékekre a 12.2. táblázatban láthatunk egy példát.

12.2. táblázat: becsült részhasznossági értékek

Dimenzió	Szint	Részhasznossági érték
Ár	Ingyenes	1,059
	Ingyenes kiszállítás 15.000 forint felett	-1,311
	990 forint	0,144
	Ingyenes hűségprogrammal	-0,153
Szállítás teljesítésének ideje	Másnap	1,059
	Harmadnap	1,068
	Legalább 3 nap, de választott szállítási dátum	0,962
	Legalább 5 nap, de választott szállítási dátum	0,110
Szállítás módja	Választott cím, hétköznapokon (9:00 és 18:00 között)	1,059
	Választott cím, kétórás időintervallumban	0,522
	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	-0,254
	Csomagautomata (0-24 órában)	0,273
Visszaküldési lehetőség	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	1,059
	Ingyenesen csomagküldőszekrényben (0-24 órában)	1,981
	990 Ft csomagküldőszekrényben (0-24)	0,506
	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	0,292

A fentebbi számok egy kiválasztott válaszadó preferenciáit tükrözik, ugyanis a conjoint elemzésnél arra törekszünk, hogy a részhasznossági modellt válaszadónként állítsuk elő. A számok értelmezésénél a „minél nagyobb, annál jobb” elv érvényesül, az előjelnek azonban nem szabad túlzott jelentőséget tulajdonítani; azaz a negatív részhasznossági értékkel bíró szintek nem jelentenek elutasítást, csupán arra utal az előjelük, hogy az „átlagnál” szintnél kevésbé kedveltek.

A szintek részhasznossági értékeinek az ismeretében meghatározható a dimenziók fontossága. A fontosság itt azt jelenti, hogy mennyire lényeges az, hogy valamely dimenzióban a legvonzóbb szintet kapja a vásárló vagy az ügyfél a legkevésbé vonzó szint helyett. A fenti példát folytatva a vizsgált válaszadónál a fontossági számok az 12.3. táblázat szerint alakulnak.

12.3. táblázat: fontossági értékek egy válaszadó esetében

Dimenzió	Fontosság
Ár	37,44%
Szállítás teljesítésének ideje	15,12%
Szállítás módja	20,74%
Visszaküldési lehetőség	26,69%

A fentiekből azt a következtetést vonhatjuk le, hogy a vizsgált válaszadó esetében a dimenzióknak a sorrendje fontosság szempontjából egyértelműen felállítható, ugyanakkor igazán domináns és igazán marginális attribútum nincs, azok mindegyikének van kisebb vagy nagyobb szerepe a vizsgált egyén döntéseiben.

Az eddigi conjoint eredmények közös jellemzője az, hogy mind a részhasznossági értékek, mind pedig a fontossági értékek leíró jellegű elemzéseket tesznek lehetővé. Ugyanakkor az nagyon lényeges, hogy a részhasznossági értékeket az egyes koncepciók összmegítéléséből származtatjuk, ami lehetővé teszi azon webshop ajánlatok összmegítélésének vagy összhazsnosságának a becslését is, amelyekkel a válaszadók az interjú során nem találkoztak. Így a kutatásba bevont szolgáltatásjellemzők szintjeiből előállítható ajánlatok közötti vásárlói döntéseket modellezni tudjuk, és ezen a téren tényleg csak a kutatói kreativitás szab korlátokat.

Az egyes webshop ajánlatok becsült összhazsnossági értéke az aktuális ajánlatot alkotó szintek részhasznossági értékeinek összegeként számítható ki. Tételizzük fel, hogy az egyik általunk modellezett webshop ajánlat 12.4. táblázat szerint alakul.

12.4. táblázat: Egy általunk modellezett webshop ajánlat

Dimenzió	Szint	Részhasznossági érték
Ár	990 forint	0,144
Szállítás teljesítésének ideje	Másnap	1,059
Szállítás módja	Csomagautomata (0-24 órában)	0,273
Visszaküldési lehetőség	Ingyenesen csomagküldő-szekrényben (0-24 órában)	1,981
<i>Összesen</i>		1,982

A kiszámított összhazsnossági érték (1,982) abszolút értékének megítélése nehézkes, a fő szabály itt is ugyanaz, mint amit a részhasznossági értékeknél megfogalmazunk; azaz minél nagyobb, annál jobb. Arra vonatkozólag azonban

nem igazán fogalmazható meg hüvelykujj szabály, hogy hol kezdődik a már valóban vonzó ajánlatok összhasznossági értéke. Sokkal inkább az a gyakorlat, hogy a különböző ajánlatokat egymással versenyeztetjük és az ajánlatok közti döntésekre dolgozunk ki valamiféle szabályrendszert. Ez különösen igaz a páros conjoint módszerekre, hiszen esetükben az interjú során a válaszadókat is relatív preferencia meghatározására kértük (a bal vagy a jobb ajánlatot vennék-e meg inkább). Tétélezzük fel, hogy a fentiekben összerakott ajánlatot (12.4. táblázat) a 12.5. táblázat szerinti ajánlattal szeretnénk megversenyeztetni.

12.5. táblázat: Versenyző ajánlat

Dimenzió	Szint	Részhasznossági érték
Ár	Ingyenes	1,059
Szállítás teljesítésének ideje	Legalább 5 nap, de választott szállítási dátum	0,110
Szállítás módja	Csomagautomata (0-24 órában)	0,273
Visszaküldési lehetőség	Ingyenesen csomagküldő-szekrényben (0-24 órában)	1,981
Összesen		3,423

A második ajánlat (12.5. táblázat) összhasznossági értéke 3,423, ami az első ajánlat (12.4. táblázat) 1,982-es teljesítményéhez képest jóval magasabb. De mégis melyik webshop ajánlatot választaná a vizsgált válaszadónk? Tradicionálisan 2 döntési logikát szokás használni a conjointban:

1. A vásárló mindig azt a termék vagy szolgáltatást vásárolja meg, amelynek az összhasznossági értéke magasabb. Ezt a megközelítést szokás *first choice* modellnek hívni. Ha ezt a megközelítést használjuk, akkor a válaszadónk az első ajánlatot választja (12.4. táblázat)
2. A vásárló mindig az ajánlatok összhasznossági értékeinek – pontosabban az összhasznossági értékek e-ad kitevőjű hatványainak – arányában vásárolja meg a termékeket vagy szolgáltatásokat. Ez a modell a share of preference logika. Ebben az esetben a két ajánlat részesedése az alábbiak szerint alakul:

- Első ajánlat (12.4. táblázat):

$$\frac{e^{3,457}}{e^{3,457} + e^{3,423}} = 0,508 \text{ azaz } 51\%$$

- Második ajánlat (12.5. táblázat):

$$\frac{e^{3,423}}{e^{3,457} + e^{3,423}} = 0,492 \text{ azaz } 49\%$$

Jelen esetben 2 ajánlat közel azonos mértékben vonzó a válaszadónak, hiszen a *share of preference* modell alapján az esetek 51%-ában az első webshop ajánlatot választaná a megkérdezett, míg a 49%-ban a második ajánlatot (12.5. táblázat). Vegyük észre, hogy a két ajánlat között a különbség részben a költségekben, részben pedig a kiszállítás teljesítésének az idejében rejlik. Lényegében itt azt vizsgáltuk meg, hogy a válaszadónk hajlandó-e 990 forintos felárat fizetni a másnapi szállításért. A válasz a *first choice* modell alapján az, hogy igen, míg a *share of preference* megközelítés alapján az, hogy nem igazán lehet eldönteni. A *share of preference* modellünk azt mondja, hogy a két ajánlat közel indifferens a válaszadónak, vagyis 990 forint felett már valószínűleg inkább az ingyenes, de lassabb kiszállítást választaná a válaszadó, míg 990 forint alatt hajlandó lenne a fizetős másnapi kiszállítást igénybe venni. A gyakorlatban azt szokás mondani, hogy ebben az esetben a 990 forint az a felár, ami maximálisan elkérhető a másnapi szállításért az 5 napos kiszállítással szemben, hiszen ennél az árpontnál válik azonosan vonzóvá a mutatott 2 ajánlat.

De melyik a jobb módszer? Erre a kérdésre nem lehet egyértelmű választ adni. A *first choice* módszer hajlamos felnagyítani a modellezett ajánlatok közötti különbséget, ami bizonyos esetekben előnyös lehet. Ha ugyanis úgynevezett *line extension* piaci helyzetet szimulálunk – amikor olyan terméket vagy szolgáltatást vezetünk be a piacra, amely nagy hasonlóságot mutat egy másik termékkel vagy szolgáltatással – a valóságot jobban leképezi a *first choice* módszer. Ilyen esetekben ugyanis a *share of preference* modell hajlamos felülbecsülni a *line extension*-t jelentő termék vagy szolgáltatás sikerét. A *share of preference* megközelítés előnye viszont abban rejlik, hogy tipikusan kisebb sztenderd hiba mellett becsüli az egyes ajánlatok preferencia részesedését, illetve ezt a módszert használva több lehetőségünk van az eredmények finomhangolására. Ha viszont a mintánk kelően nagy, akkor a *first choice* modell hibahatára is megfelelő lehet, és ha ez a módszer a pontosabb (például a jelen piaci helyzetet modellezve a valós piacrészek jobban közelíthetők), akkor érdemes ezt választani.

Nagyon lényeges kiemelni, hogy amikor a vásárlási döntéseket modellezzük a részhasznossági értékek alapján, nem csupán 2 ajánlat versenyezethető egyidejűleg, hanem azok száma tetszés szerint növelhető. Akkor is lehetőség van erre, ha olyan conjoint módszert választottunk a vásárlói preferenciák mérésére, ahol az interjúk során a válaszadók egyidejűleg csak 2 ajánlatot láttak.

A különböző conjoint módszerek

A conjoint módszernek mára már több válfaja létezik. Összességében azonban megállapíthatjuk, hogy a különböző altípusok az interjúk során alkalmazott kérdéstípusokban különböznek, a különböző módszerek outputja jóformán megegyező.

A conjoint módszereket alapvetően két dimenzió mentén csoportosíthatjuk:

- A válaszadói preferenciát milyen mérési szintű kérdéssel mérjük
- Illetve a válaszadói preferencia mérésére szolgáló kérdések adaptívak-e, azaz igazodnak-e a kérdések az aktuális válaszadó által korábban adott válaszok alapján kirajzolódó képhez vagy sem

A 4 legfontosabb technika az **12.3.** ábra mátrixán rajzolódik ki a két előbb megfogalmazott dimenzió alapján.

12.3. ábra: A négy legfontosabb conjoint technika

Igen		Adaptívak-e a válaszadónak feltett kérdések	
		Nem	
A preferencia kérdések mérési szintje	Likert skála	Adaptív conjoint analízis (ACA)	Conjoint value analysis (CVA)
	Bináris skála (választja vagy nem választja)	Adaptív choice-based conjoint (ACBC)	Choice-based conjoint (CBC)

Bármelyik módszert is választjuk a kérdőívben alapvetően az alábbiakat tesszük:

1. A dimenziók vagy attribútumok szintjeiből egy bizonyos számú ajánlatot képezünk (más néven koncepciónak nevezzük őket)
2. Azokat vagy egyesével vagy egymással versenyeztetve megmutatjuk a válaszadóknak
3. És arra kérjük őket, hogy mondják meg melyik ajánlatot (vagy ajánlatokat) preferálják és esetleg azt is, hogy milyen mértékben.

A válaszadó által ilyen módon adott válaszok alapján a későbbiekben valamilyen módszer segítségével azt fogjuk vizsgálni, hogy a nyertes ajánlatok (azaz amelyeket a válaszadók preferáltak) miben különböznek a vesztes ajánlatoktól:

1. A nyertes ajánlatokban melyek azok a szintek, amelyek nagyobb valószínűséggel előfordulnak (és egyúttal nagyobb valószínűséggel hiányoznak a vesztes ajánlatokból)
2. És melyek azok a szintek, amelyek tipikusan inkább a vesztes ajánlatokban szerepelnek (és egyúttal jellemzően hiányoznak a nyertes ajánlatokból)

Elmondható tehát, hogy minden conjoint módszer esetén a szintek részhasznossági értékeit és a dimenziók fontosságát az alapján származtatjuk, hogy *egy bizonyos számú szintkombinációról vagy más néven ajánlatról összességében kikérjük a válaszadók véleményét*. Nem véletlen, hogy a conjoint technika innen kapta a nevét is, ugyanis a conjoint elnevezés az angol *consider jointly* szókapcsolatból származik.

De akkor miben is különböznek egymástól ezek a módszertanok? A továbbiakban röviden áttekintjük ezeket.

Conjoint value analysis (CVA)

Ez a technika ma már a klasszikus conjointnak számít. Egy olyan viszonylag egyszerű módszer, ahol a válaszadó egy többfokú (Likert) skála segítségével jelöli meg, hogy az aktuális ajánlatot mennyire valószínű, hogy megvásárolná vagy 2 egymással versengő ajánlat közül melyiket vásárolná meg inkább. Ennek a módszertannak ugyanis további két altípusa létezik annak megfelelően, hogy egyidejűleg hány ajánlatot lát a válaszadó.

A korai conjoint alapú kutatásokban egyidejűleg mindig 1 darab ajánlatot látott a válaszadó és egy többfokú vásárlási hajlandóság skálán jelölte meg azt a kategóriát, amely a legjobban leírta a vásárlási szándékát azzal kapcsolatban (12.4. ábra). Természetesen egymás után minden válaszadó több ajánlatot látott.

12.4. ábra: Egyetlen értékelendő ajánlat – CVA

Ár:	990 forint
Szállítás teljesítésének ideje:	Harmadnap
Szállítás módja:	Választott cím, kétórás időintervallumban
Visszaküldési lehetőség:	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)
	Biztosan nem venném igénybe
	Biztosan igénybe venném
	1 2 3 4 5 6 7

Ennek a megközelítésnek a szépsége az egyszerűségében rejlik. Az ajánlatok generálása a dimenziók szintjeiből viszonylag könnyen megoldható, illetve az eredmények feldolgozása is egyszerű, hiszen dummy változók lineáris regressziójaként származtathatók a szintek részhasznossági értékei. Sőt a válaszadói szintű modellezésnek sincs akadálya, hiszen a lineáris regresszió viszonylag kis megfigyelésszám esetén is jól működik. A módszer hátránya ugyanakkor, hogy a sokadik ajánlatnál a válaszok érvényessége megkérdőjelezhető, ugyanis nagyon monotonná, így unalmasá válik a válaszadók feladata, így félő, hogy nem valós viselkedést mérünk.

A CVA technikának létezik egy népszerűbb válfaja, mégpedig az a módszer, amikor a válaszadó egyidejűleg két ajánlatot lát és azt a kérdést kell megválaszolni, hogy melyik ajánlatot vásárolná meg inkább egy szintén többfokozatú Likert skálán. Itt is több ajánlatpárral szembesül minden válaszadó, hiszen egy kártya-

pár alapján nem lenne beazonosítható, hogy milyen logika alapján hozza meg a vásárlási döntéseit. A módszerre példákat a könyv korábbi részeiben már láthatunk, de a teljesség kedvéért itt is bemutatunk egyet (12.5. ábra).

12.5. ábra: Egy kártyapár – CVA

Ár:	990 forint	Ingyenes
Szállítás teljesítésének ideje:	Harmadnap	Legalább 5 nap, de választott szállítási dátum
Szállítás módja:	Választott cím, kétórás időintervallumban	Választott cím, hétköznapokon (9:00 és 18:00 között)
Visszaküldési lehetőség:	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)
	Egyértelműen a bal webshopot választanám	Egyértelműen a jobb webshopot választanám
	1 2 3	4 5 6 7

Ez a módszer is viszonylag egyszerűen használható. Az ajánlatpárok előállítása kissé bonyolultabb, mint a szimpla ajánlatos CVA esetén, ugyanakkor a részhassznossági értékek meghatározása ugyanúgy dummy változók lineáris regressziójával történik, így a válaszadószintű elemzés lehetősége itt is adott. A páros CVA előnye az egyidejűleg szimpla ajánlatot tesztelő megközelítéshez képest abban rejlik, hogy a sokadik döntési feladatnál is valószínűleg pontosabban méri a válaszadói preferenciát, hiszen mindig két alternatíva közötti választásra kényszeríti a válaszadót. Ugyanakkor vegyük észre, hogy a válaszadó által adott válasz sosem abszolút vásárlási hajlandóság lesz, hanem mindig egy valamely másik alternatívával szembeni erejét fogja csak mutatni egy tetszőleges ajánlatnak, ha az abszolút értékekben vagyunk érdekeltek, akkor nem ez a módszer a megfelelő.

Choce-based conjoint (CBC)

A CBC a jelenleg legszélesebb körben használt, így legnépszerűbb conjoint technika. A legfőbb jellemzője az, hogy elengedjük a többfokú skála használatát és az egyidejűleg mutatott több ajánlat (minimum 2) közül a válaszadónak ki kell választania azt az egyet, amelyiket legnagyobb valószínűséggel megvásárolná. Amennyiben a válaszadó egy döntési feladatban nem talál olyan alternatívát, amelyik kellően vonzó lenne számára, akkor dönthet úgy, hogy az „Egyiket sem venném meg” lehetőséget választja. CBC esetében egy tipikus döntési feladatot a 12.6. ábra mutatja.

12.6. ábra: Egy tipikus döntési feladat – CBC

Ár:	990 forint	Ingyenes	Ingyenes hűség-programmal	Egyiket sem venném igénybe
Szállítás teljesítésének ideje:	Harmadnap	Legalább 5 nap, de választott szállítási dátum	Legalább 5 nap, de választott szállítási dátum	
Szállítás módja:	Választott cím, kétórás időintervallumban	Választott cím, hétköznapokon (9:00 és 18:00 között)	Csomag-automata (0-24 órában)	
Vissza-küldési lehetőség:	Ingyenes hűség-programmal, a kereskedelmi lánc üzletében (nyitvatartási időben)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomag-ponton (nyitvatartási időben)	990 Ft csomagküldő-szekrényben (0-24)	
	○	○	○	○

A CBC jóval komplexebb módszer, mint a CVA megközelítés. A conjoint döntési feladatok generálása is lényegesen nagyobb kihívás, mint amivel a legfeljebb két alternatívával operáló CVA esetén találkozunk. Az úgynevezett kísérleti design (döntési feladatok) létrehozására többféle stratégia közül lehet választani, melynek részleteibe ennek a könyvnek a keretei között nem mélyülünk el. A részhasznossági értékek számításához elsődlegesen kondicionális logisztikus regressziót használhatunk, amely érzékeny a kis esetszámokra. Ennélfogva válaszadói szintű elemzésre ennek használata esetén nincs lehetőség, ha arról mégsem akarunk lemondani, akkor jóval szofisztikáltabb módszerekhez kell folyamodnunk. Ez utóbbira a *hierarchikus bayesi modell* (HB) vagy a *latent class analysis* (LCA) kínál lehetőséget, amelyeket a jelen könyv keretei között nem tárgyalunk.

A CBC előnye abban rejlik, hogy az interjúk során mutatott döntési szituációk tényleg realizisztikusak, jól közelítik a valós vásárlási helyzetet. A vásárlók a boltban is hasonló szituációval szembesülnek; több alternatíva közül kell kiválasztani azt, amelyiket megvennének, és ne feledjük, hogy a valós vásárlási döntés is bináris változó (megveszi / nem veszi meg) és nem többfokú skálával leírt esemény.

Adaptív conjoint módszerek

Az adaptív módszerek közös jellemzője az, hogy a kérdőív összeállításakor nem tudjuk még pontosan megmondani, hogy az egy ajánlatokon milyen szintek fognak szerepelni, hiszen azok mindig attól függenek, hogy az interjú korábbi fázisában milyen válaszokat adott a válaszadó.

Az adaptív conjoint tipikusan egy szint és esetleg dimenzió előszelektáló kérdéskörrel indul (sokszor klasszikus direkt kérdésekkel, azaz mi vonzó és mi nem, mi fontos és mi nem a válaszadó részére), annak érdekében, hogy beazonosítsuk a valóban releváns szintek (és dimenziók) körét. Erre azért van szükség, mivel az adaptív conjointot akkor szokás használni, ha nagyon komplex a vizsgálandó termék- vagy szolgáltatáskategória, azaz a bevonandó dimenziók és szintek száma magas. Ilyen esetekben érdemes eliminálni azokat a szinteket, amelyek nagy valószínűséggel nem jelentenének reális alternatívát a vásárlónak és csak a valóban lényeges elemekre koncentrálni annak érdekében, hogy az interjú hossza ésszerű kereteken belül maradjon.

Az adaptív conjoint módszerek sajátja, hogy az interjúkhoz speciális szoftvereket igényelnek, hiszen egy algoritmus igazítja már a kérdézési folyamat során a mindenkori válaszadó preferenciájához a vizsgálandó szintek (és esetleg dimenziók) körét. Ezen algoritmusok komplexitása meghaladja a jelen könyv kereteit, így ezeket nem tárgyaljuk.

További conjoint módszerek

A conjointnak mára számos további alfaját fejlesztették ki, melyek kiindulási alapja elsősorban a CBC technika, ami annak tükrében nem meglepő, hogy ez a módszer képezi le a legrealisztikusabban a valós vásárlási helyzeteket. Az eddigiekben az úgynevezett *full profile* módszereket tárgyaltuk, amelyek jellemzője, hogy egy-egy döntési feladatban a kutatásba bevont valamennyi dimenzió egyidejűleg megjelenik. Ugyanakkor létezik olyan megközelítés is (*partial profile*), amely minden döntési feladatban a bevont dimenzióknak csak egy részét mutatja meg, de feladatról feladatra mindig változik, hogy pontosan mely dimenziók szerepelnek. Ennek előnye abban rejlik, hogy egyszerűsíti a döntési szituációt a válaszadók számára nagyszámú dimenzió esetében, mégis minden dimenzióra tudunk eredményeket prezentálni.

Ugyancsak a CBC továbbfejlesztéséből született az *allokáción alapuló conjoint*, amikor a válaszadó nem csupán egy alternatívát választ döntési feladatonként, hanem akár több alternatívát is megjelölhet vásárlási szokásainak megfelelően, sőt eltérő mennyiséget is hozzárendelhet az egyes ajánlatokhoz, ha a való életben is így döntene egy vásárlási helyzetben. Ennek a megközelítésnek az előnye abban rejlik, hogy a valós vásárlásokat jobban szimuláló döntési szituációba helyezi a válaszadókat, amely eredményeként a conjoint elemzés segítségével modellezett piac is jobban közelíti a valóságot.

Ezen változatok ugyanakkor túlmutatnak az SPSS keretein, így a jelen könyvben mélyebben nem tárgyaljuk őket.

A kísérleti design előállítása

Ebben a részben áttekintjük, hogy hogyan alkothatók meg az interjúk során a válaszadókkal értékelendő ajánlatok, vagy koncepciók és mire kell figyelniünk, amikor megalkotjuk őket. A legegyszerűbb conjoint módszerre, a korábban bemutatott CVA technikára fókuszálunk, amely az SPSS és MS Excel segítségével is megalkotható.

A kísérleti design legfőbb jellemzői

Ha a conjoint technika mellett döntöttünk a kutatásunkban, akkor az első lépés a megvalósítás útján az úgynevezett kísérleti design (*experimental design*) előállítása lesz. Ez lényegében azt jelenti, hogy meg kell alkotunk azokat az ajánlatokat vagy koncepciókat a kutatásba bevont dimenziók szintjeiből, amelyeket az interjúk során értékeltetünk a válaszadóinkkal.

A kísérleti design előállítása nagy figyelmet és precizitást igényel. Ezzel ugyanis meghatározzuk a conjoint modellünk a paramétereinek körét, illetve az azok közti korrelációt, amely nagy mértékben befolyásolja az eredményeinket, tehát akár nem kívánatos torzítást is vihetünk a modellünkbe. Az egyik legfontosabb szabály a kísérleti design összeállításakor az, hogy a szintekből nem csak a valós piacon aktuálisan elérhető ajánlatokat mutatjuk meg a válaszadóinknak, hanem bármely tetszőleges, akár nem létező ajánlatot is. Gyakran megesik, hogy a kísérleti designban szereplő ajánlatok többsége nem is kapható, vagy vehető igénybe a valóságban, ugyanakkor ez teljesen helyénvaló. Ezzel tesszük ugyanis lehetővé egyrészt azt, hogy a szintek részhasznossági értékeit pontosan mérni tudjuk, illetve azt, hogy piaci szimulációban a szintekből kirakható tetszőleges helyzetet fogunk tudni modellezni.

Ugyancsak lényeges, hogy bizonyos esetekben egyes szintkombinációk vagy párosítások kivehetőek a kísérleti designból, ha azok például nyilvánvaló helyzetet teremtenének (ilyen például az az eset, ha a legmagasabb szolgáltatásszint a legalacsonyabb áron jelenne meg egy ajánlaton), vagy olyan kombináció bukkan fel a kérdőívben, ami a gyakorlatban nem valósítható meg (például letöltendő szoftver postai úton történő kézbesítéssel). Nagyon fontos, hogy ezek az úgynevezett letiltások az összes kombináció marginális részét képezhetik, amelyek csak így nem gyakorolnak jelentős hatást a kísérleti design hatékonyságára. Amennyiben túl sok párosítást kellene tiltanunk, abban az esetben az a helyzet húzódhat meg a háttérben, hogy a szóban forgó dimenzió nem független egymástól, ami a hagyományos conjoint alapjaiba ütközik, hiszen itt a dimenziók függetlenségét feltételezzük. Ebben az esetben a megoldás a dimenziók összevonása lehet, azaz a két dimenzióból képezünk egy olyat, amelynek szintjei a korábbi két dimenzió „értelmezhető” kombinációi lesznek.

Ha CVA designt alkotunk, akkor a kísérleti designnak alapvetően 2 matematikai feltételnek kell megfelelniük annak érdekében, hogy hatékonyan legyen képesek mérni a válaszadói preferenciákat (Kuhfeld, W.F., 1997):

1. **Kiegyensúlyozottság:** ez azt jelenti, hogy a megalkotott ajánlatokban az egyes dimenziók szintjei azonos valószínűséggel kell, hogy előforduljanak
2. **Ortogonalitás:** ez pedig azt jelenti, hogy egyidejűleg 2 dimenziót vizsgálva bármely szint párosításnak ugyanolyan valószínűséggel kell előfordulnia, azaz a paramétereknek korrelálatlannak vagy másképp függetlenek kell lenniük

A fentiek teljesülésének mértékét kétféleképpen vizsgálhatjuk:

1. Olyan mutatószám segítségével, amely a kísérleti design összességében (valamennyi dimenziót és azok szintjeit egyidejűleg figyelembe véve) jellemzi. Erre a leggyakrabban használt módszer az úgynevezett *D-efficiency* mutató, melynek képlete az alábbi:

$$D - efficiency = \frac{|X^T X|^{1/p}}{N}$$

Ahol:

- az X a szintekből dummy változókkal képzett úgynevezett információs mátrix,
- a p a szintek összesített száma mínusz a dimenziók száma és hozzáadva 1-et,
- az N pedig a döntési feladatok száma (azaz hány darab kérdést válaszol meg a válaszadónk) az interjú során.

A fentebb definiált *D-efficiency* mutató egy normált mérőszám, amelynek értéke minden esetben 0 és 100 közé esik. A 100 jelenti a tökéletes ortogonalitást, a 0 pedig az ortogonalitás teljes hiányát. A 100-as érték a gyakorlatban sosem érhető el, sőt az esetek többségében meg sem közelíthető. A mérőszám igazából 2 kísérleti design összehasonlítására alkalmas, azaz abban tud segíteni, hogy eldöntsük melyik kísérleti design a jobb, és mennyivel jobb a másiknál.

2. Gyakorisági táblák és keresztábrák segítségével vizsgáljuk meg azt, hogy a szintek dimenzió belüli egyenletes eloszlása teljesül-e, illetve két tetszőleges dimenzió szintpárosításai azonos gyakorisággal fordulnak-e elő a kísérleti designunkban.

Összességében a gyakorlati szemszögből a 2. megközelítés teljesen jól használható a design hatékonyságának vizsgálatára és minősítésére. MS Excel segítségével a D-efficiency mutató is könnyen előállítható, de abszolút értelemben való minősítésre a gyakorisági és keresztátlák jobban használhatók.

Fontos kiemelni, hogy a design megalkotása manuálisan is lehetséges, de igen időigényes és komplex feladat. Erre a célra léteznek célszoftverek (Sawtooth Software eszközei a legnépszerűbbek), amelyek rendkívül felhasználóbarát módon csak a kutatásunk paramétereit (hány dimenziónk és azoknak hány szintje van) és a döntési feladatok számának meghatározását kérik, majd generálnak nekünk egy kísérleti designt és minősítik is azt. Az SPSS szoftver csak egy ajánlatos CVA megközelítéshez szükséges kísérleti design megalkotására képes és a döntési feladatok száma is csak limitált kereteken belül állítható. Amennyiben csak az SPSS áll rendelkezésünkre és páros ajánlatos CVA modellt szeretnénk építeni, akkor csak kiindulási pontnak jó az SPSS, de még sok manuális munka vár ránk az MS Excelben ahhoz, hogy megfelelő kísérleti design álljon a rendelkezésünkre.

Az webshopos esettanulmányunkban páros ajánlatos CVA módszert használtunk 18 döntési feladattal. Ez azt jelenti, hogy 36 darab ajánlatot vizsgáltunk. Ha a páros ajánlatos CVA megközelítés mellett döntünk, akkor a párok kialakításánál figyelniünk kell az alábbiakra:

1. Bármely ajánlat csak egyszer forduljon elő (ez igaz a szimpla ajánlatos CVA módszerre is),
2. illetve az ajánlatpárok minden dimenzióban különböző szinteket mutassanak, azaz döntési feladaton belül nincs szint ismétlődés az ajánlatokon.

Ez utóbbinak az az oka, hogy azon dimenzió tekintetében, amelyek azonos szinttel szerepelnek a döntési feladatban, semmit sem tanulunk a válaszadó preferenciáiról, és a döntési feladatok száma általában túl kevés ahhoz, hogy ilyen luxust megengedjünk magunknak.

Az esettanulmányuk kísérleti designjának D-efficiency mutatószáma *43,19* volt, ami elég messze esik a 100-as maximum értéktől. A szintek gyakorisági tábláit, illetve a keresztátlákat megvizsgálva viszont azt láthatjuk, hogy a design hatékonysága megfelelő volt.

Ár

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 990 forint	9	25,0	25,0	25,0
Ingyenes	9	25,0	25,0	50,0
Ingyenes hűségprogrammal	9	25,0	25,0	75,0
Ingyenes kiszállítás 15.000 forint fel	9	25,0	25,0	100,0
Total	36	100,0	100,0	

Szállítás teljesítésének ideje

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Harmadnap	8	22,2	22,2	22,2
Legalább 3 nap, de választott szállítási	10	27,8	27,8	50,0
Legalább 5 nap, de választott szállítási	9	25,0	25,0	75,0
Másnap	9	25,0	25,0	100,0
Total	36	100,0	100,0	

Szállítás módja

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Csomagautomata (0-24 órában)	9	25,0	25,0	25,0
Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási idő)	9	25,0	25,0	50,0
Választott cím, hétköznapokon (9:00 és 18:00 között)	9	25,0	25,0	75,0
Választott cím, kétórás időintervallumban	9	25,0	25,0	100,0
Total	36	100,0	100,0	

Visszaküldési lehetőség

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 990 Ft csomagküldőszekrényben (0-24)	10	27,8	27,8	27,8
Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időb	8	22,2	22,2	50,0
Ingyenesen csomagküldőszekrényben (0-24 órában)	9	25,0	25,0	75,0
Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási idő	9	25,0	25,0	100,0
Total	36	100,0	100,0	

Ár * Szállítás teljesítésének ideje Crosstabulation

Count

	Szállítás teljesítésének ideje				Total
	Harmadnap	Legalább 3 nap, de választott szállítási	Legalább 5 nap, de választott szállítási	Másnap	
Ár 990 forint	3	2	2	2	9
Ingyenes	1	3	3	2	9
Ingyenes hűségprogrammal	2	3	2	2	9
Ingyenes kiszállítás 15.000 forint fel	2	2	2	3	9
Total	8	10	9	9	36

Ár * Szállítás módja Crosstabulation

Count

	Szállítás módja				Total
	Csomagautomata (0-24 órában)	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási idő	Választott cím, hétköznapok (9:00 és 18:00 között)	Választott cím, kétórás időintervallumban	
Ár 990 forint	2	2	2	3	9
Ingyenes	2	2	3	2	9
Ingyenes hűségprogrammal	3	3	1	2	9
Ingyenes kiszállítás 15.000 forint fel	2	2	3	2	9
Total	9	9	9	9	36

Ár * Visszaküldési lehetőség Crosstabulation

Count		Visszaküldési lehetőség				Total
		990 Ft csomagküldő szekrényben (0-24)	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időb)	Ingyenesen csomagküldő szekrényben (0-24 órában)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időb)	
Ár	990 forint	2	2	3	2	9
	Ingyenes	3	2	2	2	9
	Ingyenes hűségprogrammal	2	2	3	2	9
	Ingyenes kiszállítás 15.000 forint fel	3	2	1	3	9
Total		10	8	9	9	36

Szállítás teljesítésének ideje * Szállítás módja Crosstabulation

Count		Szállítás módja				Total
		Csomagautomata (0-24 órában)	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időb)	Választott cím, hétköznapokon (9:00 és 18:00 között)	Választott cím, kétórás időintervallumban	
Szállítás teljesítésének ideje	Harmadnap	2	2	2	2	8
	Legalább 3 nap, de választott szállítási	3	1	3	3	10
	Legalább 5 nap, de választott szállítási	2	3	2	2	9
	Másnap	2	3	2	2	9
Total		9	9	9	9	36

Szállítás teljesítésének ideje * Visszaküldési lehetőség Crosstabulation

Count		Visszaküldési lehetőség				Total
		990 Ft csomagküldő szekrényben (0-24)	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időb)	Ingyenesen csomagküldő szekrényben (0-24 órában)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időb)	
Szállítás teljesítésének ideje	Harmadnap	1	3	2	2	8
	Legalább 3 nap, de választott szállítási	3	2	3	2	10
	Legalább 5 nap, de választott szállítási	3	2	2	2	9
	Másnap	3	1	2	3	9
Total		10	8	9	9	36

Szállítás módja * Visszaküldési lehetőség Crosstabulation

Count		Visszaküldési lehetőség				Total
		990 Ft csomagküldő szekrényben (0-24)	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időb)	Ingyenesen csomagküldő szekrényben (0-24 órában)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagpont n (nyitvatartási idő)	
Szállítás módja	Comagautomata (0-24 órában)	3	1	3	2	9
	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási idő)	2	2	2	3	9
	Választott cím, hétköznapokon (9:00 és 18:00 között)	3	2	2	2	9
	Választott cím, kétórás időintervallumban	2	3	2	2	9
Total		10	8	9	9	36

Az egyváltozós gyakorisági táblák alapján a kiegyensúlyozottság tekintetében azt a következtetést vonhatjuk le, hogy a szintek nagy többsége minden dimenzióban 9 alkalommal jelenik meg, 1-2 esetben látunk 8-as vagy 10-es gyakoriságot. Az ideális érték minden dimenzió minden szintje esetén 9 lenne, hiszen 36 ajánlatunk és minden dimenzióban 4 darab szintünk ($36/4=9$). A 8-as és 10-es egyváltozós gyakoriságok előrevetítik azt is, hogy az ortogonalitás sem lesz tökéletes. Bár nagyon fontos, hogy ebben az esetben tökéletes ortogonalitás eleve nem is érhető el, hiszen 36 ajánlatunk van és minden dimenzióban 4-4 szintünk, így a keresztábláknak 16 cellájuk van. Az elvárt gyakoriság pedig $36/16 = 2,25$ lenne, ami azt jelenti, hogy a 2-es és 3-as gyakoriság tekinthető tökéletesnek, hiszen csak egész számokban lehet gondolkodni. A tipikus érték a keresztáblák celláiban a 2 és a 3, de bizonyos esetekben 1 is előfordul, de ezen cellák számossága kifejezetten alacsony. Ne feledjük, hogy volt egy komoly megkötésünk az ajánlatokkal kapcsolatban: *minden ajánlatpárnál a két ajánlatnak minden dimenzióban eltérő szintet kell felvennie*. Ha ezt a feltételt elengednénk, akkor valószínűleg könnyebb lenne a tökéletes kiegyensúlyozottságot és az ortogonalitást közelíteni, bár akkor sem tudnánk teljesíteni tökéletesen a két feltételt. Összességében a jelen kísérleti design a lehetőségeket figyelembe véve semmi esetre sem tekinthető rossznak.

A döntési feladatok számának meghatározása

A kísérleti design kialakításánál a másik fontos szempont az az, hogy hány *döntési feladatban* kell részt vennie a válaszadónak, vagyis hány ajánlatról vagy ajánlatpárról kell véleményt nyilvánítania a válaszadónknak preferencia tekintetében.

A tradicionális CVA esetén a korábbiakban leírtaknak megfelelően válaszadói szinten készítünk egy-egy modellt és ha lineáris regresszióval határozzuk meg a szintek részhasznossági értékeit, akkor a döntési feladatok számának meghatározásakor a lineáris regresszió vonatkozó elvárásainak kell megfelelnünk. A lineáris regresszió használatakor az az általános elvárás, hogy *a megfigyelések száma legyen a többszöröse a becülendő paraméterek számának*. Ez pontosan mit is jelent a CVA esetén? A becülendő paraméterek számát az alábbi képlettel határozhatjuk meg:

Paraméterek száma = a dimenziók szintjeinek száma összesen-dimenziók száma+1

A képlet mögötti logika alapja az, hogy a kísérleti designt (azaz a dimenziók szintjeit) dummy változókként kódoljuk, amelyek tipikusan 0 és 1 értéket vesznek fel. A dummy változók képzésénél úgy járunk el, hogy annyi új változót hozunk létre, amennyi a vizsgált kiindulási változó kategóriájának száma mínusz egy. Azért kell egy változóval kevesebbet létrehozni, mert ellenkező esetben a multikollinearitás szélsőséges esete állna fenn, azaz a független változók nem lennének függetlenek. Mivel minden dimenzió szintjeit ezen az elven kódolunk, így minden dimenziónál egy szintet el kell hagynunk (azaz pontosan annyi szintet hagyunk el, ahány dimenzió van). A végén 1-et adunk hozzá a becülendő paraméterek számához, mert az úgynevezett tengelymetszetet vagy másképp a lineáris regresszió konstans paraméterét is becülni fogjuk, ami pont 1-gyel növeli a paraméterek számát.

A becsléshez szükséges megfigyelésszám pedig az alábbiak szerint alakul:

Szükséges megfigyelésszám = paraméterek száma \times N

Az *N* értéke ideális esetben minimum 3 lenne (Sawtooth Software CVA felhasználói kézikönyv), ami általában nehezen tartható. A webshopos esettanulmányunknál (12.1. táblázat) 4 dimenzió van, ahol mindegyik dimenzió 4 szinttel bír. A becülendő paraméterek száma tehát:

$$4 \times 4 - 4 + 1 = 13$$

A szükséges megfigyelésszám különböző *N*-ek esetén pedig az alábbiak szerint alakul:

12.6. táblázat

N	Megfigyelés szám
3	39
2	26
1,5	19-20

Ha ragaszkodnánk az elméleti szempontból optimális 3-as szorzóhoz, akkor minden egyes válaszadónak 39 esetben kellett volna eldöntenie, hogy az ajánlat-párok közül melyiket választaná inkább. Nagyon lényeges ugyanis, hogy a páros ajánlatos CVA esetében egy ajánlatpár közötti döntés számít egy megfigyelésnek és nem a mutatott ajánlatok száma adja a megfigyelésszámot (azaz 36 ajánlatot látott minden válaszadó, de csak 18 döntést hozott). A 39 döntési feladat időben is sok és vélhetően a válaszadó rátanulna a feladatok logikájára, így nem igazán a valós fogyasztói viselkedését figyelnénk meg, hanem az interjú során „betanult” magatartását. Ha 2-es szorzót alkalmaznánk, még akkor is 26 kérdést kellene megválaszolni, ami még mindig nagyon sok. A gyakorlati tapasztalat azt mutatja, hogy a döntési feladatok számát érdemes 20 alatt tartani, de ideálisan maximum 15-ig menjünk el. Ez viszont azt jelenti, hogy 1,5-es szorzót kell alkalmaznunk, ami elméleti szempontból messze nem ideális, de kérdezéstechnikai szempontból már egész vállalható hosszúságú interjút tesz lehetővé.

Az esettanulmányunkban 4 dimenziót vizsgálunk, ahol minden dimenziónak 4 szintje van, ami nem jelent egy túl bonyolult modellt, a megfigyelések számában már kompromisszumra kényszerülünk. Ha egyedi szinten akarjuk becsülni a szintek részhasznossági értékeit és mindezt lineáris regresszióval akarjuk megtenni, akkor nagyon gyorsan elkezdjük feszegetni a határokat. Mit tudunk tenni akkor, ha a jelenleginél jóval komplexebb modellt szeretnénk építeni?

1. Elengedjük a válaszadói szintű elemzést és a teljes mintán vagy előre meghatározott almintákon építünk egy-egy regressziós modellt
2. Nem lineáris regressziót, hanem *hierarchikus bayesi modellt* használunk a részhasznossági értékek becslésére. A *hierarchikus bayesi modell* úgy képes adni „kvázi” válaszadói szintű becslést, hogy sokkal megengedőbb a megfigyelésszám és paraméter arányában. Ugyanakkor az SPSS-ben jelenleg nem elérhető ilyen technika
3. Adaptív conjointra váltunk, amely minden válaszadónál beazonosítja az igazán releváns dimenziókat és szinteket, és csak arra fókuszálva alkalmazza a klasszikus *trade-off* kérdéseket

Összességben egy tradicionális CVA modell határa a 4-5 dimenzió, dimenzióként 3-4 szinttel. Arra a jelen esettanulmány a példa, hogy ezt a határt mennyire egyszerű elérni.

Felmerülhet a kérdés, hogy véletlenszám függvényrel (például az MS Excel *VÉLETLEN.KÖZÖTT()*) függvényével generálható-e kísérleti design? Amennyiben minden válaszadó ugyanazt az ajánlatszettet látja, akkor nem javasolt ennek a megközelítésnek a használata, ugyanis ilyen kis bázison (ugye legfeljebb $2 \times 20 = 40$ ajánlatról beszélünk) nem garantált a kiegyensúlyozott és közel ortogonális design létrejötte pusztán véletlenszám generátor használatával. Ezt a megközelítést legfeljebb akkor alkalmazhatjuk, ha nagy a mintánk (például 1000 fő már nagy mintának tekinthető), minden válaszadó más és más kísérleti designnal szembesül, illetve nem akarunk válaszadói szinten becsülni a részhasznossági értékeket. Viszont ebben az esetben is érdemes némi kontrollt alkalmazni a véletlenszámok felett, ugyanis elkerülendő az a szituáció, hogy az egy döntési feladatban szereplő ajánlatok azonosak legyenek vagy nyilvánvaló döntésre kérjük a válaszadót.

A CBC kísérleti design sajátosságai

Ha a CBC módszertant választjuk eszközzül, akkor több szempontot kell mérlegelni a kísérleti design összeállításakor és kevesebb szigorú szabálynak kell megfelelnünk. Egyrészt amiatt, hogy a CBC esetén tisztán válaszadói szinten nem tudunk részhasznossági értékeket becsülni, így a megfigyelések számára vonatkozólag nincs olyan szigorú szabály, mint amivel a CVA esetében találkoztunk. Arra persze oda kell figyelni, hogy a legkisebb elemzési egység (alminta) szintjén az az elvárás teljesüljön, hogy a megfigyelési feladatok száma (ami itt a döntési feladatok számát jelenti) nagyságrendekkel legyen nagyobb, mint a becsülendő paraméterek száma).

Ami a legfontosabb különbség a CVA modellhez képest, hogy a CBC-nél tipikusan többféle kísérleti designt, azaz kérdőívváltozatot használunk a mintán úgy, hogy minden kérdőívváltozattal közel azonos számú válaszadó találkozzon. A kérdőívváltozatok tehát értelemszerűen csak abban különböznek egymástól, hogy az egyes döntési feladatokban eltérő ajánlatok szerepelnek. Így biztosítjuk azt, hogy a lehetséges szintkombinációkból minél többet fedjen le az adatfelvételünk. A CVA módszernél a többféle kérdőívváltozatnak általában nincs relevanciája abban az esetben, ha lineáris regressziót használunk válaszadói szinten, mert akkor minden válaszadónk részhasznossági értékeinek becslésére kizárólag a saját válaszait használjuk, így semmi előnyünk nem származik a többféle kérdőívből. CBC esetén viszont még abban az esetben is, ha becslést adunk a válaszadói részhasznossági értékekre, minden válaszadó modelljében többé kevésbé benne van teljes mintánk válaszai is, így előnyös, ha több kérdőívváltozatot használunk.

CBC esetén annyi elvárás azért van, ha rendelkezésünkre áll az *LCA* vagy *hierarchikus bayesi modell* a részhasznossági értékek „kvázi” egyéni szintű becslésére, hogy minden dimenzió minden szintje 3-4 alkalommal előforduljon a kísérleti designban. Viszont nagyon fontos, hogy a szintek előfordulása CBC esetén nem csupán a döntési feladatok számának növelésével érhető el, hanem úgy is, ha az egy döntési feladatban szereplő ajánlatok számát növeljük. Emlékezzünk vissza, hogy CBC esetén nem csupán 2 ajánlatot mutathatunk egy-egy döntési feladatban, hanem többet, sőt sokszor jóval többet. Valamint fontos megjegyezni, hogy kifejezetten nem javasolt csupán 2 ajánlatot szerepeltetni egy-egy kérdésnél, hiszen ebben az esetben nagyon keveset tanulnánk egy-egy döntési feladat eredményéből. Az ajánlatok komplexitása (a bevont dimenziók száma) természetesen határt szab az egy döntési feladatban szereplő alternatívák számának, illetve az is kerülendő, hogy egyszerre egy-egy dimenzió összes szintjét megjelenítsük a döntési feladatokban. A kísérleti design kiegyensúlyozottsága és ortogonalitása itt is olyan cél, amire törekednünk kell a pontos mérés érdekében. Illetve a 15 döntési feladatot, mint ideális felső határt itt is érdemes megtartani.

Amennyiben csak a teljes mintán vagy előre definiált alcsoportokon fogunk részhasznossági értékeket becsülni kondicionális logisztikus regresszió segítségével és nagy minta (például 1000 fő) áll rendelkezésünkre, akkor jóval könnyebb dolgunk van a kísérleti design megalkotására. Ebben az esetben a korábban említett véletlenszám generátor elfogadható eszköz abban az esetben, ha lehetőségünk van minden válaszadó számára egyedi kísérleti design létrehozni (erre általában akkor van lehetőség, ha CAPI vagy CAWI interjúkat készítünk). Természetesen a válaszadónkénti maximum 15 döntési feladat ebben az esetben is a javasolt felső határ az interjú hosszára.

Hold-out döntési feladatok

A modellünk előrejelző képességének pontosságát az úgynevezett *hold-out* kérdések segítségével tudjuk jellemezni, illetve ezen kérdések segítségével tudjuk finomhangolni a modellünket (például meghozhatjuk azt a döntést, hogy a *first choice* vagy *share of preference* megközelítést használjuk-e a piaci szimulációk során). A *hold-out* kérdések a conjoint elemzés során hasonló szerepet töltenek be, mint a prediktív modellezés esetén a *test sample*, csupán *hold-out* kérdések aránya tradicionálisan jóval alacsonyabb, mint a *big data* elemzésben megszoktuk. Ennek oka az, hogy a válaszadó figyelme és ideje nem korlátlan, így arányaiiban kénytelenek vagyunk több időt fordítani a conjoint modellünk betanítására szolgáló döntési feladatokra, amit a big data logikájánál maradván nevezhetünk *training* feladatoknak nevezünk.

A *hold-out* kérdések szintén olyan döntési feladatok, ahol a válaszadó a preferenciáját nyilvánítja ki. A döntés tárgyát képező ajánlatok (ami lehet egy vagy több) pedig ugyanúgy a kutatásba bevont dimenziókból és azok szintjeiből épülnek fel, mint a modell betanítására szolgáló döntési feladatokban. Fontos, hogy a CVA módszernél is alkalmazhatunk olyan *hold-out* kérdéseket, ahol nem csupán 2, hanem több ajánlat is szerepel egyidejűleg és nem többfokozatú skálán kérdezzük a preferenciát, hanem a válaszadó feladata annyi, hogy kiválassza azt az egy ajánlatot, amelyet legnagyobb valószínűséggel megvásárolna. A CVA modell eredményei alapján ezek a döntési feladatok is lemodellezhetők, azaz meg tudjuk határozni azt, hogy a *hold-out* kérdésekben szereplő alternatívákat milyen arányban választanák a válaszadóink, s végül a tény és modellezett adatok összevetésével jellemezhető a modellünk pontossága.

Tipikusan 2-3 *hold-out* kérdést szokás kérdezni a válaszadóktól, általában a modellépítéshez használt döntési feladatokkal vegyítve. Nagyon fontos, hogy a *hold-out* kérdés vagy kérdések minden válaszadónál ugyanazok még abban az esetben is, ha többféle kérdőívet használunk. Ennek oka az, hogy a modell előrejelző képességének tesztelését nem válaszadói szinten vizsgáljuk (és finomítjuk), hanem a minta egészén.

Egy ajánlatos CVA kísérleti design létrehozása SPSS segítségével

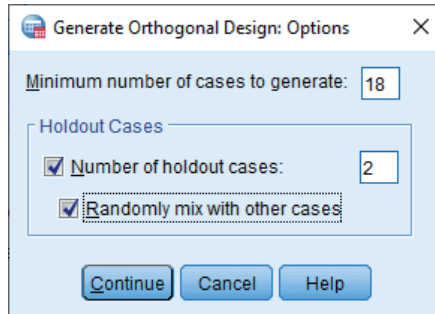
Az SPSS segítségével generálhatunk garantáltan ortogonális designt egy ajánlatos CVA modell számára, amennyiben rendelkezünk az SPSS *Conjoint* moduljával, amely külön megvásárolható a programcsomaghoz. A modul nem túl rugalmas, kevés beállítási lehetőségre ad lehetőséget (például bizonyos kombinációk letiltására nincsen lehetőségünk), viszont kiindulási pontnak egy kísérleti design megalkotásához alkalmas lehet. Ami nehézséget és többletmunkát okozhat az SPSS design generátora használatakor az az, hogy ez a modul a *kiegyensúlyozottságot* nem figyeli egy-egy design előállításakor, így ha ennek a feltételnek is meg akarunk felelni, akkor nem kevés többletmunkára számítsunk.

A design generálás modult az **SPSS Data** főmenü pontjában találjuk, mégpedig a beszédes **Orthogonal Design** almenüpont alatt. Itt két lehetőség áll rendelkezésünkre:

- *Generate*: itt tudjuk létrehozni a kívánt kísérleti designunkat, annak paramétereit megadva. A program a létrehozott designt egy külön adatbázisba vagy fájlba rakja
- *Display*: ennek segítségével meg tudjuk jeleníteni a megalkotott ajánlatokat. Gyakorlati haszna ma már kevés van, mivel manapság tipikusan CAWI vagy CAPI módszerekkel megy a terepmunka, ahol kérdőív szer-

kesztő és interjú menedzselő programokhoz általában más formátumban van szükség az ajánlatokra

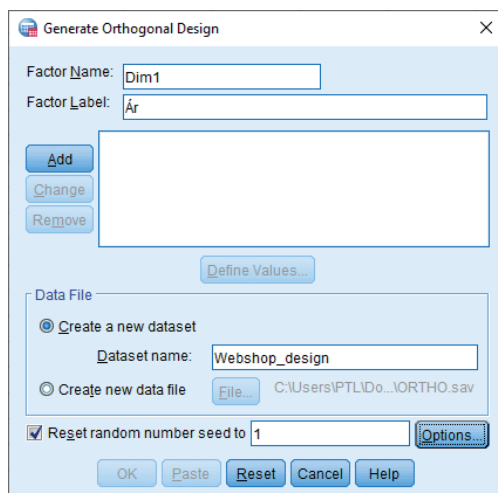
A *Generate* gombra kattintva egy új ablak nyílik, ahol meg tudjuk adni a kísérleti designunk paramétereit. Első körben kattintsunk az *Option* gombra, ahol a kívánt feladatok számát tudjuk meghatározni, illetve azt tudjuk beállítani, hogy hány darab *hold-out* kérdést szeretnénk kérdezni a válaszadótól.



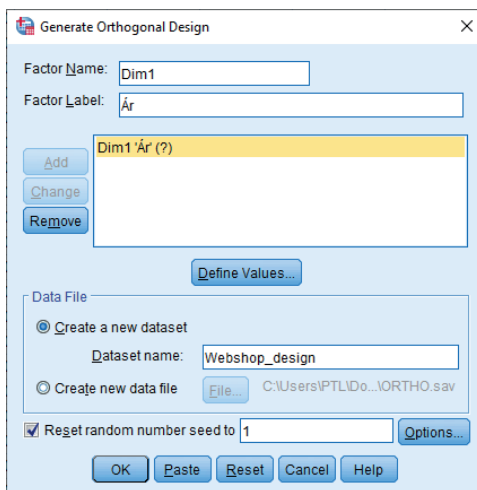
Az SPSS-ben a döntési feladatok számát csak hozzávetőlegesen tudjuk meghatározni, pontosabban azt a számot tudjuk megadni, hogy mennyinél ne legyen kevesebb a kérdezett koncepciók száma. Ha a mezőt üresen hagyjuk, akkor az SPSS azt az abszolút minimum ajánlatszámot fogja generálni, ami szerinte szükséges az adott kutatás végrehajtásához. Tételezzük fel, hogy a webshopos kutatásunkhoz szeretnénk design generálni, ahol a becsülendő paraméterek száma 13 volt. Így 18 ajánlat (ugye továbbra is 1 ajánlat döntési feladatonként) már elfogadható kompromisszumnak tűnik. A *hold-out* blokk opcionális, vagyis nem kötelező *hold-out* kérdéseket feltenni a válaszadónak. Ha mégis így döntünk és bejelöljük az opciót, akkor a *Number of holdout cases* szövegmezőben tudjuk specificálni a kérdések számát (ami minimum 2), illetve alatta tudjuk bejelölni, hogy ezeket a kérdéseket keverni akarjuk-e a többi kérdéssel vagy a modellépítéshez felhasznált kérdések (*training*) után szeretnénk azokat elhelyezni. Ha a *hold-out* kérdések használata mellett döntünk, akkor érdemes a vegyítés mellett dönteni. Ha befejeztük a paraméterezést, akkor a *Continue* gomb megnyomásával tudunk visszalépni a *Generate* menü főablakára.

A főablakban először állítsuk be azt, hogy új adatbázisba vagy inkább fájlba szeretnénk rakni a generált designt (*Data File* blokk). Majd érdemes a *Reset random number seed to* opciót aktiválni és beírni egy közel tetszőleges számot a mellett lévő szövegmezőbe. Ezzel azt érjük el, hogy reprodukálni tudjuk majd a későbbiekben a designunkat szükség esetén. Ha ezt nem kattintjuk be, akkor az

SPSS maga fog választani egy kiindulópontot a véletlenszám generáláshoz, de mindig más, ami miatt minden alkalommal más és más kísérleti designt hoz létre még az egyéb paraméterek változatlansága esetén is.



Ezt követően a *Factor Name* és a *Factor Label* mezőkben tudjuk megadni az egyes dimenziók nevét és szövegét. A *Factor Label* opcionális, abban az esetben érdemes feltölteni, ha az SPSS *Display* funkcióját is szeretnénk használni. Ha végeztünk az aktuális dimenzióknak specifikálásával, akkor kattintsunk az *Add* gombra, majd a létrehozott *Dim1* sort kijelölve a *Define Values* gomb segítségével tudunk átlépni a dimenzió szintjeinek meghatározására.



A *Define Values* ablakban vagy manuálisan megadjuk a szintek számát és szövegét, vagy az *Auto-Fill* lehetőséget használva az SPSS-re bízunk az adatok feltöltését. Megint csak akkor érdemes időt áldozni a *Label* adatok specifikálására, ha az SPSS *Display* funkcióját használni szeretnénk, a szintek szövegeinek megadása nem kötelező. A beállítások után itt is a *Continue* gombra kattintva véglegesíthetjük a paramétereket és léphetünk vissza a *Generate* főablakra.

	Value	Label
1:	1	Ingyenes
2:	2	Ingyenes kiszállítás 15.000 forint fe
3:	3	990 forint
4:	4	Ingyenes hűségprogrammal
5:		
6:		
7:		
8:		
9:		

Ha minden dimenzió minden szintjét specifikáltuk, akkor a *Generate Orthogonal Design* ablak *OK* gombjára kattintva indíthatjuk el a design generálást. Az SPSS a kérésünket felülbírálvá 27 darab ajánlatot (25 *training* és 2 darab *hold-out* ajánlatot) generál a frissen létrehozott *Webshop_design* adatbázisban. A dimenziók számának megfelelő *Dim* változónk van az adatbázisunkban, illetve két technikai mezőnk: *STATUS_* és *CARD_*. A *CARD_* egy egyszerű futóindex, amely az ajánlatok beazonosíthatóságát segíti, a *STATUS_* változó jelöli meg 1-es értékkel a *hold-out* kérdéseket, amelyeket a későbbiekben a modell előrejelző képességének jellemzésére, illetve a finomhangolásra használhatunk.

	Dim1	Dim2	Dim3	Dim4	STATUS_	CARD_
1	4,00	2,00	2,00	3,00	1	1
2	1,00	3,00	4,00	3,00	0	2
3	4,00	4,00	4,00	1,00	0	3
4	4,00	1,00	2,00	1,00	0	4
5	4,00	1,00	1,00	4,00	0	5
6	1,00	2,00	1,00	4,00	0	6
7	3,00	1,00	4,00	4,00	0	7
8	2,00	1,00	1,00	3,00	0	8
9	1,00	2,00	4,00	1,00	0	9
10	1,00	3,00	2,00	4,00	0	10
11	2,00	3,00	1,00	1,00	0	11
12	4,00	3,00	1,00	2,00	0	12
13	2,00	1,00	4,00	2,00	0	13
14	1,00	1,00	3,00	2,00	0	14
15	3,00	2,00	1,00	2,00	0	15
16	1,00	4,00	1,00	3,00	0	16
17	4,00	2,00	3,00	3,00	0	17
18	1,00	1,00	3,00	1,00	0	18
19	1,00	4,00	2,00	2,00	0	19

Az SPSS a design hatékonyságáról nem ad információt. Ha meg szeretnénk nézni, hogy a létrehozott designunk mennyiben felel meg a kiegyensúlyozottság, illetve ortogonalitás feltételének, akkor azt nekünk kell kiszámolni a korábban bemutatott képlet segítségével vagy gyakorisági táblák és kereszt táblák lekérésével.

Ha a létrehozott ajánlatok számát soknak találjuk, akkor véletlenszerűen vegyünk ki annyi rekordot, amennyit feleslegesnek ítélünk. Ugyanakkor ezt mindenképp úgy tegyük meg, hogy közben gyakorisági táblák és kereszt táblák futtatásával ellenőrizzük azt, hogy a kiegyensúlyozottság és ortogonalitás elvárásoknak többé kevésbé megfelelünk-e.

Páros CVA kísérleti design létrehozása

Ha a két ajánlatos vagy páros CVA mellett döntünk, akkor kicsit nehezebb dolgunk lesz, ugyanis az SPSS ilyen design generálására nem képes, ugyanakkor ez a CVA megközelítés, ahogy azt korábban láttuk, kérdezéstechnikai szempontból sok esetben szerencsésebb, mint az egyajánlatos módszer. A kiinduló megoldás előállításában viszont tud segíteni az SPSS, ami nagy előrelépést jelent a design generálásban.

A páros design generálásának lépései az SPSS segítségével az alábbiak:

- i. Az SPSS *Generate Orthogonal Design* modulja segítségével generáljunk minimum annyi ajánlatot, amennyire szükségünk lesz a döntési feladatokhoz. Ez ugye pontosan a kétszerese a döntési feladatok számának. Az így kapott ajánlathalmaz az ortogonalitási feltételnek megfelel (bár általában nem kellően kiegyensúlyozott).

2. A kapott ajánlathalmazból képezzünk párokat, pontosan annyi párt, ahány döntési feladatot szeretnénk kérdezni a válaszadóktól. Arra ügyeljünk, hogy a párok kialakításánál ne kerüljenek azonos szintek a két ajánlatba, illetve az ajánlatpároknál lehetőleg minden dimenziónál minden szintpárosítás előforduljon. Ha az SPSS által generált ajánlatokból már nem találunk olyat, ami ezeknek a feltételeknek megfelel, akkor manuálisan módosítsuk a kiválasztandó ajánlatot úgy, hogy megfelelő legyen az aktuális ajánlatpár.
3. Ellenőrizzük gyakorisági táblák és keresztábrák segítségével (például az SPSS-be átrakva az ajánlatokat) azt, hogy a kiegyensúlyozottság és az ortogonalitás tekintetében hogyan teljesít a designunk. Ha szükséges, akkor manuálisan módosítsuk a megfelelő ajánlatokat azon dimenziók tekintetében, ahol olyan szint fordul elő, amelyből az egyenletes megoszláshoz képest többlet figyelhető meg olyan szintre, amelyből hiány van. Természetesen az ajánlatpárok teljes különbözőségét most is tartsuk.
4. Ismételjük a 3. lépést mindaddig, amíg a design nem lesz olyan kiegyensúlyozottság és ortogonalitás tekintetében, amely már elfogadható eltéréseket mutat az egyenletes megoszlástól.

Amennyiben MS Excellel szeretnénk kísérleti designt építeni, akkor a véletlenszám generátor megfelelő lehet kiindulási pontnak. Ugyanakkor fontos kiemelni, hogy a véletlenszám generátorral létrehozott design nagyon nagy valószínűséggel nem lesz tökéletes kiegyensúlyozottság és ortogonalitás tekintetében, mivel a tesztelt ajánlatok száma túl kevés a nagy számok törvényeinek a teljesüléséhez. Ennélfogva szinte biztosra vehetjük, hogy manuálisan korrigálni kell a designunkat az előbbiekhöz hasonló lépéseken keresztül.

1. Az MS Excel *VÉLETLEN.KÖZÖTT()* függvényével készítsünk annyi ajánlatot, amennyire szükségünk a döntési feladatokhoz. Ez ugye pontosan a kétszerese a döntési feladatok számának. Érdeemes másolni és értéként beilleszteni az ajánlathalmazt egy új területre, ugyanis a *VÉLETLEN.KÖZÖTT()* függvény minden egyes cellamódosításnál új véletlenszámokat generál.
2. A kapott ajánlathalmazból képezzünk párokat, pontosan annyi párt, ahány döntési feladatot szeretnénk kérdezni a válaszadóktól. Arra ügyeljünk, hogy a párok kialakításánál ne szerepeljenek azonos szintek a két versengő ajánlatban, illetve az ajánlatpároknál lehetőleg minden dimenziónál minden szintpárosítás előforduljon. Ha már nem találunk olyan ajánlatot a véletlenszám generátor által kreált ajánlatok között,

akkor valamelyik ajánlatot manuálisan módosítsuk úgy, hogy az teljesítse az előbbi feltételeket.

3. Ellenőrizzük gyakorisági táblák és keresztábrák segítségével (például az SPSS-be átrakva az ajánlatokat) azt, hogy a kiegyensúlyozottság és az ortogonalitás tekintetében hogyan teljesít a designunk. Ha szükséges, akkor manuálisan módosítsuk a megfelelő ajánlatokat azon dimenziók tekintetében, ahol olyan szint fordul elő, amelyből az egyenletes megoszláshoz képest többlet figyelhető meg olyan szintre, amelyből hiány van. Természetesen az ajánlatpárok teljes különbözőségét most is tartjuk.
4. Ismételjük a 3. lépést mindaddig, amíg a design nem lesz olyan kiegyensúlyozottság és ortogonalitás tekintetében, amely már elfogadható eltéréseket mutat az egyenletes megoszlástól.

Összességében elmondható, hogy jó minőségű kísérleti design megalkotása megfelelő célszoftver nélkül egy komplex és időigényes feladat. A célra alkalmas szoftverek viszont költségesek, így komoly dilemma, hogy melyik irányba érdemes elindulni.

A részhasznossági értékek előállítására CVA esetén

Az egy ajánlatos CVA módszer kísérleti designjának kódolása

Ahogy azt már korábban említettük, legyen szó akár az egy ajánlatos, akár a páros CVA módszerről a részhasznossági értékeket leggyakrabban lineáris regresszió becsüljük. A becslést válaszadói szinten végezzük, ami annyit tesz, hogy annyi modellt építünk, ahány válaszadónk van. A modellben a függő változó az egyes ajánlatokra adott értékelés lesz, míg a független változók pedig azok a dummy változók, amelyek azt írják le, hogy az egyes kártyákon a dimenziók mely szintje szerepel.

Az SPSS *Generate Orthogonal Design* moduljával a 3.4 fejezetben előállítottuk a webshopos esettanulmányunkhoz a kísérleti designt, amely 4 dimenzióval bír és minden dimenzióban 4 szint közül mindig az egyik jelent meg.

Az SPSS által generált design az alábbiak szerint alakul a 12.7. táblázat szerint alakul. A *Feladat* változó a döntési feladatok beazonosításához szükséges futóindex, az azt követő 4 változó (*Dim. 1 – Dim. 4*) mutatja, hogy az aktuális kártyán a dimenzió mely szintje jelenik meg, míg az utolsó 12 darab változó a *Dim. 1 – Dim. 4* változókból képzett dummy változók.

12.7. táblázat: Az SPSS által generált design

Feladat	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim_1_1	Dim_1_2	Dim_1_3	Dim_2_1	Dim_2_2	Dim_2_3	Dim_3_1	Dim_3_2	Dim_3_3	Dim_4_1	Dim_4_2	Dim_4_3
1	1	3	4	3	1	0	0	0	0	1	0	0	0	0	0	1
2	4	4	4	1	0	0	0	0	0	0	0	0	0	1	0	0
3	4	1	2	1	0	0	0	1	0	0	0	1	0	1	0	0
4	4	1	1	4	0	0	0	1	0	0	1	0	0	0	0	0
5	1	2	1	4	1	0	0	0	1	0	1	0	0	0	0	0
6	3	1	4	4	0	0	1	1	0	0	0	0	0	0	0	0
7	2	1	1	3	0	1	0	1	0	0	1	0	0	0	0	1
8	1	2	4	1	1	0	0	0	1	0	0	0	0	1	0	0
9	1	3	2	4	1	0	0	0	0	1	0	1	0	0	0	0
10	2	3	1	1	0	1	0	0	0	1	1	0	0	1	0	0
11	4	3	1	2	0	0	0	0	0	1	1	0	0	0	1	0
12	2	1	4	2	0	1	0	1	0	0	0	0	0	0	1	0
13	1	1	3	2	1	0	0	1	0	0	0	0	1	0	1	0
14	3	2	1	2	0	0	1	0	1	0	1	0	0	0	1	0
15	1	4	1	3	1	0	0	0	0	0	1	0	0	0	0	1
16	4	2	3	3	0	0	0	0	1	0	0	0	1	0	0	1
17	1	1	3	1	1	0	0	1	0	0	0	0	1	1	0	0
18	1	4	2	2	1	0	0	0	0	0	0	1	0	0	1	0
19	3	1	2	3	0	0	1	1	0	0	0	1	0	0	0	1
20	1	1	1	1	1	0	0	1	0	0	1	0	0	1	0	0
21	3	4	1	1	0	0	1	0	0	0	1	0	0	1	0	0
22	2	4	3	4	0	1	0	0	0	0	0	0	1	0	0	0
23	2	2	2	1	0	1	0	0	1	0	0	1	0	1	0	0
24	1	1	1	1	1	0	0	1	0	0	1	0	0	1	0	0
25	3	3	3	1	0	0	1	0	0	1	0	0	1	1	0	0

A dummy változók képzésének elve rendkívül egyszerű: minden dimenziót annyi dummy változóba fogunk rögzíteni, ahány szintje van mínusz 1. Tehát a webshopos esettanulmány esetében minden dimenziót 3 dummy változóval veszünk fel, mivel minden dimenziónak 4 szintje van. Az első dimenziót a *Dim_1_1*, *Dim_1_2* és *Dim_1_3* dummy változókkal írjuk le. *Dim_1_1* akkor vesz fel 1-es értéket, ha az aktuális kártyán az első dimenzióban az 1-es szint jelenik meg, minden más esetben 0-s értéket vesz fel. A *Dim_1_2* dummy változó akkor vesz fel 1-es értéket, ha az aktuális kártyán a 2-es szint szerepel, minden más esetben 0 értéket fog kapni. Ugyanezen logika alapján kódoljuk a *Dim_1_3* dummy változót az első dimenzió 3-as szintje alapján. Az első dimenzió 4-es szintje úgy jelenik meg a 3 dummy változónkban, hogy a *Dim_1_1*, *Dim_1_2* és *Dim_1_3* változók mindegyike 0-át vesz fel. A kódolási logikát az első dimenzióra vonatkozólag a 12.8. táblázat foglalja össze.

12.8. táblázat: kódolási logika az első dimenzió esetében

1. dimenzióban megjelenített szint	Dim_1_1	Dim_1_2	Dim_1_3
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

Az utolsó szint azért nem rendelkezik saját dummy változóval, mert az első dimenzió 3 dummy változója egyértelműen leírja, hogy mikor szerepel az aktuális ajánlatban ez a szint (ugye akkor lesz mind a három dummy változó értéke 0). Így az eredményekben ennek a szintnek a hasznossági értékét is látni fogjuk a látszólag elhagyott szintet.

A páros CVA módszer kísérleti designjának kódolása

A páros CVA esetén is dummy kódolást végzünk azzal a különbséggel az egy-ajánlatos CVA-vel összehasonlítva, hogy ha az adott szint a bal ajánlatban jelenik meg, akkor -1-et vesz fel a megfelelő dummy változó, ha jobb oldali ajánlatban szerepel az adott szint, akkor pedig +1-et.

A webshopos esettanulmányban a páros CVA megközelítésnél 18 döntési feladat volt, amelyek dummy kódolása a 12.9. táblázat szerint alakult.

12.9. táblázat: A 18 döntési feladat dummy kódolása

Feladat	Bal oldali ajánlat				Jobb oldali ajánlat				Dim_1_1	Dim_1_2	Dim_1_3	Dim_2_1	Dim_2_2	Dim_2_3	Dim_3_1	Dim_3_2	Dim_3_3	Dim_4_1	Dim_4_2	Dim_4_3
	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 1	Dim. 2	Dim. 3	Dim. 4												
1	3	2	2	4	1	4	1	1	1	0	-1	0	-1	0	1	-1	0	1	0	0
2	4	4	4	3	3	3	3	2	0	0	1	0	0	1	0	0	1	0	1	-1
3	4	2	3	4	3	1	2	1	0	0	1	1	-1	0	0	1	-1	1	0	0
4	3	3	4	2	2	2	1	1	0	1	-1	0	1	-1	1	0	0	1	-1	0
5	1	2	4	2	4	3	2	4	-1	0	0	0	-1	1	0	1	0	0	-1	0
6	1	3	2	1	4	4	3	2	-1	0	0	0	0	-1	0	-1	1	-1	1	0
7	2	3	1	3	1	4	2	2	1	-1	0	0	0	-1	-1	1	0	0	1	-1
8	2	3	2	3	3	1	1	4	0	-1	1	1	0	-1	1	-1	0	0	0	-1
9	3	2	1	2	1	1	3	3	1	0	-1	1	-1	0	-1	0	1	0	-1	1
10	4	1	2	2	3	4	4	3	0	0	1	-1	0	0	0	-1	0	0	-1	1

Feladat	Bal oldali ajánlat				Jobb oldali ajánlat				Dim_1_1	Dim_1_2	Dim_1_3	Dim_2_1	Dim_2_2	Dim_2_3	Dim_3_1	Dim_3_2	Dim_3_3	Dim_4_1	Dim_4_2	Dim_4_3
	Dim. 1	Dim. 2	Dim. 3	Dim. 4	Dim. 1	Dim. 2	Dim. 3	Dim. 4												
11	2	4	1	4	4	2	4	3	0	-1	0	0	1	0	-1	0	0	0	0	1
12	3	2	3	1	2	1	4	2	0	1	-1	1	-1	0	0	0	-1	-1	1	0
13	3	4	2	3	1	3	4	4	1	0	-1	0	0	1	0	-1	0	0	0	-1
14	4	1	3	1	2	2	2	4	0	1	0	-1	1	0	0	1	-1	-1	0	0
15	4	3	4	1	2	1	3	3	0	1	0	1	0	-1	0	0	1	-1	0	1
16	2	1	4	1	1	3	1	3	1	-1	0	-1	0	1	1	0	0	-1	0	1
17	1	4	3	4	4	3	1	2	-1	0	0	0	0	1	1	0	-1	0	1	0
18	1	1	1	3	2	4	3	1	-1	1	0	-1	0	0	-1	0	1	1	0	-1

A páros CVA esetén nagyon fontos, hogy annak ellenére, hogy döntési feladatonként 2 ajánlatunk van, nem duplázódik sem a szintek kódolásához használt változók száma, sem a döntési feladatok száma.

Az előbb bemutatott dummy kódolású mátrixot érdemes átrakni SPSS-be, mivel a későbbiekben itt fogjuk felhasználni a benne szereplő dummy változókat a regressziós elemzések során független változóként. A változók elnevezését érdemes az 12.7. ábrának megfelelően megtenni.

12.7. ábra: A változók elnevezése

	TASK	Dim_1_1	Dim_1_2	Dim_1_3	Dim_2_1	Dim_2_2	Dim_2_3	Dim_3_1	Dim_3_2	Dim_3_3	Dim_4_1	Dim_4_2	Dim_4_3
1	1	1	0	-1	0	-1	0	1	-1	0	1	0	0
2	2	0	0	1	0	0	1	0	0	1	0	1	-1
3	3	0	0	1	1	-1	0	0	1	-1	1	0	0
4	4	0	1	-1	0	1	-1	1	0	0	1	-1	0
5	5	-1	0	0	0	-1	1	0	1	0	0	-1	0
6	6	-1	0	0	0	0	-1	0	-1	1	-1	1	0
7	7	1	-1	0	0	-1	-1	1	0	0	1	-1	0
8	8	0	-1	1	1	0	-1	1	-1	0	0	0	-1
9	9	1	0	-1	1	-1	0	-1	0	1	0	-1	1
10	10	0	0	1	-1	0	0	0	-1	0	0	-1	1
11	11	0	-1	0	0	1	0	-1	0	0	0	0	1
12	12	0	1	-1	1	-1	0	0	0	-1	-1	1	0
13	13	1	0	-1	0	0	1	0	-1	0	0	0	-1
14	14	0	1	0	-1	1	0	0	1	-1	-1	0	0
15	15	0	1	0	1	0	-1	0	0	1	-1	0	1
16	16	1	-1	0	-1	0	1	1	0	0	-1	0	1
17	17	-1	0	0	0	0	1	1	0	-1	0	1	0
18	18	-1	1	0	-1	0	0	-1	0	1	1	0	-1

Az elemzéshez szükséges adatbázis előállítás

Az előbbieken áttekintettük, hogy a regressziós modellünkhöz a *független változókat* hogyan definiáljuk és hogyan kódoljuk. Viszont még hiányzik a *függő vál-*

tozó, amely a conjoint esetében a válaszadói preferencia vagy annak mértéke a döntési feladatban szereplő ajánlatokra vonatkozólag.

A conjoint elemzéshez használt adatbázis felépítése kissé eltér a szokásostól, ami annyit jelent, hogy egy tetszőleges válaszadót nem egyetlen sor fog reprezentálni az adatbázisban, hanem annyi, ahány döntési feladatot kérdeztünk az interjú során. És ez egyaránt igaz az egyajánlatos és a páros CVA modellekre, de bármely más conjoint módszer esetén is hasonló jellegű adatokból fogunk dolgozni.

Az adatbázis felépítése a webshopos esettanulmány esetében a 12.10. táblázatban látható. Az *ID* változó a válaszadók azonosításához szükséges úgynevezett *futóindex*, míg a *TASK* mutatja meg, hogy mely döntési feladatra vonatkozólag tartalmaz adatokat az adott rekord vagy sor az adatbázisunkban. A *Preferencia* oszlop azt írja le, hogy a válaszadónk milyen választ adott az adott döntési feladatban, míg a *Dim* változóink azt írják le, hogy az egyes döntési feladatokban milyen szintek szerepeltek az egyes ajánlatokban. Ennek megfelelően a 12.10. táblázat 1. sora az alábbiakat tartalmazza:

- A 8-as azonosítójú válaszadót (*ID*)
- Neki is az 1. döntési feladatra (*Task*) vonatkozó értékelését és a feladat paramétereit
- Ebben a döntési feladatban a válaszadónk 6-os értékelést adott a 7 fokú skálán (*Preferencia*), azaz a jobboldali ajánlatot preferálta a baloldallal szemben
- A *Dim* változók írják le, hogy pontosan mi is szerepelt az 1. döntési feladat két ajánlatában

A 8-as azonosítójú válaszadóra még 17 sort látunk, melyek az előbbi logika szerint tartalmaznak információt a további döntési feladatokra. Majd következik a 9-es azonosítójú válaszadó a 18 sorával, hiszen ő is 18 döntési feladatot értékelt. Az ő 18 sorának a logikája teljesen azonos a fentiekben leírtakkal. Mivel az esettanulmány során 390 válaszadót kérdeztünk meg, így a conjoint adatbázisunk összesen 18×390, azaz 7020 sort vagy rekordot fog tartalmazni.

12.10. táblázat: Az adatbázis felépítése

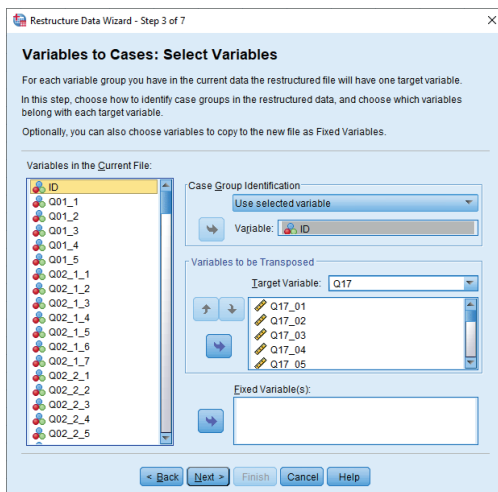
Task	ID	Preferencia	Dim_1_1	Dim_1_2	Dim_1_3	Dim_2_1	Dim_2_2	Dim_2_3	Dim_3_1	Dim_3_2	Dim_3_3	Dim_4_1	Dim_4_2	Dim_4_3
1	8	6	1	0	-1	0	-1	0	1	-1	0	1	0	0
2	8	6	0	0	1	0	0	1	0	0	1	0	1	-1
3	8	6	0	0	1	1	-1	0	0	1	-1	1	0	0

Task	ID	Preferencia	Dim_1_1	Dim_1_2	Dim_1_3	Dim_2_1	Dim_2_2	Dim_2_3	Dim_3_1	Dim_3_2	Dim_3_3	Dim_4_1	Dim_4_2	Dim_4_3
4	8	2	0	1	-1	0	1	-1	1	0	0	1	-1	0
5	8	1	-1	0	0	0	-1	1	0	1	0	0	-1	0
6	8	2	-1	0	0	0	0	-1	0	-1	1	-1	1	0
7	8	6	1	-1	0	0	0	-1	-1	1	0	0	1	-1
8	8	7	0	-1	1	1	0	-1	1	-1	0	0	0	-1
9	8	3	1	0	-1	1	-1	0	-1	0	1	0	-1	1
10	8	1	0	0	1	-1	0	0	0	-1	0	0	-1	1
11	8	6	0	-1	0	0	1	0	-1	0	0	0	0	1
12	8	4	0	1	-1	1	-1	0	0	0	-1	-1	1	0
13	8	5	1	0	-1	0	0	1	0	-1	0	0	0	-1
14	8	4	0	1	0	-1	1	0	0	1	-1	-1	0	0
15	8	2	0	1	0	1	0	-1	0	0	1	-1	0	1
16	8	7	1	-1	0	-1	0	1	1	0	0	-1	0	1
17	8	7	-1	0	0	0	0	1	1	0	-1	0	1	0
18	8	1	-1	1	0	-1	0	0	-1	0	1	1	0	-1
1	9	6	1	0	-1	0	-1	0	1	-1	0	1	0	0
2	9	6	0	0	1	0	0	1	0	0	1	0	1	-1
3	9	6	0	0	1	1	-1	0	0	1	-1	1	0	0
4	9	6	0	1	-1	0	1	-1	1	0	0	1	-1	0
5	9	6	-1	0	0	0	-1	1	0	1	0	0	-1	0
6	9	2	-1	0	0	0	0	-1	0	-1	1	-1	1	0
7	9	6	1	-1	0	0	0	-1	-1	1	0	0	1	-1
8	9	6	0	-1	1	1	0	-1	1	-1	0	0	0	-1
...

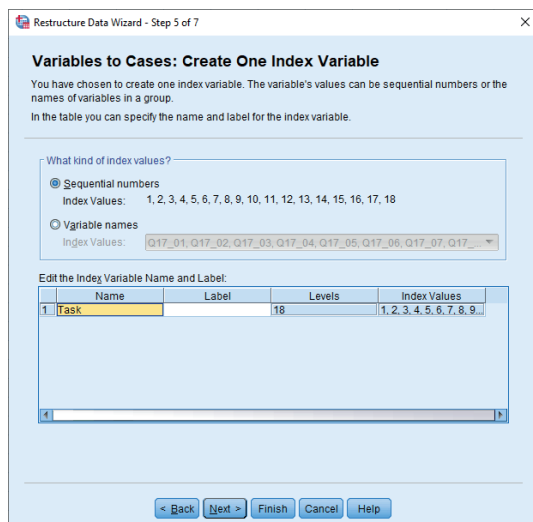
Az itt bemutatott, a szokásostól kissé eltérő logikájú adatbázisunkat tipikusan vagy MS Excelben, vagy pedig SPSS segítségével állítjuk elő. Amennyiben az MS Excel választjuk, akkor a kész adatbázist majd importálni kell SPSS-be, hogy a lineáris regressziót le tudjuk futtatni az adatbázisunkon.

Ha az adatbázis előállításához az SPSS-t szeretnénk használni, akkor egy kevésbé gyakran használt menüponttal kell dolgoznunk. Először töltsük be az esettanulmány „eredeti” adatbázisát, ahol 18 változóban van eltárolva a conjoint blokk 18 döntési feladatára vonatkozó válasz (Q17_01-től Q17_18-ig), hiszen minden válaszadónk egy sort vagy rekordot reprezentál az adatbázisunkban. Az SPSS főmenüjében a *Data* menüpontra kattintva válasszuk ki a *Restructure* almenüpontot.

1. Az SPSS egy új ablakot fog nyitni, ahol kiválaszthatjuk, hogy milyen jellegű átalakítást szeretnénk végrehajtani az adatokon. Itt maradunk az alapbeállításnál (*Restructure selected variables into cases*), majd kattintsunk a *Next >* gombra.
2. Ezt követően azt kell meghatároznunk, hogy hány változócsoportot szeretnénk átstrukturálni (*Number of Variable Groups* ablak), ahol ugyancsak az alapbeállítás, vagyis a *One* opció a megfelelő választás számunkra.
3. A következő lépésben kell definiálnunk, hogy pontosan mely változókat érinti az átstrukturálás.
 - a. A *Case Group Identification* blokkban kell meghatároznunk, hogy a válaszadók azonosítására használt változó vagy futóindex melyik változó. Itt a lenyíló menüből válasszuk ki a *Use selected variable* opciót, majd a *Variable* beviteli mezőbe tegyük át a baloldali változó listából az *ID* változót (amely az adatbázisunkban a válaszadók azonosítására szolgáló információt tartalmazza).
 - b. A *Variables to be Transposed* blokkban a *Target Variable* beviteli mezőbe írjuk be a létrehozandó változó nevét, ami legyen például a *Q17*, hiszen az új változót a *Q17_01 – Q17_18* változókból fogjuk képezni. Majd tegyük át a baloldali változó listából a kis nyílra kattintva a *Q17_01, Q17_02, Q17_03, Q17_04, Q17_05, Q17_06, Q17_07, Q17_08, Q17_09, Q17_10, Q17_11, Q17_12, Q17_13, Q17_14, Q17_15, Q17_16, Q17_17, Q17_18* változókat (amelyekből egyetlen változót szeretnénk készíteni) az előbbi beviteli mezőbe.
 - c. Majd kattintsunk a *Next >* gombra



4. A *Create Index Variables* lépésben az alapbeállítás megfelelő számunkra (*One*), így kattintsunk a *Next > gombra*.
5. A *Create One Index* lépésben állíthatjuk be a korábbiakban leírt, a döntési feladat beazonosítására szolgáló *Feladat* vagy *Task* változót, mint a döntési feladatok futóindexét. A *What kind of index values?* blokkban maradhat az alapbeállítás, azaz a *Sequential numbers*. Az *Edit the Index Variable Name and Label* blokkban az *Index1* változó nevet írjuk át a beszédesebb *Task* névre, majd kattintsunk a *Next > gombra*.



6. Az *Options* ablakban a *Handling of Variables not Selected* blokkban választjuk a *Drop variable(s) from the new data file* opciót. Ezzel a lépéssel az átstrukturálás definiálásának végére is értünk. Ha nem célunk szkript készítése az előbbi lépésekből, akkor kattintsunk a *Finish* gombra.

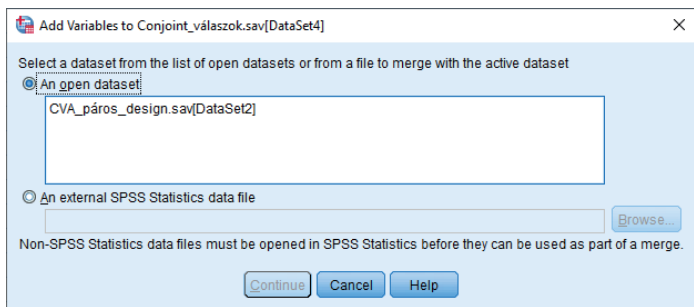
Ezzel egy olyan rendkívül kevés változót (*ID*, *Task*, *Q17*) tartalmazó adatbázishoz jutottunk, amelyben viszont relatíve sok, azaz 7020 rekord van, ami ugye a 390×18 számok szorzatából adódik. Ez az adatbázis tartalmazza a regressziós modellünk *függő* változóját.

	ID	Task	Q17
1	10	1	5
2	10	2	5
3	10	3	2
4	10	4	5
5	10	5	2
6	10	6	2
7	10	7	6
8	10	8	5
9	10	9	4
10	10	10	2
11	10	11	5
12	10	12	6
13	10	13	6
14	10	14	2
15	10	15	2
16	10	16	3
17	10	17	6
18	10	18	3
19	100	1	3
20	100	2	1
21	100	3	1

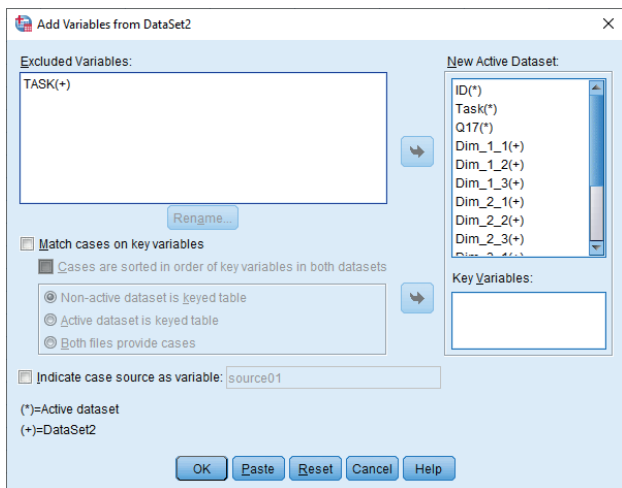
Kísérleti design adatbázis és a válaszokat tartalmazó adatbázis összefűzése

Ha modellünkhöz szükséges adatbázis előállítását az SPSS segítségével állítjuk elő, akkor ennél a pontnál két külön adatbázisban található a regressziós modellhez szükséges *függő változó (Q17)* és a *független változók (DIM)*. Ebben a lépésben nem teszünk mást, mint összefűzzük a két adatbázist.

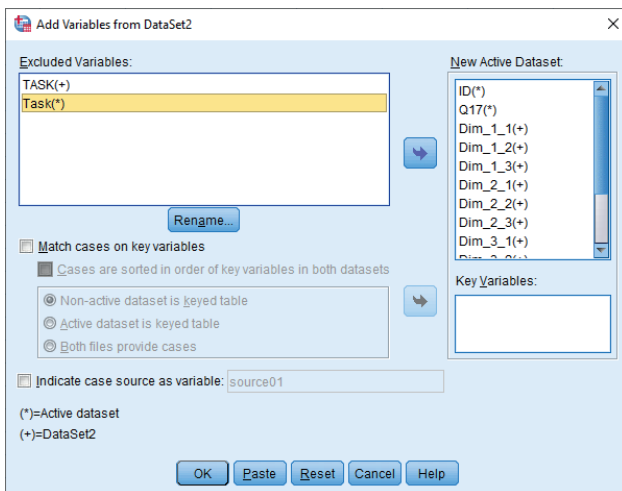
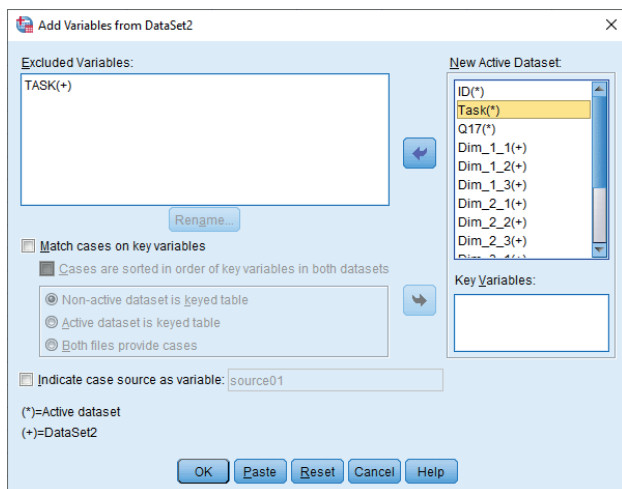
Mindenekelőtt bizonyosodjunk meg róla, hogy az aktív adatbázisunk a *Q17* változót tartalmazó adatbázis legyen, vagyis az SPSS-nek abban az ablakban tegyük meg a következő lépéseket, ahol ezt az adatbázist látjuk. Az adatbázisok összefűzését a *Data* főmenü *Merge Files* pontján találjuk, ahol kétféle lehetőség kínálkozik, amelyből nekünk az *Add variables* opciót kell választanunk. Ennek segítségével lehet egy meglévő adatbázishoz egy másik adatbázisból új változókat hozzáfűzni, amire nekünk pont szükségünk van, hiszen azt pontosan tudjuk, hogy az egyes döntési feladatokban mi volt a válasza a válaszadónknak, de azt nem tudjuk még ebben az adatbázisban, hogy mi szerepelt az egyes kártyákon. Miután az *Add variables* opcióra kattintottunk, egy új ablakban választhatjuk ki azt az adatbázist, ami a kísérleti designt tartalmazza. Amennyiben a kísérleti designt tartalmazó adatbázist már korábban megnyitottuk, akkor az *An open dataset* listában fog szerepelni, ha nincs nyitva, akkor az *An external SPSS Statistics data file* blokkon segítségével tudjuk megkeresni a fájljaink között. Esetünkben egy nyitott adatbázis, a *CVA_páros_design.sav* tartalmazza a kísérleti design, amelyet tehát szeretnénk hozzáfűzni a válaszokat tartalmazó adatbázishoz.



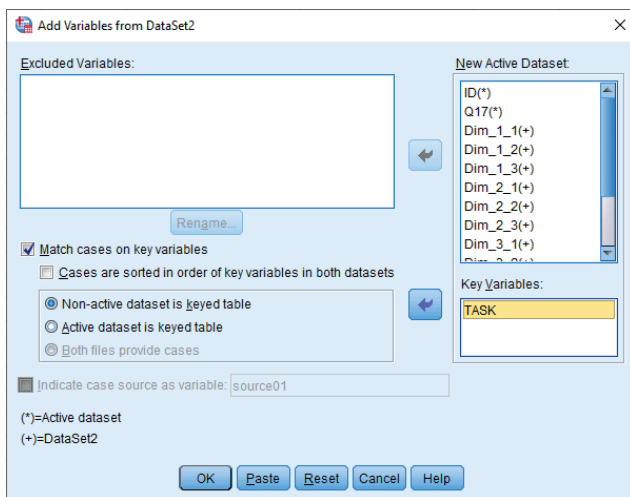
Ha kiválasztottuk a megfelelő adatbázist, akkor az alább látható ablak jelenik meg, ahol definiálhatjuk az adatbázisok összefűzésének logikáját. A *New Active Dataset* lista tartalmazza az új adatbázis összes változóját. Itt szerepelnek a válaszokat tartalmazó (*Q17*) adatbázisnak a változói, illetve a kísérleti design adatbázis azon változói, amelyek a változók elnevezése alapján nem okoznak duplikációt.



Az SPSS a kísérleti design adatbázisból a *TASK* változót szeretné kihagyni az összefűzésből, ami azért van, mert a *TASK* változó – szerencsére – mindkét adatbázisban szerepel, így redundáns. Ugyanakkor nagyon fontos, hogy az összefűzés „kulcsa” pont ez a változó, így definiálnunk kellene, hogy ennek segítségével rendezze össze az SPSS a válaszok és a kísérleti design adatbázisát annak, érdekében, hogy a válaszok a kísérleti design megfelelő sorára hivatkozzanak. Ezt úgy tudjuk megtenni, hogy kattintsunk a *New Active Dataset* listában a *Task* változóra és tegyük át a nyíl gomb segítségével az *Excluded Variables* blokkba.

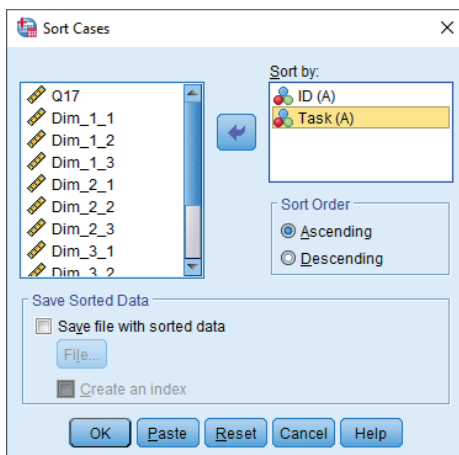


Ezt követően jelöljük ki a **CTRL** billentyű nyomva tartása mellett kattintással az *Excluded Variable* listában mind a *Task* és a *TASK* változót, illetve jelöljük be a *Match cases on key variables* opciót, majd kattintsunk a *Key variables* melletti gombra, amely lépéssel definiáljuk, hogy a két adatbázis összefűzésének alapja a *Task* változó lesz.



Végül kattintsunk az **OK** gombra.

Annak érdekében, hogy megfelelő struktúrában lássuk az adatbázist, a *Data* főmenüpontra kattintsunk a *Sort* pontra, majd válasszuk ki az *ID* és a *Task* változókat és a kis nyíl gombbal tegyük át *Sort by* listába, amivel definiáljuk, hogy az adatbázis rendezésének alapja ez a két változó legyen a megadott sorrendben.

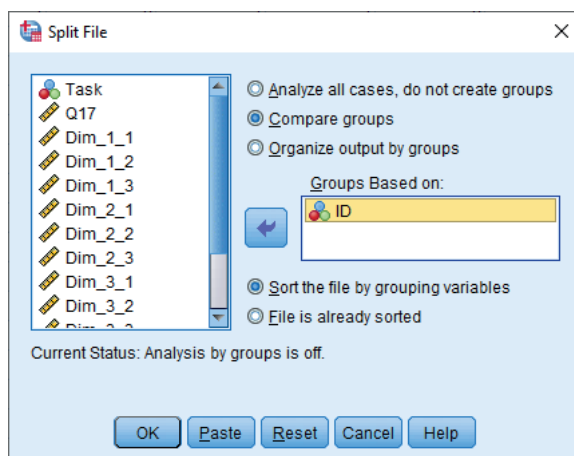


Az OK gombra kattintva az SPSS átrendezi az adatbázist, és az adat nézetben az alábbi adatbázist kell látnunk.

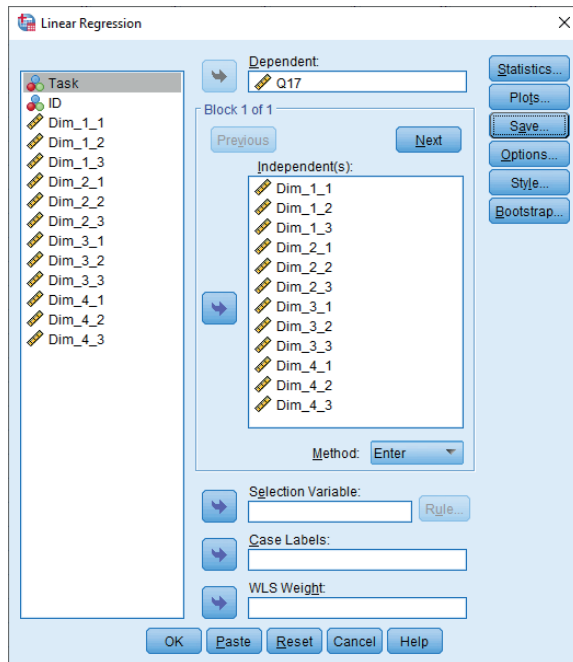
	Task	ID	Q17	Dim_1_1	Dim_1_2	Dim_1_3	Dim_2_1	Dim_2_2	Dim_2_3	Dim_3_1	Dim_3_2	Dim_3_3	Dim_4_1	Dim_4_2	Dim_4_3
1	1	8	6	1	0	-1	0	-1	0	1	-1	0	1	0	0
2	2	8	6	0	0	1	0	0	1	0	0	1	0	1	-1
3	3	8	6	0	0	1	1	-1	0	0	1	-1	1	0	0
4	4	8	2	0	1	-1	0	1	-1	1	0	0	1	-1	0
5	5	8	1	-1	0	0	0	-1	1	0	1	0	0	-1	0
6	6	8	2	-1	0	0	0	0	-1	0	-1	1	-1	1	0
7	7	8	6	1	-1	0	0	0	-1	-1	1	0	0	1	-1
8	8	8	7	0	-1	1	1	0	-1	1	-1	0	0	0	-1
9	9	8	3	1	0	-1	1	-1	0	-1	0	1	0	-1	1
10	10	8	1	0	0	1	-1	0	0	0	-1	0	0	-1	1
11	11	8	6	0	-1	0	0	1	0	-1	0	0	0	0	1
12	12	8	4	0	1	-1	1	-1	0	0	0	-1	-1	1	0
13	13	8	5	1	0	-1	0	0	1	0	-1	0	0	0	-1
14	14	8	4	0	1	0	-1	1	0	0	1	-1	-1	0	0
15	15	8	2	0	1	0	1	0	-1	0	0	1	-1	0	1
16	16	8	7	1	-1	0	-1	0	1	1	0	0	-1	0	1
17	17	8	7	-1	0	0	0	0	1	1	0	-1	0	1	0
18	18	8	1	-1	1	0	-1	0	0	-1	0	1	1	0	-1
19	1	9	6	1	0	-1	0	-1	0	1	-1	0	1	0	0
20	2	9	6	0	0	1	0	0	1	0	0	1	0	1	-1
21	3	9	6	0	0	1	-1	-1	0	0	-1	-1	1	0	0
22	4	9	6	0	1	-1	0	1	-1	1	0	0	1	-1	0
23	5	9	6	-1	0	0	0	-1	1	0	1	0	0	-1	0
24	6	9	2	-1	0	0	0	0	-1	0	-1	1	-1	1	0
25	7	9	6	1	-1	0	0	0	-1	-1	1	0	0	1	0

A részhasznossági értékek becslése

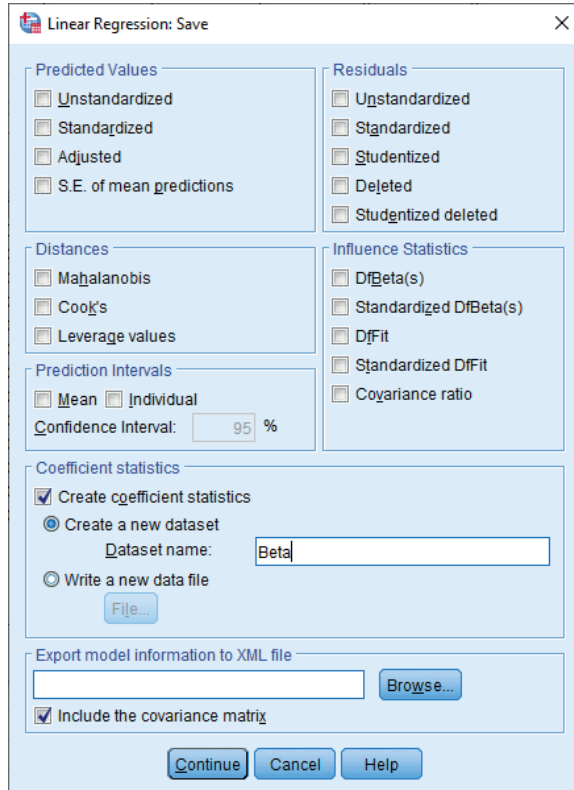
Ha előállt a *függő* és *független* változókat is tartalmazó adatbázisunk, akkor úgy le kell futtatnunk a lineáris regresszió modelljét, hogy az SPSS azt minden válszadóra külön-külön tegye. Szerencsére az SPSS-ben van lehetőségünk arra, hogy egyidejűleg hajtsunk végre egy adatbázis összes alcsoportján egy bizonyos elemzést. Ezt úgy tudjuk beállítani, hogy a *Data* főmenü ponton belül a *Split file* pontra kattintunk, majd a felugró ablakon belül jobb oldalon kiválasztjuk a *Compare groups* opciót, majd az *ID* változót átrakjuk a kis nyíl segítségével a *Groups Based on* listába.



Ezt követően az SPSS-t arra kell utasítanunk, hogy futasson lineáris regressziót a megfelelő változók alapján minden válaszadóra külön-külön. Kattintsunk az *Analyze* fő menüpontra, majd a *Regression* ponton belül válasszuk ki a *Linear* menüpontot. A *Dependent* változók a *Q17* lesz, míg az *Independent(s)* változóink az ajánlatokon szereplő szintekből képzett dummy változók, azaz a Dim előtaggal rendelkező változók.



A függő és független változók definiálása után még be kell állítanunk azt, hogy az SPSS adatbázis formájában mentse el a regressziós együtthatókat, a bétákat. Ezt úgy tehetjük meg, hogy kattintsunk a *Save* gombra, majd azt követően jelöljük be a *Create coefficient statistics* opciót, végül adjuk meg a létrehozandó adatbázis nevét (például *Beta*). A regressziós béták elmentésére más lehetőség is van (például az SPSS output MS Excelbe történő átkopizása), de az előbb felvázolt megközelítés mindenképpen a legegészségesebb eljárás. Sőt annak az az előnye is megvan, hogy a béták adatbázisával a későbbiekben további transzformációk az SPSS-szel egyből végrehajthatók mindenféle adatimport nélkül, amennyiben ezeket a transzformációkat SPSS-szel szeretnénk végrehajtani. Ezt követően a *Linear Regression* ablakon kattintsunk a *Continue* gombra.



Ezzel visszalépünk a *Linear regression* modul fő ablakára, ahol más alapbeállítás módosítására nincs szükség. A lineáris regresszió alapbeállításai ugyanis tökéletesen megfelelnek nekünk, azaz az *Enter* módszerrel „kényszerítsünk” minden független változót a modellbe, illetve tengelymetszetre is kérjük becslést (*Options* ablak *Include constant in equation* opció legyen bekattintva). Nem marad más hátra, mint kattintsunk a fő ablak OK gombjára, és az SPSS lefuttatja esetünkben a 390 darab lineáris regressziót – hiszen ennyi válaszadónk van – az adatainkon.

A conjoint elemzésnél a regressziós output egyik legfontosabb része a modellek illeszkedését leíró *Model Summary* tábla. Ebben kerül kiírásra az a korrelációs koefficiens (R), illetve determinációs együttható (R Square), amelyek azt írják le, hogy mennyire mozognak együtt a valós válaszok a modellek által megbecsült értékelésekkel. Mivel válaszadói szinten a becslendő paraméterszámhoz képest arányaiban kevés a megfigyelések száma (emlékezzünk vissza, hogy ez hányados sokszor 1,5 értéket jelent), így nagyon magas korrelációs együtthatókat várunk el. Egy kis adathalmazra illesztünk lineáris regressziót (arányaiban

sok paraméterrel), így abban az esetben, ha a válaszadók kellő figyelemmel töltötte ki a kérdőívet, a modelleknek jól kell illeszkedniük. Ha megnézzük, akkor szembejövő, hogy az *R Square* értékek többsége 0,8 felett van, sőt nem ritka a 0,9 feletti együttható, ami már-már szinte függvényyszerű kapcsolatról tanúskodik. 0,7 alatti detarminációs együtthatóval elvétve találkozunk, azaz összességében levonhatjuk azt a következtetést, hogy a mintánk olyan szempontból jó minőségű, hogy a válaszadók többsége valamilyen mintát vagy logikát követve – azaz odafigyelve – válaszolta meg a kérdőív conjoint blokkjának döntési feladatait. A gyakorlatban általában nem zárjuk ki azokat a válaszadókat a conjoint elemzésekben, akiknél a determinációs együttható alacsony, mivel egy ésszerű arány esetén (például 20% alatt) feltételezhető, hogy vannak olyan vásárlók vagy ügyfelek, akik nem igazán involváltak egy vásárlási döntésben.

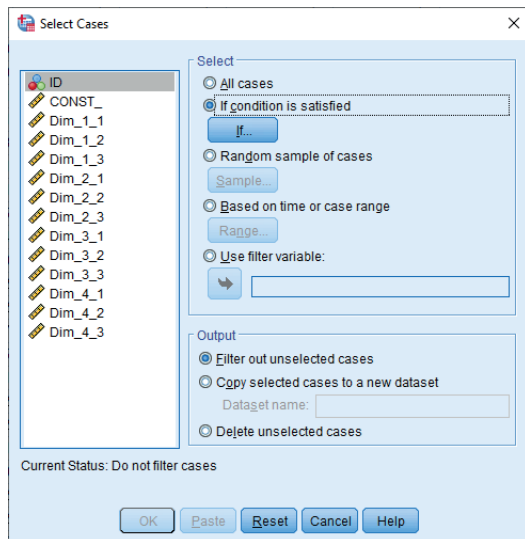
Model Summary					
Sorszám	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
8	1	,967 ^a	,935	,778	1,067
9	1	,924 ^a	,855	,506	1,343
10	1	,900 ^a	,811	,356	1,334
12	1	,979 ^a	,959	,860	,772
13	1	,953 ^a	,909	,690	,978
14	1	,906 ^a	,821	,391	1,999
15	1	,937 ^a	,877	,582	1,454
16	1	,926 ^a	,857	,512	1,821
17	1	,950 ^a	,903	,671	1,394
18	1	,913 ^a	,833	,431	1,869
19	1	,969 ^a	,940	,795	1,364
20	1	,862 ^a	,743	,125	1,168
21	1	,992 ^a	,985	,948	,463
22	1	,983 ^a	,966	,883	,747
23	1	,961 ^a	,923	,739	1,196
24	1	,915 ^a	,838	,450	1,619
25	1	,936 ^a	,875	,576	,996
26	1	,929 ^a	,862	,532	1,082
27	1	,912 ^a	,832	,428	1,444
28	1	,940 ^a	,883	,602	1,337
29	1	,921 ^a	,848	,484	1,207
30	1	,946 ^a	,895	,643	1,604
31	1	,969 ^a	,940	,795	,965

Fontos kiemlni, hogy conjoint elemzés esetén azzal nem szokás foglalkozni, hogy az ANOVA próbák alapján szignifikánsak-e a válaszadói modellek. Ennek oka megint csak az, hogy az alacsony megfigyelésszámok miatt félrevezető következtetést vonhatnánk le az eredményekre nézve. Ugyancsak nem szoktuk vizsgálni a regressziós bétáknál a T-próba eredményét sem, tehát azon változókat is benn tartjuk az egyes modellekben, amelyek ezen próba alapján nem tekinthetők szignifikánsnak.

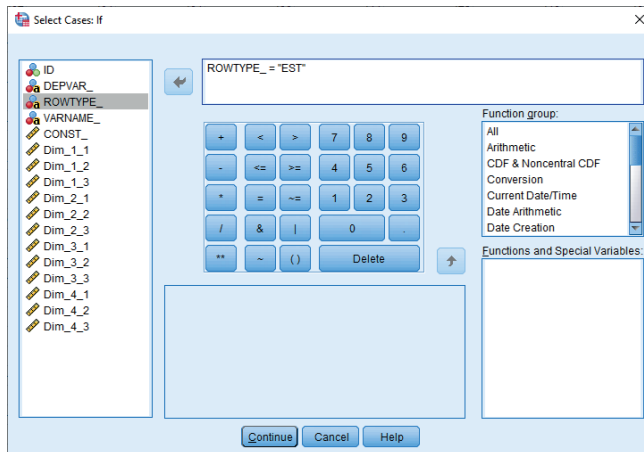
Ha a lineáris regresszió modul *Create coefficient statistics* opcióját beaktítottuk a *Save* ablakban, akkor az SPSS létrehozott nekünk egy új adatbázist, amely a regressziós együtthatókat, illetve az azokra vonatkozó statisztikákat tartalmazza.

	ID	DEPVAR	ROWTYPE	VARNAME	CONST	Dim_1_1	Dim_1_2	Dim_1_3	Dim_2_1	Dim_2_2	Dim_2_3	Dim_3_1	Dim_3_2	Dim_3_3	Dim_4_1	Dim_4_2	Dim_4_3	
1				CONST_1	-.07	-.01	-.01	.00	.00	.02	.00	-.01	.00	.01	.01	-.01	.00	
2				Dim_1_1	-.01	.20	.10	.09	.00	.01	.01	.00	.00	.01	.01	.03	.01	
3				Dim_1_2	-.01	.10	.21	.11	.01	.01	.03	-.01	-.03	.00	.02	.04	.02	
4				Dim_1_3	.00	.09	.11	.20	.02	.00	.02	-.02	-.04	-.01	.01	.01	.01	
5				Dim_2_1	.00	.00	.01	.02	.23	.14	.10	-.01	-.02	.01	.00	-.03	-.01	
6				Dim_2_2	.02	.01	.01	.00	.14	.25	.10	-.01	-.01	.01	.03	.00	.01	
7				Dim_2_3	.00	.01	.03	.02	.10	.10	.18	-.02	-.02	.03	.01	.00	.00	
8				Dim_3_1	-.01	.00	-.01	-.02	-.01	-.01	-.02	.20	.10	.10	.01	.02	.01	
9				Dim_3_2	.00	.00	-.03	.04	-.02	.01	-.02	.10	.10	.21	.10	.01	.02	.01
10				Dim_3_3	-.01	.01	.00	.01	.01	.01	.03	.10	.10	.21	.02	.04	.02	
11				Dim_4_1	.01	.01	.02	.01	.00	.03	.01	.01	.01	.01	.02	.21	.12	.11
12				Dim_4_2	-.01	.03	.04	.01	-.03	.00	.00	.00	.02	.04	.12	.24	.13	.13
13				Dim_4_3	.00	.01	.02	.01	-.01	.01	.00	.01	.01	.02	.11	.13	.20	.20
14				EST	4.24	1.21	-1.16	.30	.95	.96	.85	.79	.25	-.53	.77	1.69	.21	
15				SE	.26	.45	.46	.45	.48	.50	.42	.45	.45	.46	.49	.49	.45	
16				SB	.00	.04	.05	.04	.10	.11	.10	.14	.01	.38	.15	.02	.05	
17				DFE	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	
18				CONST_1	.11	-.02	-.02	.00	.01	.03	.00	-.02	.00	.02	.01	-.02	.00	
19				Dim_1_1	-.02	.32	.17	.15	.01	.02	.02	.01	.00	.02	.00	.05	.01	
20				Dim_1_2	-.02	.17	.33	.17	.02	.02	.05	-.02	-.05	-.01	.03	.06	.03	
21				Dim_1_3	.00	.15	.17	.32	.03	.00	.03	-.03	-.05	.02	.01	.01	.02	
22				Dim_2_1	.01	-.01	.00	.01	.01	.01	.01	-.01	-.01	.01	.01	-.01	.00	

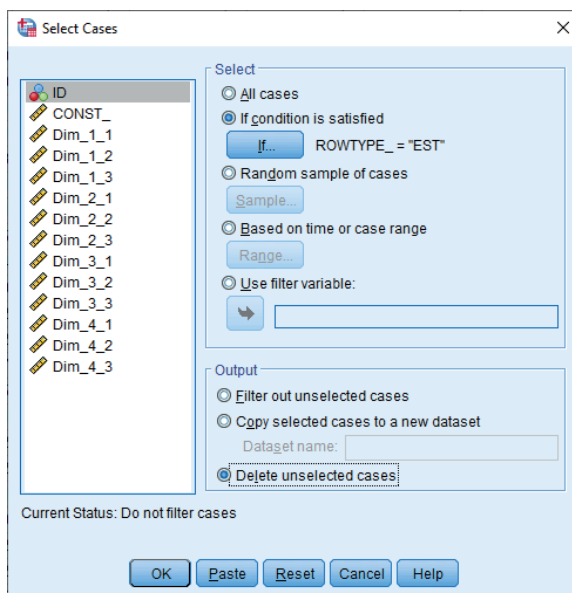
Ha a determinációs együtthatók alapján megvizsgáltuk azt, hogy a mintánk nagy többségén jól működött-e modellünk, akkor ebből az adatbázisból csak a regressziós bétákat tartalmazó sorokra (*ROWTYPE_* változó szerint *EST* értéket felvevő sorok) lesz szükségünk, a többi rekord törölhető. Ezt legegyszerűbben az SPSS *Select Cases* parancsával tudjuk megtenni, amelyet a *Data* főmenüpontban találunk. A felugró ablakban kattintsunk az *If condition is satisfied* opcióra, majd az *If* gombra.



Az újabb felugró ablakon definiáljuk a szűrési logikát; itt kell specificálni azt a szűrési logikát, hogy az SPSS csak a ROWTYPE_ változó EST értéket tartalmazó sorokat tartsa meg.



Miután a *Continue* gombra kattintva visszalépünk a *Select Cases* főablakára, annak az *Output* blokkjában válasszuk ki a *Delete unselected cases* opciót, ugyanis ezzel tudjuk utasítani az SPSS-t arra, hogy az adatbázis többi sorát törölje az adatbázisunkból.



Ezzel létrehoztunk egy olyan adatbázist , amely minden válaszadóra (ID változó) tartalmazza a regressziós bétákat. Ezekkel azonban nehéz dolgozni, illetve sok limitációjuk van. A két legfontosabb az alábbi:

1. Az egyes dimenziókban elhagyott szintekre vonatkozólag nem tartalmaznak részhasznossági értéket,
2. A válaszadónként más skálán mértek, azaz például a válaszadók összehasonlítására nem adnak lehetőséget, illetve a minta egészére vagy alcsoportjaira számított átlagértékeket sem tudunk belőle számolni

Ugyanakkor ezek a limitációk feloldhatók a regressziós béták transformációjával, amelyeket vagy az SPSS *Compute* parancsának használatával, vagy pedig MS Excelben hajthatjuk végre. A béták adatbázisa az SPSS-ből egyszerűen exportálható MS Excelbe, majd szükség esetén a transformált értékek adatbázisa visszaimportálható SPSS-be. Tehát az a döntés, hogy mivel hajtjuk végre a transformációkat, leginkább csak ízlés kérdése. Ebben a könyvben a transformáció logikáját mutatjuk be és rábízunk az olvasóra, hogy milyen szoftver használatával hajtja azt végre.

	ID	DEPVAR_	ROWTYPE_	VARNAME_	CONST_	Dim_1_1	Dim_1_2	Dim_1_3	Dim_2_1	Dim_2_2	Dim_2_3	Dir
1		8 Q17	EST		4.2374	1,2124	-1,1575	2967	,9489	,9571	8515	
2		9 Q17	EST		4.3539	-,1612	,1297	-,0424	2023	-,1772	-,3370	
3		10 Q17	EST		3.9523	,3249	-,5905	-1,0821	,9627	1,0038	6087	
4		12 Q17	EST		4.0459	2,7672	,4316	,9977	,1604	-,0762	5978	
5		13 Q17	EST		3.5544	,1083	-,4532	-1,0715	,9750	,6108	8548	
6		14 Q17	EST		4.8914	-,4099	-2,0149	-1,4384	1,2897	-,5125	-,6036	
7		15 Q17	EST		3.5734	-,4245	-1,7878	-2,8632	,4835	-,6967	-,0383	
8		16 Q17	EST		3.4156	2,3451	-,1894	-1,5355	,0804	-,0887	5004	
9		17 Q17	EST		4.5179	1,0768	,1446	-,4322	3,0231	2,0224	1,2816	
10		18 Q17	EST		4.2116	1,4995	-,1715	-1,5043	1,4545	,8954	6024	
11		19 Q17	EST		3.5391	1,5729	,1548	-,4180	1,3670	,8577	4068	
12		20 Q17	EST		3.4362	,3037	-,5412	-,3449	1,2053	,4468	5488	
13		21 Q17	EST		4.4115	-,1186	-1,5645	,2207	,5808	1,0205	2421	

Az átskálázott részhasznossági értékek előállítás

A nyers béták átskálázása olyan formára, hogy azok alkalmasak legyenek aggregált válaszadói elemzések futtatására, általában úgy hajtjuk végre, hogy a kapott átskálázott részhasznossági értékek az alábbi feltételeknek feleljenek meg:

1. Nullára legyenek centrálva, azaz minden dimenzióban a szintek részhasznossági értékeinek átlaga nulla legyen. Innentől kezdve egy negatív részhasznossági érték azt fogja jelenteni, hogy a hozzá tartozó szint az „átlagosnál” kevésbé vonzó, míg a pozitív azt, hogy az adott szint az „átlagosnál” vonzóbb.
2. A részhasznossági értékek összesített terjedelme az alábbi konstanssal legyen egyelő: Az összesített terjedelmet úgy definiáljuk, hogy minden dimenzióra kiszámítjuk külön-külön a részhasznossági értékek maxi-

mumának és minimumának a különbségét – vagyis a részhasznossági értékek dimenziókénti terjedelmét –, majd ezeket dimenziókénti terjedelem mutatókat egyszerűen összegezzük.

A továbbiakban vegyük a 8-as ID számmal vagy sorszámmal rendelkező válaszadó nyers részhasznossági értékeit vagy bétáit (12.II. táblázat), és azokon keresztül fogjuk bemutatni az átskálázás pontos menetét.

12.11. táblázat: A 8-as ID nyers részhasznossági értékei

Paraméter	Elnevezés	Béta
CONST_	Tengelymetszet	4,2374
Dim_1_1	Ingyenes	1,2124
Dim_1_2	Ingyenes kiszállítás 15.000 forint felett	-1,1575
Dim_1_3	990 forint	0,2967
Dim_2_1	Másnap	0,9489
Dim_2_2	Harmadnap	0,9571
Dim_2_3	Legalább 3 nap, de választott szállítási dátum	0,8515
Dim_3_1	Választott cím, hétköznapokon (9:00 és 18:00 között)	0,7864
Dim_3_2	Választott cím, kétórás időintervallumban	0,2487
Dim_3_3	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	-0,5266
Dim_4_1	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	0,7673
Dim_4_2	Ingyenesen csomagküldőszekrényben (0-24 órában)	1,6894
Dim_4_3	990 Ft csomagküldőszekrényben (0-24)	0,2144

Az **első lépés** rendkívül egyszerű, ugyanis minden dimenzióba be kell szűrni a lineáris regresszió futtatásakor technikai okok miatt kihagyott utolsó szintet (12.12. táblázat). Ezeknek az elhagyott szinteknek a bétája minden esetben pontosan 0. Valójában a lineáris regresszió által a szintekhez kiszámolt minden egyes béta ehhez az elhagyott 0 értékű szinthez képest értelmezendő, másképp fogalmazva, ha valamely szint bétája nagyobb, pozitív szám, akkor az a szint vonzóbb, mint az elhagyott szint, ha pedig negatív, akkor kevésbé vonzó, mint az elhagyott szint.

12.12. táblázat: A lineáris regresszió futtatásakor kihagyott utolsó szint

Paraméter	Elnevezés	Béta
CONST_	Tengelymetszet	4,2374
Dim_1_1	Ingyenes	1,2124
Dim_1_2	Ingyenes kiszállítás 15.000 forint felett	-1,1575
Dim_1_3	990 forint	0,2967
Dim_1_4	Ingyenes hűségprogrammal	0,0000
Dim_2_1	Másnap	0,9489
Dim_2_2	Harmadnap	0,9571
Dim_2_3	Legalább 3 nap, de választott szállítási dátum	0,8515
Dim_2_4	Legalább 5 nap, de választott szállítási dátum	0,0000
Dim_3_1	Választott cím, hétköznapokon (9:00 és 18:00 között)	0,7864
Dim_3_2	Választott cím, kétórás időintervallumban	0,2487
Dim_3_3	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	-0,5266
Dim_3_4	Csomagautomata (0-24 órában)	0,0000
Dim_4_1	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	0,7673
Dim_4_2	Ingyenesen csomagküldőszekrényben (0-24 órában)	1,6894
Dim_4_3	990 Ft csomagküldőszekrényben (0-24)	0,2144
Dim_4_4	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	0,0000

A **második lépésben** megszüntetjük a tengelymetszetet (CONST_ sor). Ezt úgy tesszük meg, hogy a tengelymetszet értékét elosztjuk a dimenziók számával – esetünkben négygel –, majd a kapott értéket ($4,2374 / 4 = 1,0594$) minden dimenzió minden szintjének bétájához hozzáadjuk. Az eredményt a 12.13. táblázat mutatja.

12.13. táblázat: Tengelymetszet nélküli eredmények

Paraméter	Elnevezés	Béta
Dim_1_1	Ingyenes	2,2718
Dim_1_2	Ingyenes kiszállítás 15.000 forint felett	-0,0982
Dim_1_3	990 forint	1,3561
Dim_1_4	Ingyenes hűségprogrammal	1,0594
Dim_2_1	Másnap	2,0083

Paraméter	Elnevezés	Béta
Dim_2_2	Harmadnap	2,0165
Dim_2_3	Legalább 3 nap, de választott szállítási dátum	1,9109
Dim_2_4	Legalább 5 nap, de választott szállítási dátum	1,0594
Dim_3_1	Választott cím, hétköznapokon (9:00 és 18:00 között)	1,8458
Dim_3_2	Választott cím, kétórás időintervallumban	1,3081
Dim_3_3	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	0,5328
Dim_3_4	Csomagautomata (0-24 órában)	1,0594
Dim_4_1	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	1,8267
Dim_4_2	Ingyenesen csomagküldőszekrényben (0-24 órában)	2,7488
Dim_4_3	990 Ft csomagküldőszekrényben (0-24)	1,2738
Dim_4_4	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	1,0594

A **harmadik lépésben** a dimenziók szintjeihez tartozó béták nullára centrálása történik, ami gyakorlatilag annyit jelent, hogy kiszámoljuk minden dimenzióra külön-külön a béták átlagát (12.14 táblázat), majd ezeket az átlag értékeket kivonjuk a megfelelő szintek béta értékeiből – azaz minden átlagot azokból a bétákból vonjuk ki, amelyek alapján az egyes átlagok kalkulálódtak. A nullára centrált béta értékeket a 12.14 táblázat mutatja. Ezen a ponton fontos kiemelni, hogy amennyiben olyan piaci szimulációkat szeretnénk készíteni, ahol a vásárlási döntéseket a *share of preference* modellel fogjuk modellezni, akkor ahhoz, az itt kiszámolt nullára centrált részhasznossági értékeket kell használnunk (12.15. táblázat).

12.14. táblázat: Nullára centrált béták

Dimenzió	Béták átlaga
Dim_1	1,1473
Dim_2	1,7487
Dim_3	1,1865
Dim_4	1,7271

12.15. táblázat: Nullára centrált részhasznossági értékek

Paraméter	Elnevezés	Béta
Dim_1_1	Ingyenes	1,1245
Dim_1_2	Ingyenes kiszállítás 15 000 forint felett	-1,2454
Dim_1_3	990 forint	0,2088
Dim_1_4	Ingyenes hűségprogrammal	-0,0879
Dim_2_1	Másnap	0,2595
Dim_2_2	Harmadnap	0,2677
Dim_2_3	Legalább 3 nap, de választott szállítási dátum	0,1621
Dim_2_4	Legalább 5 nap, de választott szállítási dátum	-0,6894
Dim_3_1	Választott cím, hétköznapokon (9:00 és 18:00 között)	0,6593
Dim_3_2	Választott cím, kétórás időintervallumban	0,1216
Dim_3_3	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	-0,6537
Dim_3_4	Csomagautomata (0-24 órában)	-0,1271
Dim_4_1	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	0,0995
Dim_4_2	Ingyenesen csomagküldőszekrényben (0-24 órában)	1,0216
Dim_4_3	990 Ft csomagküldőszekrényben (0-24)	-0,4534
Dim_4_4	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	-0,6678

A **negyedik lépés** az előzőekben kiszámolt nullára centrált részhasznossági értékek (12.15. táblázat) dimenziókénti minimumának és maximumának meghatározása, majd ezek alapján a dimenziók külön-külön vett terjedelmének kiszámítása, amely nem más, mint a dimenziókénti maximum és minimum értékek különbsége (12.16. táblázat). Végül annyit kell tennünk, hogy a dimenziók terjedelem mutatóinak értékét összegezzük (esetünkben ez 6,3294), majd kiszámoljuk egy osztással azt a szorzót, amellyel majd megszorozzuk a nullára centrált részhasznossági értékeket. A felszorzással az a célunk, hogy az összesített terjedelem megegyezzen a *dimenziók számának százszorosával*. Esetünkben az alkalmazandó szorzó az alábbiak szerint alakul:

$$\frac{4 \times 100}{6,3294} = 63,1971$$

A fenti hányadossal megszorított a 12.15. táblázatból vett nullára centrált részhasznossági értékeket a 12.17. táblázat mutatja.

12.16. táblázat: Dimenziókénti maximum és minimum értékek különbsége

Dimenzió	Minimum	Maximum	Terjedelem
Dim_1	-1,2454	1,1245	2,3699
Dim_2	-0,6894	0,2677	0,9571
Dim_3	-0,6537	0,6593	1,3130
Dim_4	-0,6678	1,0216	1,6894
Összesen	-	-	6,3294

12.17. táblázat: Felszorzott (átskálázott) nullára centrált részhasznossági értékek

Paraméter	Elnevezés	Béta
Dim_1_1	Ingyenes	71,0652
Dim_1_2	Ingyenes kiszállítás 15.000 forint felett	-78,7057
Dim_1_3	990 forint	13,1956
Dim_1_4	Ingyenes hűségprogrammal	-5,5550
Dim_2_1	Másnap	16,4012
Dim_2_2	Harmadnap	16,9195
Dim_2_3	Legalább 3 nap, de választott szállítási dátum	10,2458
Dim_2_4	Legalább 5 nap, de választott szállítási dátum	-43,5665
Dim_3_1	Választott cím, hétköznapokon (9:00 és 18:00 között)	41,6643
Dim_3_2	Választott cím, kétórás időintervallumban	7,6832
Dim_3_3	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	-41,3136
Dim_3_4	Csomagautomata (0-24 órában)	-8,0339
Dim_4_1	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	6,2897
Dim_4_2	Ingyenesen csomagküldőszekrényben (0-24 órában)	64,5638
Dim_4_3	990 Ft csomagküldőszekrényben (0-24)	-28,6520
Dim_4_4	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	-42,2015

Nagyon fontos, hogy az itt bemutatott lépéseket minden válaszadóra külön-külön kell elvégeznünk és ha ezeket megtettük, akkor az átskálázott részhasznossági értékeket már könnyedén felhasználhatjuk a mintánk, illetve tetszőleges alminták jellemzésére, de akár a minta klaszterezésre is. Az átskálázott részhasznossági értékek aggregálására átlagot használjunk, ha szig-

nifikancia vizsgálatra van szükségünk, akkor pedig ANOVA vagy kétmintás t-teszt végezhető.

A 12.18. táblázat a webshopos esettanulmány az aggregált átskálázott részhasznossági értékeket mutatja be nem szerinti bontásban. A férfiak és nők között csak kisebb különbség fedezhető fel a szintek részhasznossági értékei tekintetében, hiszen a szintek sorrendje azonos. Az a következtetés viszont levonható, hogy a teljesen ingyenes kiszállítás kissé jobban elvárt a nők körében, a férfiak viszont kissé türelmetlenebbek, hiszen a másnapi kiszállítást jobban elvárják egy webshoptól. Ugyancsak elmondható, hogy a hölgyek kevésbé szívesen fizetnek 990 forint a visszaküldés esetén, ami összhangban azzal, hogy az ingyenes kiszállítást is inkább elvárják, mint a férfiak.

12.18. táblázat: Aggregált, átskálázott részhasznossági értékek nem szerinti bontásban

Dimenzió	Szint	Férfiak	Nők	Anova p érték
		Átlag	Átlag	
Ár	Ingyenes	59,25	75,06	0,001
	Ingyenes kiszállítás 15.000 forint felett	-27,91	-40,47	0,006
	990 forint	-37,79	-48,40	0,016
	Ingyenes hűségprogrammal	6,45	13,82	0,090
Szállítás teljesítésének ideje	Másnap	30,66	17,24	0,000
	Harmadnap	2,08	-2,29	0,121
	Legalább 3 nap, de választott szállítási dátum	-3,75	-1,04	0,278
	Legalább 5 nap, de választott szállítási dátum	-28,99	-13,91	0,000
Szállítás módja	Választott cím, hétköznapokon (9:00 és 18:00 között)	9,10	11,36	0,612
	Választott cím, kétórás időintervallumban	19,38	15,40	0,348
	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	-17,30	-15,46	0,654
	Csomagautomata (0–24 órában)	-11,19	-11,30	0,980

Dimenzió	Szint	Férfiak	Nők	Anova p érték
		Átlag	Átlag	
Vissza-küldési lehetőség	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	11,86	16,34	0,094
	Ingyenesen csomagküldő-szekrényben (0-24 órában)	7,75	9,61	0,516
	990 Ft csomagküldő-szekrényben (0-24)	-20,36	-27,07	0,033
	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	,74	1,11	0,895

A dimenzió fontossági értékeinek előállításá

Amennyiben kiszámoltuk az átskálázott részhasznossági értékeket (12.17. táblázat), akkor a dimenziók fontossági számainak előállításá már egyszerű feladat. Ehhez ugyanis a dimenziók szintjeihez tartozó részhasznossági értékek terjedelem mutatóira van szükség (12.16. táblázat), mivel a dimenziók fontossága ezeknek a terjedelem mutatóknak *egy konstans összegű skálára vetített megfelelői*, ahol a konstans érték a 100%. Összesen tehát annyit kell tennünk, hogy a dimenziók terjedelem mutatóit összegezzük (6,3294), majd az így kapott számmal elosztjuk a dimenziók terjedelem értékeit (például az első dimenzió esetén a 2,3699 osztva 6,3294-gyel adja a 37,4%-ot). A 8-as ID értékkel rendelkező válaszadó esetén a vizsgált 4 dimenzió fontossági értékeit a 12.19. táblázat mutatja. A vizsgált válaszadóról elmondható, hogy kiszállítási ára a legfontosabb szempont, sőt a szállítási teljesítésének idejéhez képest 2,5-szer olyan fontos.

12.19. táblázat: 8-as ID értékkel rendelkező válaszadó fontossági értékei a négy dimenzió szerint

Dimenzió	Dimenzió neve	Minimum	Maximum	Terjedelem	Fontosság
Dim_1	Ár	-1,2454	1,1245	2,3699	37,4%
Dim_2	Szállítás teljesítésének ideje	-0,6894	0,2677	0,9571	15,1%
Dim_3	Szállítás módja	-0,6537	0,6593	1,3130	20,7%
Dim_4	Visszaküldési lehetőség	-0,6678	1,0216	1,6894	26,7%
Összesen		-	-	6,3294	100,0%

Ugyanúgy, mint a részhasznossági értékek átskálázásánál, a fontossági számok esetén is minden válaszadóra külön-külön el kell végeznünk az előbbi bekezdésben leírt lépést. Az így kapott válaszadói szintű százalékos fontossági számok a minta egészére vagy tetszőleges almintára átlagolással aggregálhatók. Az alminták közötti különbségeket a gyakorlatban ANOVA vagy kétmintás t-teszt segítségével szoktuk vizsgálni.

A 12.20. táblában foglaltuk össze az átlagolással aggregált dimenzió fontossági számokat nem alapján megbontva. A nemek között levonható különbség összhangban van azzal, amit a részhasznossági értékeknél láttunk; a nők kissé érzékenyebbek az árra és válogatósabbak a visszaküldési lehetőség tekintetében, a férfiaknak kissé fontosabb a szállítás teljesítésének ideje, vagyis az, hogy minél gyorsabban jussanak hozzá a megrendelt termékhez.

12.20. táblázat: Átlagolással aggregált dimenzió fontossági értékek nemek szerint

Dimenzió neve	Férfiak	Nők	Anova p érték
	Átlag	Átlag	
Ár	34,71%	38,53%	0,017
Szállítás teljesítésének ideje	22,24%	17,95%	0,000
Szállítás módja	25,39%	24,06%	0,332
Visszaküldési lehetőség	17,66%	19,46%	0,039

A piacmodellezés alapjai

Ahogy az 1. fejezetben bemutattuk, a conjoint elemzéseknél általában nem állunk meg a szintek részhasznossági értékeinek, illetve a dimenziók fontosságának leírásánál, hanem „mi lenne, ha” típusú szimulációkat készítünk. Ezt szokás piacmodellezésnek vagy szimulációnak nevezni, hiszen egy jól megválasztott mintával pontosan azt tudjuk vizsgálni, hogy mi történne a piacon, ha ott az általunk specifikált termékek vagy szolgáltatások lennének elérhetők, azaz azt tudjuk *előrejelezni*, hogy mit vennének a fogyasztók vagy milyen szolgáltatásokat vennének igénybe az ügyfelek.

A piacmodellezés során a kutatásunkba bevont szintekből tetszőleges ajánlatokat alkothatunk és az azok közötti fogyasztói döntéseket a teljes mintán vagy tetszőleges almintákon vizsgálhatjuk, ha a szintek részhasznossági értékeit válasz-

adói szinten becsültük meg. A piacmodellezés során általában nagy számú szcenáriót vizsgálunk meg, amelyek leképezése SPSS-ben kissé nehézkes, ezért érdekesebb inkább az MS Excel programot használni egy szimulátor felépítéséhez. Ennek oka az, hogy az MS Excel gyakorlatilag azonnal újraszámolja az eredményeket a vizsgált ajánlatok módosítása után, és ez a dinamika gyorsabbá teszi a szcenáriók futtatását. Ez a sebesség azért előnyös, mivel a piacmodellezésénél általában sok szcenáriót vizsgálunk meg és egy-egy szcenárió alapján a kapott eredmények tükrében sokszor újabb, módosított piaci helyzetet vizsgálunk meg. A modellezés alapjául szolgáló *share of preference* vagy *first choice modell* rendkívül egyszerűen számolható leképezését jelenti a fogyasztói döntéseknek, így a piacmodellezéshez átfogó matematikai ismeretek helyett inkább némi gyakorlat és kreativitás szükséges.

Az 1. fejezetben egy válaszadón keresztül bemutattuk a *share of preference* és *first choice* modelleket, most ugyanazt a példát felhasználva a teljes mintán, illetve nem bontásban nézzük meg a 12.2. és 12.3. táblában szereplő két ajánlat preferenciáját. Az ottani példa azt vizsgálta, hogy hajlandó-e a szóban forgó válaszadó a másnapi szállításért 990 forintos felárat fizetni. Most azt nézzük meg tehát, hogy a mintánk összességében, illetve a nők és férfiak külön-külön hogyan döntenének a két ajánlat között. Amikor aggregált szinten vizsgáljuk a fogyasztói döntéseket, akkor a válaszadók választásait továbbra is külön-külön modellezzük, majd az aggregálás szintjén a választások arányát (*first choice* modell), vagy átlagát (*share of preference* modell) mutatjuk be. Nagyon fontos kiemelni, hogy ha a *share of preference* modell alapján szeretnénk előrejelezni a fogyasztói döntéseket, akkor a piacmodellezéshez azokat a nullára centrált részhasznossági értékeket használjuk (12.15. táblázat), amelyeket még nem skáláztunk át az összesített terjedelem mutatók alapján.

A 12.21. táblázatban láthatjuk a teljes minta választásának vagy preferenciájának megoszlását az ingyenes, de 5 napot igénylő kiszállításra, illetve a másnapi kiszállítást ígérő, de 990 forintba kerülő szolgáltatások esetében. A döntési modellek közül a *share of preference* megközelítést használva azt látjuk, hogy a mintánk 64,74%-ban az ingyenes, de hosszabb várakozást igénylő alternatívát választaná, ami annak tükrében nem meglepő, hogy a legfontosabb dimenzió az ár lett, azaz a mintánk elég árérzékeny. A pénzbe kerülő, de másnapi kiszállítást igénylő szolgáltatást az esetek körülbelül harmadában venné igénybe a célcsoport, ha csak megjelölt két szolgáltatás lenne elérhető a piacon, vagyis a 990 forintos árat a gyorsaság nem igazán tudja kompenzálni, ami ugyancsak nem meglepő, hiszen a dimenziók fontosságánál azt láttuk, hogy a szállítás teljesítésének ideje jóval kevésbé fontos aspektus, mint a szállítási költség.

A nemek preferenciái tekintetében (12.22. tábla) visszaköszön az a korábban kimutatott eredmény, hogy a nők kissé árérzékenyebbeknek bizonyultak, mint a

férfiak, illetve a férfiak esetében a kiszállítás gyorsasága fontosabb volt, mint a nők esetében. A férfiak az esetek 41,77%-ban választaná a költségesebb, de gyorsabb kiszállítást, azaz még esetükben sem tudná a vizsgált árat kompenzálni a gyorsaság. Ez tehát azt jelenti, hogy a másnapi kiszállításért kérhető felár 990 forint alatt van még a férfiak esetében is. A hölgyeknél a fizetős alternatíva végképp nem versenyképes, hiszen ez az alminta csak az esetek kevesebb, mint harmadában választaná ezt a megoldást. Az ő esetükben jóval nagyobb árengedményt kellene tenünk, ha szeretnénk nagyobb preferencia részesedést elérni ezzel az ajánlattal.

12.21. táblázat: A teljes minta választásának vagy preferenciájának megoszlása az ingyenes, de 5 napot igénylő kiszállításra, illetve a másnapi kiszállítást ígérő, de 990 forintba kerülő szolgáltatások esetében.

Dimenzió	1. ajánlat	2. ajánlat
Ár	990 forint	Ingyenes
Szállítás teljesítésének ideje	Másnap	Legalább 5 nap, de választott szállítási dátum
Szállítás módja	Csomagautomata (0-24 órában)	Csomagautomata (0-24 órában)
Visszaküldési lehetőség	Ingyenesen csomagküldőszekrényben (0-24 órában)	Ingyenesen csomagküldőszekrényben (0-24 órában)
Share of preference modell – Teljes minta	35,26%	64,74%

12.22. táblázat: Nemek szerinti preferencia megoszlása az ingyenes, de 5 napot igénylő kiszállításra, illetve a másnapi kiszállítást ígérő, de 990 forintba kerülő szolgáltatások esetében.

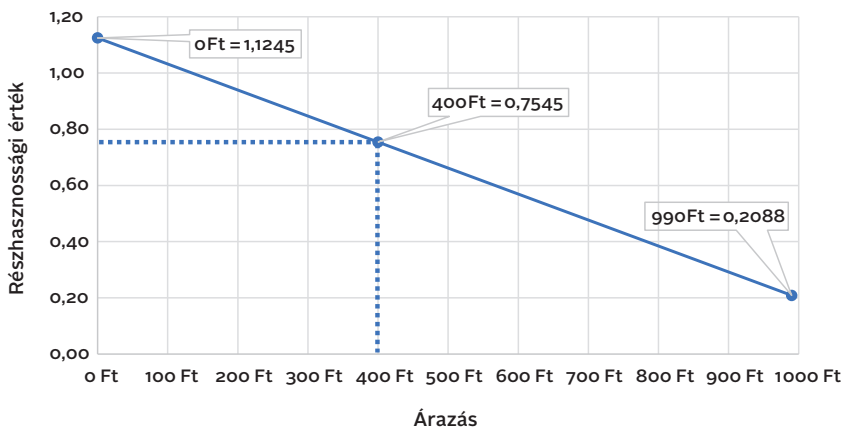
Dimenzió	1. ajánlat	2. ajánlat
Ár	990 forint	Ingyenes
Szállítás teljesítésének ideje	Másnap	Legalább 5 nap, de választott szállítási dátum
Szállítás módja	Csomagautomata (0-24 órában)	Csomagautomata (0-24 órában)
Visszaküldési lehetőség	Ingyenesen csomagküldőszekrényben (0-24 órában)	Ingyenesen csomagküldőszekrényben (0-24 órában)
Share of preference modell – Férfiak	41,77%	58,23%
Share of preference modell – Nők	28,26%	71,74%

A lineáris interpoláció használata

Az előbbieken azt láttuk, hogy 990 forint egy túlzott felár a másnapi kiszállítáért, de mekkora az a felár, amit elkérhet egy webshop ezért a szolgáltatásért a célcsoportban? Ebben a részhasznossági értékek közötti *lineáris interpoláció* tud segítséget adni. Ennek lényege annyi, hogy minden válaszadónál ismerjük az *ingyenes (0 forint)* és a *990 forintos kiszállítás* nullára centrált részhasznossági értékeit, de a 0 és 990 forint közötti árszintekre nem rendelkezünk megfigyeléssel, ezért ezeket becsülnünk kell.

Szemléltetésként ismét vegyük példaként ismét a 8-as azonosítójú válaszadót (12.19. táblázat)! Azt a számot keressük, amit úgy kapunk meg, mintha a 0 forint és a 990 forint részhasznossági értékét összekötjük egy egyenes vonallal, és ennek a vonalnak a 400 forintos árponthoz tartozó Y tengely szerinti koordinátáját vesszük (12.8.).

12.8. ábra: Részhasznossági érték megállapítása a 8-as azonosítójú kitöltő esetében



A 400 forintos árhoz tartozó részhasznossági érték tehát a fentieknek megfelelően az alábbi módon számolható ki a 0 és 990 forintos árszintekhez tartozó részhasznossági értékekből (0,2088 és 1,1245) lineáris interpolációt alkalmazva:

$$1,1245 + (0,2088 - 1,1245) \times \frac{400}{990 - 0} = 0,7545$$

Ha ezt a lépést minden válaszadóra elvégezzük, akkor a teljes mintán modellezni tudjuk a preferencia részesedések alakulását a 400 forintos másnapi és az ingyenes 5 nap várakozást igénylő kiszállítás között. Ebben az esetben a teljes mintán a preferencia részesedések a 12.23. szerint alakulnak.

12.23. táblázat

Dimenzió	1. ajánlat	2. ajánlat
Ár	400 forint	Ingyenes
Szállítás teljesítésének ideje	Másnap	Legalább 5 nap, de választott szállítási dátum
Szállítás módja	Csomagautomata (0-24 órában)	Csomagautomata (0-24 órában)
Visszaküldési lehetőség	Ingyenesen csomagküldő-szekrényben (0-24 órában)	Ingyenesen csomagküldő-szekrényben (0-24 órában)
Share of preference modell – Teljes minta	50,02%	49,98%

Ebben az esetben a másnapi kiszállítást ígérő ajánlat pontosan ugyanakkora preferencia részesedéssel rendelkezik, mint az ingyenes alternatíva. Ez alapján tehát azt a következtetést vonhatjuk le, hogy a másnapi kiszállításért elkérhető maximális felár 400 forint. Ezeket a nevezetes árpontokat általában többszöri próbálkozással tudjuk beazonosítani, ami miatt ez az elemzési technika elég számításintenzív feladat.

A részhasznossági értékek előállítása CBC esetén

A CBC sajátossága, hogy döntési feladatok kevesebb információt szolgáltatnak a válaszadói preferenciákról, hiszen csak azt tanuljuk meg minden döntési feladatban, hogy a válaszadónk mit preferál, de azt nem, hogy milyen mértékben kedvel jobban egy alternatívát a másíknál. Ennek következménye az, hogy válaszadói szinten nagyon nehéz meghatározni a részhasznossági értékeket, arra csak áthidaló technikákkal van lehetőség. Ha megelégszünk a részhasznossági értékek teljes mintán vagy előre definiált almintákon történő becslésével, akkor a *kondicionális logisztikus regresszió* a megfelelő módszer, amely az SPSS-ben sajnos nem elérhető. A jó hír az, hogy az SPSS ezen hiányosságát *Cox-regresszióval* át tudjuk hidalni, ugyanis ez

a módszer helyettesíteni tudja a kondicionális logisztikus regressziót (Hajdú Ottó, A diszkrét kiválasztási modell becslése Cox regresszióval, Szigma, XXXVI. (2005)). Ebben a fejezetben röviden azt fogjuk áttekinteni, hogy az adatainkat milyen struktúrába kell rendezni, illetve a Cox-regressziót hogyan paraméterezzük.

Az adatok előkészítése

A CBC modell kissé más adatstruktúrát igényel a részhasznossági értékek becsléséhez, mint a CVA. Ami közös a két modellben az az, hogy a kísérleti design ajánlatait dummy változókkal kell rögzíteni, ugyanakkor a CBC esetén minden ajánlatot külön sorban kell rögzíteni. A döntési feladatok azonosítására egy futóindexet kell létre hoznunk, amely nem indul újra 1-ről minden válaszadónál, hiszen nem válaszadói szintű becslésben gondolkodunk. Így ennek a futóindexnek a kezdeti értéke értelemszerűen 1, de végső értéke a webshopos esettanulmánynál az alábbi:

$$390 \times 18 = 7020$$

Ennek magyarázata az, hogy 390 válaszadónk van és minden válaszadó 18 döntési feladatot értékelt és ahogy azt az előbbieken jeleztük, minden döntési feladatnak egyedi azonosítóval kell rendelkeznie, ami azt jelenti tehát, hogy a futóindexünk 1-től 7020-ig vesz fel értékeket.

A CBC modell bináris változókat vár függő változóként, ami a gyakorlatban azt jelenti, hogy 1-re kódoljuk a kiválasztott ajánlatot vagy ajánlatokat és 0-ra azt vagy azokat, amelyeket nem választott ki az aktuális döntési feladatban. A jelen esettanulmányban viszont egy 7 fokú Likert skálán jelölte meg a válaszadó azt, hogy a megmutatott 2 ajánlat közül melyiket választott volna inkább. Vagyis a jelen páros CVA modellünket első körben át kell „transzformálnunk” CBC adatokká, ami a válaszok egy egyszerű átkódolási feladata a *CHOICE* változóba az alábbi logika alkalmazásával:

- Ha a válaszadó egy döntési feladatnál 1-3 értéket jelölte meg, akkor a baloldali ajánlatnál a *CHOICE* változónk 1-es értéket vesz fel, míg a jobboldali ajánlatnál 0-ás értéket
- Ha a válaszadó egy döntési feladatnál a 4-es értéket jelölte meg, akkor mind a baloldali, mind pedig a jobboldali ajánlatnál 1-es értéket vesz fel a *CHOICE* változónk. Fontos kiemelni, hogy ebben az esetben jó megoldás nincs, hiszen nem tudjuk valójában azt, hogy a válaszadónk mindkét ajánlatot megvásárolná vagy egyiket sem
- Ha a válaszadó egy döntési feladatnál 5-7 értéket jelölte meg, akkor a baloldali ajánlatnál a *CHOICE* változónk 0-ás értéket vesz fel, míg a jobboldali ajánlatnál 1-es értéket

A Cox-regresszióhoz létre kell hoznunk technikai okok miatt egy úgynevezett *TIME* változót a *CHOICE* változó alapján a következő képlet felhasználásával:

$$TIME = 1 - CHOICE$$

Az adatbázis struktúráját a 12.24. tábla mutatja. Itt jól látszik, hogy a 8-as válaszadóra (*ID* változó) 36 sort hoztunk létre, hiszen minden döntési feladatban 2 ajánlat szerepel. Az eredeti válaszokat az *ANSWER* oszlop mutatja, az abból képzett bináris kódolású választás változó pedig a *CHOICE*. A döntési feladatok futóindexe a *TASK* és jól látszik, hogy a 9-es válaszadóra lépve (*ID*) az index folytatólagos értéket vesz fel, nem indul újra 1-gyel. Az *ATT* változók a szintek dummy változói, amelyek a független változóink lesznek a modellben. Végül az *OFFER* változó az alternatívák beazonosításában segítséget nyújtó futóindex.

Ha valóban CBC megközelítést használtunk volna a webshopos esettanulmányban, akkor nagy valószínűséggel lett volna *NONE* alternatíva a döntési feladatokban. Ennek kódolása az adatbázisban az alábbiak szerint történt volna:

1. Minden döntési feladatban (*TASK*) még egy rekord vagy sor szerepelt volna (*OFFER* = 3)
2. Ezekben a sorokban az *ATT* változók 0-s értéket vettek volna fel
3. Viszont az *ATT* változók kiegészültek volna egy *NONE* változóval, amely minden más alternatívánál (azaz valós ajánlat esetén) 0-ás értéket vett volna fel, míg a *NONE* alternatíva sorában 1-es értéket.

12.24. táblázat

ID	TASK	OFFER	ANSWER	CHOICE	ATT1_1	ATT1_2	ATT1_3	ATT2_1	ATT2_2	ATT2_3	ATT3_1	ATT3_2	ATT3_3	ATT4_1	ATT4_2	ATT4_3	TIME
8	1	1	6	0	0	0	1	0	1	0	0	1	0	0	0	0	1
8	1	2	6	1	1	0	0	0	0	0	1	0	0	1	0	0	0
8	2	1	6	0	0	0	0	0	0	0	0	0	0	0	0	1	1
8	2	2	6	1	0	0	1	0	0	1	0	0	1	0	1	0	0
8	3	1	6	0	0	0	0	0	1	0	0	0	1	0	0	0	1
8	3	2	6	1	0	0	1	1	0	0	0	1	0	1	0	0	0
8	4	1	2	1	0	0	1	0	0	1	0	0	0	0	1	0	0
8	4	2	2	0	0	1	0	0	1	0	1	0	0	1	0	0	1
8	5	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0	0
8	5	2	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1

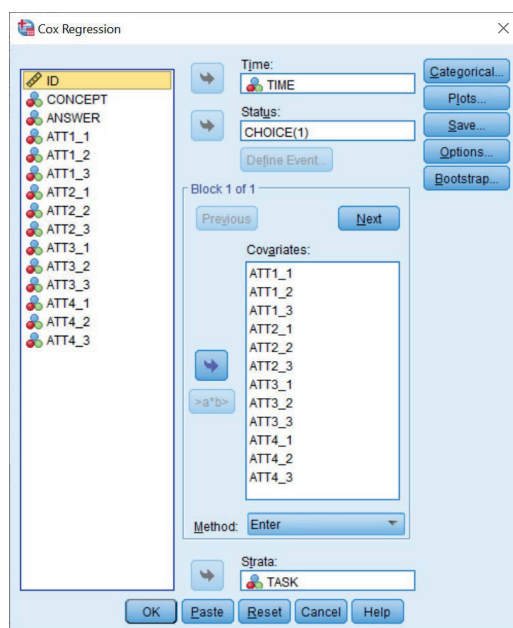
ID	TASK	OFFER	ANSWER	CHOICE	ATT1_1	ATT1_2	ATT1_3	ATT2_1	ATT2_2	ATT2_3	ATT3_1	ATT3_2	ATT3_3	ATT4_1	ATT4_2	ATT4_3	TIME
8	6	1	2	1	1	0	0	0	0	1	0	1	0	1	0	0	0
8	6	2	2	0	0	0	0	0	0	0	0	0	1	0	1	0	1
8	7	1	6	0	0	1	0	0	0	1	1	0	0	0	0	1	1
8	7	2	6	1	1	0	0	0	0	0	0	1	0	0	1	0	0
8	8	1	7	0	0	1	0	0	0	1	0	1	0	0	0	1	1
8	8	2	7	1	0	0	1	1	0	0	1	0	0	0	0	0	0
8	9	1	3	1	0	0	1	0	1	0	1	0	0	0	1	0	0
8	9	2	3	0	1	0	0	1	0	0	0	0	1	0	0	1	1
8	10	1	1	1	0	0	0	1	0	0	0	1	0	0	1	0	0
8	10	2	1	0	0	0	1	0	0	0	0	0	0	0	0	1	1
8	11	1	6	0	0	1	0	0	0	0	1	0	0	0	0	0	1
8	11	2	6	1	0	0	0	0	1	0	0	0	0	0	0	1	0
8	12	1	4	1	0	0	1	0	1	0	0	0	1	1	0	0	0
8	12	2	4	1	0	1	0	1	0	0	0	0	0	0	1	0	0
8	13	1	5	0	0	0	1	0	0	0	0	1	0	0	0	1	1
8	13	2	5	1	1	0	0	0	0	1	0	0	0	0	0	0	0
8	14	1	4	1	0	0	0	1	0	0	0	0	1	1	0	0	0
8	14	2	4	1	0	1	0	0	1	0	0	1	0	0	0	0	0
8	15	1	2	1	0	0	0	0	0	1	0	0	0	1	0	0	0
8	15	2	2	0	0	1	0	1	0	0	0	0	1	0	0	1	1
8	16	1	7	0	0	1	0	1	0	0	0	0	0	1	0	0	1
8	16	2	7	1	1	0	0	0	0	1	1	0	0	0	0	1	0
8	17	1	7	0	1	0	0	0	0	0	0	0	1	0	0	0	1
8	17	2	7	1	0	0	0	0	0	1	1	0	0	0	1	0	0
8	18	1	1	1	1	0	0	1	0	0	1	0	0	0	0	1	0
8	18	2	1	0	0	1	0	0	0	0	0	0	1	1	0	0	1
9	19	1	6	0	0	0	1	0	1	0	0	1	0	0	0	0	1
9	19	2	6	1	1	0	0	0	0	0	1	0	0	1	0	0	0
9	20	1	6	0	0	0	0	0	0	0	0	0	0	0	0	1	1
9	20	2	6	1	0	0	1	0	0	1	0	0	1	0	1	0	0
...

A Cox-regressziót az SPSS *Analyze* főmenüpontjának *Survival* alcsoportjában találjuk. Ha rákattintunk a felugró ablakban (12.9. ábra) az alábbi beállításokat kell megtennünk:

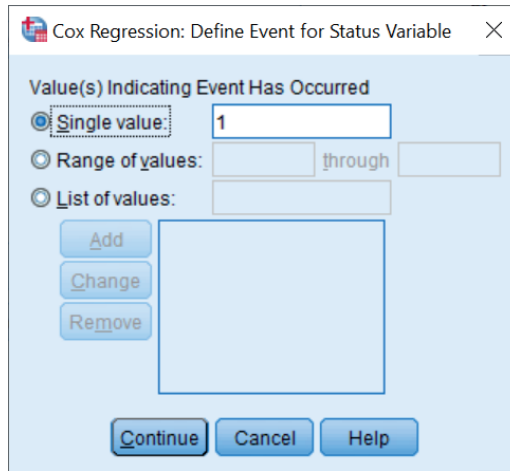
- *Time* beviteli mezőbe rakjuk be a *TIME* változót
- *Status* beviteli mezőbe tegyük be a *CHOICE* változót, majd a *Define Event* gombra kattintva az új ablakon (*Define Event for Status Variable*) a *Single value* opciót válasszuk ki és írjunk 1-et a mellette lévő beviteli mezőbe (12.10. ábra). Végül a *Continue* gombra kattintva léphetünk vissza a Cox regresszió főablakába
- A *Covariates* változók lesznek az *ATT* változók (és ide kerülne a *NONE* változó, ha lenne ilyen a modellünkben)
- Végül a *Strata* alatti beviteli mezőbe kerüljön a *TASK* változó

Ha a felsorolt beállításokat végrehajtottuk, akkor kattintsunk az *OK* gombra és ezt követően lefut a modell.

12.9. ábra



12.10. ábra



A Cox regresszió outputjából a *Variables in the Equation* táblára lesz szükségünk (12.25.). A B oszlop tartalmazza a nyers részhasznossági értékeket, amelyekkel ugyanazon műveleteket kell végrehajtanunk, mint a CVA megközelítés esetén a lineáris regresszió bétáival (kivéve a konstanssal végzendő műveletet, mert itt nincsen konstans a regressziós egyenletben):

1. Minden dimenzióban az elhagyott szintet 0 részhasznossági értékkel vissza kell raknunk
2. Minden dimenzióban a nyers részhasznossági értékeket nullára kell centrálni az átlag érték kivonásával
3. Meg kell határozni a részhasznossági értékek terjedelmét minden dimenzióra, majd azok összegét át kell skáláznunk a *dimenziók száma* $\times 100$ értékre. A két előbbi szám hányadosával meg kell szoroznunk a 2. pontban kiszámított nullára centrált részhasznossági értékeket
4. A terjedelem mutatók alapján meg tudjuk határozni az egyes dimenziók fontossági értékeit

Az átskálázott részhasznossági értékeket a teljes mintára nézve a 12.26. tábla mutatja, míg a fontossági értékeket a 12.27. tábla. Mivel a CBC-nél konfidencia intervallum a 12.25. táblázat *SE (standard error)* oszlopa és a sztenderd normális eloszlásból ismert 1,96-os szorzó segítségével adható. Az 1. dimenzió 1. szintjének nyers részhasznossági értékéhez tartozó a konfidencia intervallum az alábbiak szerint alakul:

$$0,659 \pm 0,045 \times 1,96$$

Ha a fenti 3. lépés alapján átskálázott részhasznosági értékekre szeretnénk konfidencia intervallumot számolni, akkor az előbbi intervallum alsó és felső határát még meg kell szoroznunk az ott meghatározott hányadossal is. Lényeges, hogy a becslésből kihagyott szintekre nem számítható konfidencia intervallum. Ha az alminták különbözőségét szeretnénk vizsgálni, akkor a nyers részhasznosági értékekhez érdemes visszatérni (12.25. táblázat) és azokra kiszámolni a konfidencia intervallumokat, végül megnézni azt, hogy az egyes alminták részhasznosági értékekre számolt konfidencia intervallumai átfednek-e. A nem átfedő konfidencia intervallumok esetén beszélhetünk szignifikáns különbségekről.

12.25. táblázat

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
ATT1_1	,659	,045	213,791	1	,000	1,933
ATT1_2	-,591	,047	157,762	1	,000	,554
ATT1_3	-,746	,049	229,312	1	,000	,474
ATT2_1	,587	,052	126,947	1	,000	1,798
ATT2_2	,275	,050	30,211	1	,000	1,317
ATT2_3	,210	,045	21,493	1	,000	1,234
ATT3_1	,352	,045	61,878	1	,000	1,422
ATT3_2	,510	,054	88,639	1	,000	1,665
ATT3_3	-,015	,046	,112	1	,738	,985
ATT4_1	,131	,045	8,499	1	,004	1,139
ATT4_2	,132	,049	7,103	1	,008	1,141
ATT4_3	-,392	,046	73,150	1	,000	,675

12.26. táblázat

Paraméter	Paraméter neve	Béta
Dim_1_1	Ingyenes	108,98
Dim_1_2	Ingyenes kiszállítás 15.000 forint felett	-55,44
Dim_1_3	990 forint	-75,81
Dim_1_4	Ingyenes hűségprogrammal	22,27

Paraméter	Paraméter neve	Béta
Dim_2_1	Másnap	41,95
Dim_2_2	Harmadnap	0,92
Dim_2_3	Legalább 3 nap, de választott szállítási dátum	-7,62
Dim_2_4	Legalább 5 nap, de választott szállítási dátum	-35,25
Dim_3_1	Választott cím, hétköznapokon (9:00 és 18:00 között)	18,47
Dim_3_2	Választott cím, kétórás időintervallumban	39,22
Dim_3_3	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	-29,84
Dim_3_4	Csomagautomata (0-24 órában)	-27,84
Dim_4_1	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	21,45
Dim_4_2	Ingyenesen csomagküldőszekrényben (0-24 órában)	21,61
Dim_4_3	990 Ft csomagküldőszekrényben (0-24)	-47,34
Dim_4_4	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	4,28

12.27. táblázat

Dimenzió	Dimenzió neve	Fontosság
Dim_1	Ár	46,2%
Dim_2	Szállítás teljesítésének ideje	19,3%
Dim_3	Szállítás módja	17,3%
Dim_4	Visszaküldési lehetőség	17,2%

13 ESETTANULMÁNYOK

Minőségmérés a fesztiválpiacon – Az EFOTT fesztivál példája

Az EFOTT (Egyetemisták és Főiskolások Országos Turisztikai Találkozója) Magyarország egyik nagyobb múltú fesztiválja. Az első EFOTT 1976-ban volt Abaligeten, és a rendezvény 2015-ig vándorfesztiválként működött. Ez idő alatt bejárta az országot a minden évben a házigazda szerepet vállaló egyetem vagy főiskola által választott különböző helyszíneknek köszönhetően. 2015-től azonban a többi hazai fesztiválhoz hasonlóan az EFOTT is egy állandó helyszínen, a Velencei-tó partján kerül megrendezésre.

A fesztivál életében mindig is nagy hangsúlyt fektettek a résztvevők véleményének megértésére, melyet a fesztivál ideje alatt megvalósított kérdőíves kutatással támogattak. A fejlődés, a stratégiai döntések szempontjából kiemelkedő a tapasztaltok, a minőségészlelés megértése. Egy vándorfesztivál életében, ahol minden évben új helyszíni adottságokkal kell a szervezőknek szembenézniük, pedig különösen fontos egy ilyen jellegű mérés, hogy a következő évi új helyszínen a kedvelt dolgok biztosan megjelenjenek, a fejlesztendő területekre pedig nagyobb hangsúlyt fektessenek. Az állandó jellegű helyszíni kérdőíves kutatás mellett ad hoc jellegű kvalitatív és kvantitatív kutatások is készültek a fesztivállal kapcsolatban a nyári, fesztiválszezonon kívül is.

Az EFOTT ideje alatt kezdetben papír alapú kérdőívekkel dolgoztak (PAPI), melyek adatfelvitele jelentősen meghosszabbította a feldolgozási időt. Emellett rengeteg volt a hiányosan kitöltött, vagy félbehagyott kérdőív is. 2014-ben a szervezők azonban számítógéppel (tablettel) támogatott adatfelvételtre váltottak (CAPI), melynek köszönhetően a kitöltések minősége is jelentősen javult, valamint a kutatási elemzések és beszámolók sokkal gyorsabban elkészültek.

A kvantitatív elemzést 500 résztvevő válasza alapján készítették el, melyet a fesztiválra érkezéskor felvett demográfiai adatok alapján (nem, kor, régió) súlyoztak. A kérdőív főbb témái:

- A fesztiválozók leíró jellemzői
- A fesztivál általános értékelése
- A fesztivált támogató cégek, egyesületek láthatósága
- A különböző helyszínek értékelése
- Médiafogyasztási szokások

Ez a fejezet a 2014-es miskolctapolcai EFOTT helyszíni kutatásának eredményeit foglalja össze, mely az utolsó olyan évben készült kutatás, amikor a fesztivál még vándorfesztiválként működött. Ez volt az első év, amikor CAPI adatfelvétellel dolgozott a szervezőcsapat.

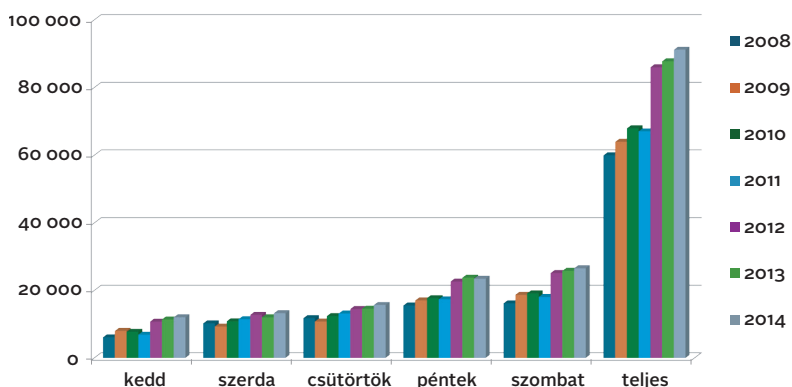
A bemutatott esettanulmány egyszerűbb leíró statisztikai elemzéseket tartalmaz, emellett az összefüggések vizsgálatára keresztábra-elemzés, egyszempontos varianciaelemzés és a könyvben részleteiben nem tárgyalt független mintás t-próba alkalmazását mutatja be.

A 2014-es EFOTT-ozók jellemzése

A jegyeladások alapján látható, hogy a fesztivál folyamatos növekedésen esett át az évek során (13.1. ábra). Az EFOTT-ot ebben az időben egy igazi hallgatói „csúcseseményként” definiálták, hisz a fesztiválon zömében az egyetemista korosztály képviseltette magát: legtöbbször a 19-24 éves korosztályból (65%) vettek részt a rendezvényen, azonban jelentős a 25 éven felüli résztvevők aránya is (29%). Őket az egyetem után is visszajáró, „ALUMNI” szegmensként definiálhatjuk.

A résztvevők 53%-a tanul még felsőoktatási intézményben. Az EFOTT-on résztvevők többsége, a helyszínek köszönhetően Kelet-Magyarországról (28%) érkezett, de kimagasló volt a Budapestről érkezők aránya is (22%).

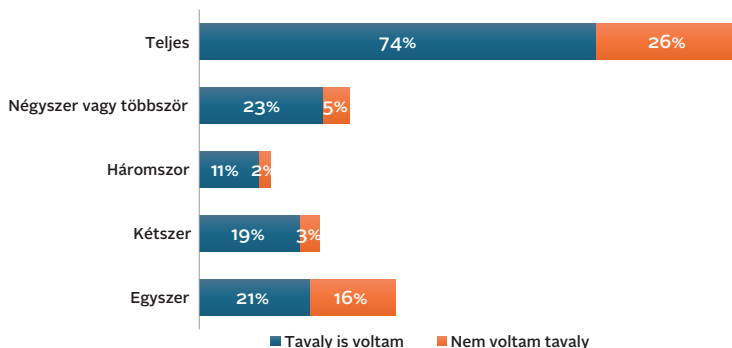
13.1. ábra: Résztvevői adatok alakulása



Forrás: jegyeladások száma

A válaszadók 47%-ának nem ez az első EFOTT-ja: a többségüknek a 2014-es volt a második EFOTT-ja (37%), azonban kiemelkedő az ötször vagy annál többször résztvevők aránya is (28%). A visszatérő vendégek 74%-a tavaly is járt a fesztiválon. (13.2. ábra), tehát magas azoknak az aránya, akik a tavalyi évhez képest valóban értékelni tudják a változásokat, esetleges fejlesztéseket.

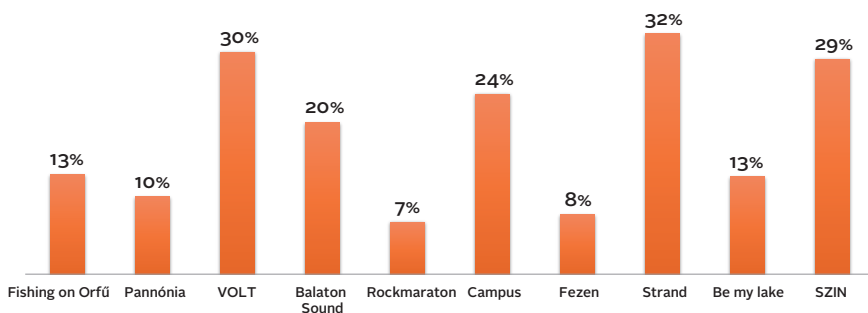
13.2. ábra: Résztétel a korábbi EFOTT-okon



Forrás: helyszíni kérőíves kutatás, N=233 fő.

Az EFOTT-ozók többsége más fesztiválokra nem érhető el, hisz az előző évekhez eredményeihez hasonlóan többségük (51%) idén semmilyen egyéb hazai fesztiválon nem vesz részt. Az egyéb fesztiválokra is meglátogatók körében legnépszerűbbnek a Strand (32%), a VOLT (30%) és a SZIN (29%) számítanak (13.3. ábra). Az előző évekhez képest az EFOTT-on résztvevők körében a Balaton Sound veszített a népszerűségéből (0%-os visszaesés).

13.3. ábra: Egyéb hazai fesztiválok népszerűsége



Forrás: helyszíni kérőíves kutatás, N=235 fő

A 2014-es EFOTT résztvevőit bemutató eredmények alapján megállapítható, hogy a látogatószáma az előző évhez képest tovább nőtt: 2013-hoz képest megközelítőleg 3.400 fővel többen, 91.193-an buliztak az EFOTT-on. A fesztivál fiatalsága és egyetemi jellege továbbra is tart: a résztvevők többsége (65%) a 19-24 éves korosztály tagja, 54%-a pedig még a felsőoktatásban tanul. Még mindig jelentős bázist jelentenek a visszatérő 25 éven felüliek, akik az állandó munkájuk mellett is már évek óta visszatérnek az EFOTT-ra. Az EFOTT-on résztvevők fele (51%) más fesztiválra nem látogat el: kizárólag erre a fesztiválra koncentrálnak a nyár során.

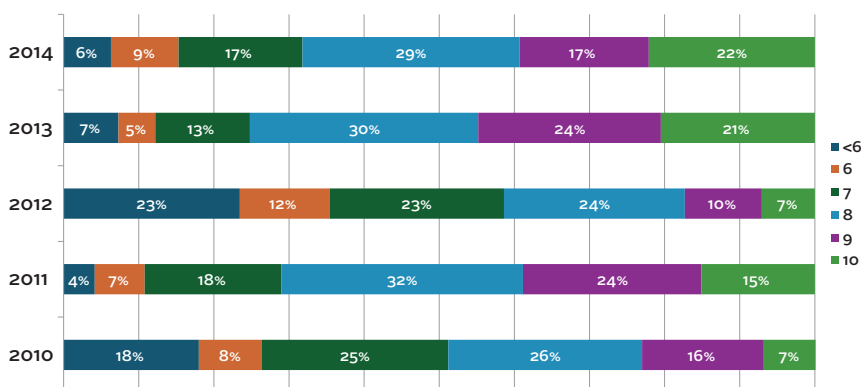
A fesztivál általános értékelése

A 2014-es miskolctapolcai EFOTT-tal összességében a fesztiválozók elégedettek voltak. 1-től 10-ig skálán (1 – teljesen csalódott vagyok, 10 – teljesen elégedett vagyok) az átlagos értékelés 8,02 (szórás=1,62) volt, mely a tavalyi, zánkai értékeléshez hasonló eredményt mutat. A független mintás t-próba alapján sem tekinthető a különbség szignifikánsnak.

Az 1-től 10-ig skálán a megkérdezettek 94%-a adott legalább 6-os értéket. 2013-ban ez 93%-os arányt mutatott, míg 2012-ben 77%-t (13.4. ábra).

A korcsoportok között szignifikáns különbség figyelhető meg az átlagos értékelésben ($F=2,42$, $szig=0,035$). Általánosságban megállapítható, hogy a 28 év feletiek, vagyis az „ALUMNI” szegmens egy része, átlag alatti elégedettséget mutat.

**13.4. ábra: Az EFOTT összegzett megítélése
(1 – teljesen csalódott vagyok, 10 – teljesen elégedett vagyok)**



Forrás: helyszíni kérőíves kutatások

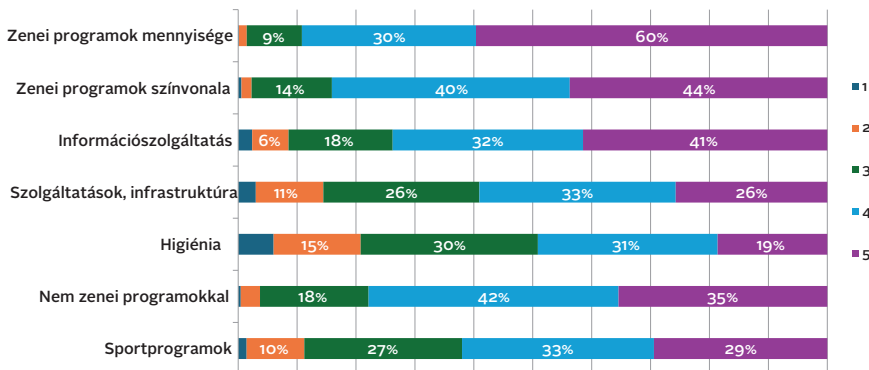
A megkérdezett EFOTT-ozók leginkább a programok mennyiségével és minőségével voltak megelégedve (átlagok 4,47 és 4,25, szórások 0,73 és 0,79), emellett azonban az egyéb, nem zenei és nem sport programok, valamint a helyszíni információs szolgáltatás megítélése is kiválóan mutatkozik (4 feletti átlagok). A 2014-es EFOTT két leggyengébbnek ítélt pontja az infrastruktúra/szolgáltatások és higiénia volt (átlagok 3,67 és 3,4, szórások 1,07 és 1,13), azonban a 2013-as évhez képest a higiénia értékelése ismét jelentős javulást mutat (átlagok₂₀₁₃ 3,01). A független mintás t-próba alapján ez a különbség szignifikánsnak tekinthető ($t=-5,26$, $szig<0,001$).

A zenei programok mennyiségének megítélés minden évben kiváló eredményeket mutat. 2014-ben a megkérdezettek 60%-a volt teljesen megelégedve ezzel a részterülettel. A zenei programok színvonalának megítélése is az előző évhez hasonlóan alakult: a megkérdezettek 84%-a adott legalább 4-es értéket.

Az EFOTT fesztiválon mindig nagy hangsúlyt kaptak a napközbeni, nem zenei programok is, és 2012-től kezdve a sportprogramok különösen kiemelt fókusszal szerepeltek a Sportpont helyszínhez kapcsolódva. 2014-ben a fesztiválozók 18%-a vett részt valamilyen sporteseményen az EFOTT ideje alatt. Az átlagos elégedettség 2014-ben a sportprogramokkal a 2012-es évhez hasonlóan alakult (átlag₂₀₁₄=3,79, szórás=0,84), mely érték 2013-hoz képest kisebb visszaesést jelent (átlag₂₀₁₃=4,18). A független mintás t-próba alapján ez a különbség is szignifikánsnak tekinthető ($t=6,32$, $szig=0,001$).

Az egyéb, nem zenei és nem sport programok (pl. Civil falu, szponzori kitelepülések) megítélése a 2013-as évhez hasonlóan alakult: a megkérdezettek 77%-a elégedett volt ezekkel a programokkal (legalább 4-es érték), mely arány az előző évben 79%-os volt. Az átlagos elégedettség 4,09 (szórás=0,84).

**13.5. ábra: Az EFOTT részterületeinek megítélése
(1 – egyáltalán nem elégedett, 5 – teljes mértékben elégedett)**



Forrás: helyszíni kérőíves kutatás, N=500 fő

Az információszoigáltatás megítélésben javulás figyelhető meg: 2014-ben a megkérdezettek 41%-a volt teljesen elégedett ezzel a részterülettel a tavalyi 32%-kal szemben, az átlagos elégedettség pedig 4,04 volt, mely érték azonban az előző évhez képest nem tekinthető szignifikánsan magasabbnak ($t=-0,83$, $szig=0,205$).

A 2013-as zánkai helyszínhez képest a fesztiválózók kevésbé voltak megelégedve a miskolctapolcai infrastruktúrával, és szolgáltatásokkal: míg 2013-ban a megkérdezettek 71%-a adott legalább 4-es értéket, addig 2014-ben ez mindössze 59% volt. Az átlagos elégedettség 3,67 (szórás=1,07) volt, mely az előző évhez képest szignifikánsan alacsonyabb érték ($t=2,57$, $szig=0,005$).

Összegzésként megállapítható, hogy sikeres EFOTT-ot zártak a szervezők 2014-ben, mivel a 2013-ban hozott jelentős fejlődések nagy részét sikerült szinten tartaniuk. Kivételnek a sportprogramok, illetve a helyszíni infrastruktúra és szolgáltatások számítanak, tehát a 2015-ös fesztiválon ezekre a területekre nagyobb hangsúlyt célszerű fektetni. Évek óta a fesztivál egy kritikus pontja az helyszíni információszoigáltatás minősége, mely 2014-ben ismét javuláson esett át a fesztiválózók értékelése alapján (13.6. ábra).

13.6. ábra: Az EFOTT megítélésének változása

	2010	2011	2012	2013	2014	Különbőség 2014- 2013	Különbőség 2014- 2012	Különbőség 2014- 2011	Különbőség 2014- 2010
Az EFOTT-ról alkotott szummázott vélemény* (1-10 skála)	7,17	8,052	6,848	8,134	8,024	-0,11 ●	1,18 ↑	-0,03 ●	0,85 ↑
Zenei programok mennyisége	4,205	4,337	4,339	4,468	4,472	0,00 ●	0,13 ↑	0,14 ↑	0,27 ↑
Zenei programok színvonala	3,952	4,088	4,031	4,298	4,25	-0,05 ●	0,22 ↑	0,16 ↑	0,30 ↑
Nem zenei programokkal	-	3,798	3,563	4,146	4,091	-0,05 ●	0,53 ↑	0,29 ↑	-
Sportprogramok	-	-	3,79	4,178	3,788	-0,39 ↓	0,00 ●	-	-
Szolgáltatások, infrastruktúra	3,072	3,528	2,81	3,845	3,674	-0,17 ↓	0,86 ↑	0,15 ↑	0,60 ↑
Információszoigáltatás	3,592	4,012	3,339	3,988	4,043	0,06 ●	0,70 ↑	0,03 ●	0,45 ↑
Higiénia	2,657	2,916	2,271	3,001	3,409	0,41 ↑	1,14 ↑	0,49 ↑	0,75 ↑

Csökkenés: különbég < -0,1 (1-5-ig skála) ↓ Stagnálás: különbég < |0,1| (1-5-ig skála) ● Fejlődés: különbég > 0,1 (1-5-ig skála) ↑
< -0,2 (1-10-ig skála) < |0,2| (1-10-ig skála) > 0,2 (1-10-ig skála)

Forrás: helyszíni kérdőíves kutatások

Az EFOTT összegzett megítéléshez hasonlóan szintén szignifikáns különbség figyelhető meg a korcsoportok között a sportprogramok és az egyéb programok megítélés alapján ($F=3,80 / 2,55$, $szig= 0,001 / 0,027$): a 28 év felettek átlag alatti értékelést adtak. Stratégiai kérdésként érdemes elgondolkodni azon, hogy érdemes-e a korcsoportnak az igényeit is magasabb szinten kiszolgálni, vagy továbbra is a fő cél az elsődleges célcsoport, az egyetemisták és főiskolások kiszolgálása legyen.

Az EFOTT helyszíneinek értékelése

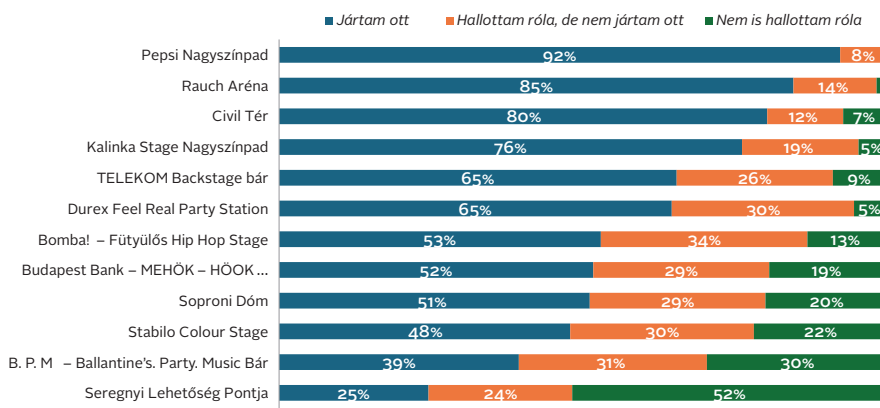
A 2014-es EFOTT több mint 50 helyszínnel és aktivitással várta a fesztiválozókat, melyek három nagy csoportba sorolhatóak:

- zenei programhelyszínek,
- jellemzően „frissítő” helyszínek, és
- egyéb szórakoztató helyszínek.

A kutatás során a megkérdezetteket megkérték, hogy a szervezők által kiemelt 12 helyszínt értékeljék az alapján, hogy jártak-e ott, és amennyiben igen, milyen minőségűnek tartják azt.

A kiemelt helyszínek közül legismertebbnek a Pepsi Nagyszínpad számított: a megkérdezettek 92%-a járt legalább egyszer ott. Második legnépszerűbb helyszín 2014-ben a Rauch Aréna volt: a megkérdezettek 85%-a járt ott legalább egyszer. Azoknak az aránya, akik hallottak az adott helyszínről, de nem jártak ott, a Bomba! – Füttyölős Hip-Hop Stage esetében a legmagasabb (34%), míg a Seregnyi Lehetőség Pontja esetében mondták a legtöbben (52%), hogy nem is hallottak az adott helyszínről.

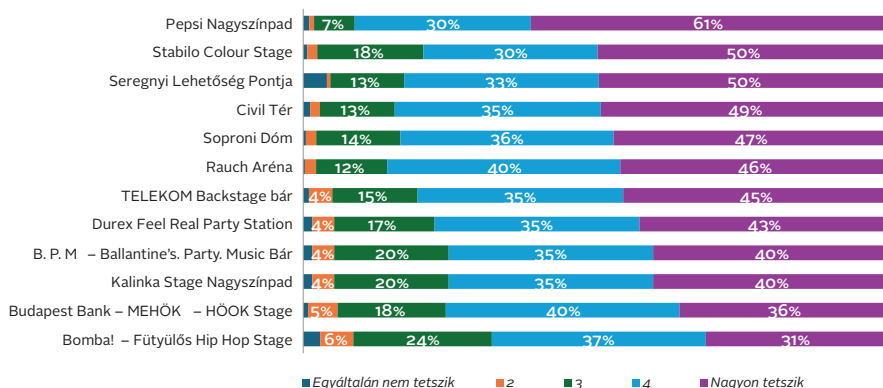
13.7. ábra: Az EFOTT kiemelt helyszíneinek ismertsége



Forrás: helyszíni kéréses kutatás, N=500 fő

Azok, akik legalább egyszer jártak az adott helyszínen, értékelték azt 1-től 5-ig skálán (1 – egyáltalán nem tetszik, 5– nagyon tetszik). Az adott helyszínt ismerőknek a Pepsi Nagyszínpad (61%), a Stabilo Colour Stage, illetve a Seregnyi Lehetőség Pontja (50-50%) tetszett kimagaslóan (5-ös érték).

13.8. ábra: Az EFOTT kiemelt helyszíneinek kedveltsége



Forrás: helyszíni kérdőív kutatás, N= az adott helyszínen járók száma (461 / 239 / 123 / 401 / 423 / 255 / 327 / 380 / 197 / 323 / 258 / 264 fő)

A kutatás legfontosabb tapasztalatai

Az EFOTT szervezőinek célja, hogy a minden évben megismétlődő kutatásuk segítségével megismerjék a fesztiválózók jellemzőit és preferenciáit, valamint megtudják, melyek azok a pontok, amelyeket a jövőben erősíteni kell, vagy a népszerűségük, vagy pedig a nem tetszésük miatt.

Összességében megállapítható, hogy a 2014-es EFOTT értékelése az elégedettség tekintetében az előző évhez hasonló szintet mutat. A részterületek közül idén is leginkább a zenei programok mennyiségével és minőségével voltak elégedettek a résztvevők, emellett kiemelkedően jó eredményt értek el az egyéb nem zenei és nem sport programok, valamint az információszolgáltatás. A fesztivál higiéniai állapotaival való elégedettség növekedése az idén is megfigyelhető volt. Az eredmények alapján továbbra is fejlesztésre szorulnak a sportprogramok, valamint az infrastruktúra.

Marketing mesterszakos hallgatók munkahelyi preferenciáinak elemzése adaptív conjoint módszerrel

Ezzel az esettanulmánnyal az a célunk, hogy conjoint elemzés segítségével megpróbáljuk felmérni a különböző pénzügyi és nem pénzügyi munkahelyválasztással kapcsolatos jellemzők fontosságát marketing mesterszakos hallgatók esetében. Az esettanulmány során az állásajánlatokra úgy tekintünk, mint termékekre, amelyet a hallgatók számára a munkaerőpiacon el lehet adni. Az állások jellemzői így olyan termékjellemzőknek tekinthetők, amelyeket azon jelentkezőknek kínálnak, akik az adott állást betöltik. Az állásért cserébe a munkáltató elköteleződést és a feladatok teljesítéséhez használt készségek meglétét és alkalmazását kéri.

Az esettanulmány során készült mintát egy magyarországi egyetem marketing MSc hallgatói adják. A kérdőívet online küldték a hallgatók számára. Az adatgyűjtés négy félév során történt: a 2017 és 2020 között, a kitöltésért 3 pontot ajánlottak fel a kurzus teljes 100 pontján felül. Az extra motivációs pontoknak köszönhetően a hallgatók 79-86%-a minden félévben részt vett a felmérésben. A teljes minta 273 fő, akik 24,9%-a férfi és 75,1%-a nő volt. Az átlagéletkor 24,20 év volt, 2,26 év szórással.

A kérdőív felépítése

A kutatás során nyolc attribútumot választottunk. Az „otthontól való távolság” és annak szintjei Budapest méretéhez igazodtak. A „nettó jövedelem EUR/hó” és annak öt kategóriális szintje az uniós statisztikák alapján kerültek meghatározásra. A „munkaidő”, a „home office” lehetőségek, a „képzés” és az „extra programok” attribútumok esetében egy háromszintű ordinális skálát használtunk, hogy részletesebben láthassuk a fontosságukat. A „munkáltató” változó szintjei a három tipikus ágazatból származnak, amelyeket a start-upok gyorsan fejlődő munkáltatói típusa szélesített ki. Ez ugyanakkor az egyetlen olyan attribútum, ahol a skála nominális és nem ölt ordinális jelleget. A nyolcadik attribútum a család szerződés-kötési és legális/illegális fizetési módokat mérte. Az attribútumok és szintek részleteit lásd a 13.1 táblázatban.

Mivel az állásra való jelentkezés egy összetett döntés, hiszen a jelentkezők számos szempontot vesznek figyelembe a döntés meghozatalakor, ezért az ACBC conjoint megközelítést választottuk. Egy conjoint lekérdezés során, a válaszadók kénytelenek (kissé irreálisan) gyors döntések sorozatát meghozni, az ACBC módszer pedig segít abban, hogy a kevésbé releváns jellemzőket és alternatívákat kiiktassuk a választási feladatokból, más szóval könnyebbé és reálisabbá teszi a válaszadók számára a döntéshozatalt, ami egy komplex termék-kategória esetében egyértelműen előnyös. Az így keletkező kevesebb „zaj” pedig a pontosabb mérést segíti elő.

Az ACBC 3+1 különböző fázisból építi fel a preferenciamodellt. Az úgynevezett „Építsd meg a sajátod” (Build Your Own – BYO) első fázisban a válaszadónak egy ideális terméket kell leírnia, ami az egyes attribútumok egyéni preferenciaszintjeit határozza meg. A választások alapján több profil is készül a további értékeléshez. Kutatásunkban ez az első fázis módosul, mivel attribútumként ordinális skálákat használunk. A második fázis a „szűrési” fázis, amelyben a válaszadónak el kell dönteniük, hogy az egyes profilokat választani szeretnék-e vagy sem. Ebben a fázisban nincs még kényszerített döntés, hiszen minden profilt elfogadhatnak vagy elutasíthatnak. A „döntési feladatok szakasza” a harmadik fázis, ahol az eddig túlélő profilok (amelyeket a második fázisban választottak ki) képernyőnként harmasával csoportokba kerülnek, ahol a termékek a legtöbb paraméter esetében nem különböznek, így a válaszadónak csak a különböző tulajdonságokra kell összpontosítaniuk. Az itt nyertes koncepciók a harmadik fordulóban versenyeznek egymással, amíg a győztes koncepciót ki nem választják. A +1 választható fázis az úgynevezett „kalibrációs szakasz”, ahol a válaszadónak értékelnie kell a BYO fázisban létrehozott koncepciót és az előző fázis győztesét, valamint 4 elfogadott és elutasított koncepciót. Ezeket a koncepciókat egy 5 pontos skálán kell értékelnie, hogy mennyire valószínű, hogy az adott profilt választaná, vagy sem.

Mindazonáltal fontos megjegyezni, hogy az ebben az esettanulmányban a 8 attribútumból 7 ordinális mérési szinttel rendelkezik, így az ACBC megközelítés „klasszikus” BYO része nem volt alkalmazható a kérdőívünkben. Egy ordinális skála esetében a legvonzóbb szint nem az egyedi válaszadó preferenciáitól függ, hanem nyilvánvalóan a legmagasabb a vágyott szint. Az egyetlen kivételt a munkáltató típusa jelentette, ahol (a hagyományos ACBC megközelítésnek megfelelően) nominális változót alkalmaztunk.

Az első fázisban a nyilvánvalóan ideális, legmagasabb szintek helyett a „már elfogadható” szinteket határoztuk meg. Az ordinális attribútumok szintjeit egymás után mutattuk meg (a legkevésbé vonzóval kezdve) egészen addig a pontig, amikor a válaszadónk azt állította, hogy az attribútum adott szintje már elfogadható az állásajánlat részeként az ő részéről. Minden kérdés egy attribútumra és egy szintre koncentrál, például: 60, 45 és 30 perces utazás a munkahelyre. Legfeljebb 24 kérdés (az összes szint száma, de csak az elfogadott szintig) koncentrál arra, hogy lefedje a hallgatók egyedi preferenciaszintjét az attribútumokkal kapcsolatban. A releváns szintek ismeretében az interjúkat az ACBC-módszertan hagyományos lépéseivel folytattuk.

A második fázisban, az ingerek (állásajánlatok) bemutatása írásbeli prezentációval történt, ahol minden egyes conjoint kártyát egy digitális képernyőn mutattunk meg a válaszadónak. Minden képernyő 3 különböző állásajánlatot mutatott be a válaszadó első fázisban adott válaszai alapján. Összesen 10 képernyőt kellett tehát értékelni, ami 30 nem redundáns állásajánlatot jelent. Minden

egy képernyőn négy attribútum azonos szintű volt (az egyének első fázisbeli választása alapján), négy pedig eltérő. Meg kell jegyezni, hogy a válaszadó akár mindhárom állásajánlatot is elfogadhatta, ha szimpatikusak voltak neki.

A harmadik fázisban, a mérlegelési halmazba előválogatott minden egyes állásajánlatot újra felajánlottunk véletlenszerűen 2-3 ajánlatban/képernyőn, hogy lássuk a kiválasztott csomagok közötti preferenciáikat.

Az eredmények

A conjoint-elemzés tipikus kimenete egy táblázat, amely tartalmazza a relatív fontosságot százalékos arányban az attribútumok teljes halmazához képest. A hasznossági értékek pedig megbecsülik a preferencia szintjét egy-egy egyedi attribútum esetében. A következő táblázat mutatja be ezeket az eredményeket a mi vizsgálatunkban.

Az attribútumok három kategóriába sorolhatóak. Az első kategória csak egy attribútumból, a „nettó jövedelem” attribútumból áll, mivel ez sokkal fontosabb a hallgatók számára, mint a többi. Ennek az attribútumnak a fontossága több mint kétszerese a következő attribútumnak, a „Lakhelytől való távolságnak”. A második kategória három olyan attribútumból áll, amelyek a válaszadók számára közepesen fontosak, mint az „otthontól való távolság”, a „fizetési mód” és a „munkaadó”. A harmadik kategória a legkevésbé fontos kategória, amelyet a „home office” és a „munkaidő”, valamint az „extra programok” és a „képzés” attribútumok alkotnak. Úgy tűnik, hogy ezek a jellemzők kevésbé számítanak a marketing mesterszakos hallgatók számára.

13.1. táblázat: Fontosságok és hasznosságok

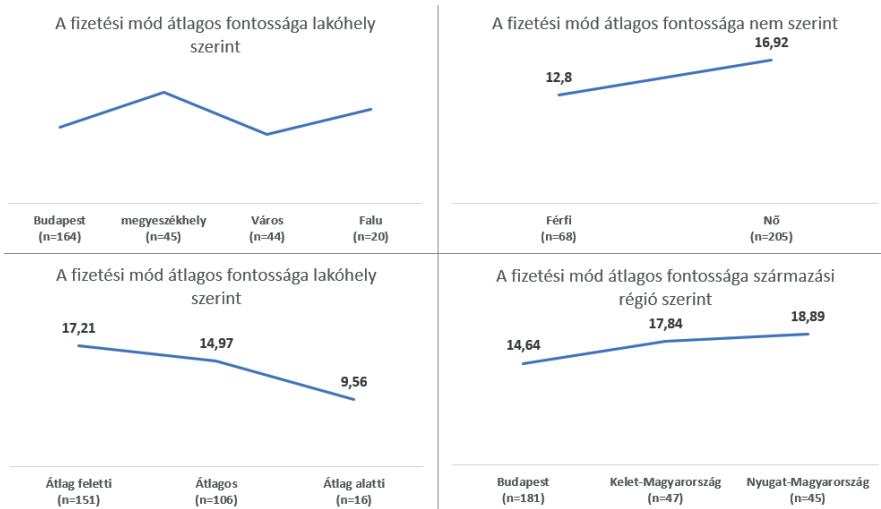
	ATTRIBÚTUM	FONTOSSÁG		SZINTEK	HASZNOSSÁG	
		átlag %	szórás		átlag %	szórás
MAGAS	Nettó jövedelem EUR/hó	39,46	11,06	516	-185,87	52,95
				581	-49,56	34,03
				645	36,38	24,37
				710	69,26	31,20
				774	129,79	43,89
KÖZEPES	Otthontól való távolság (perc)	16,52	5,83	90	-89,28	34,58
				60	6,75	20,41
				45	39,63	14,49
				30	42,90	14,10
				Fizetési módszer	15,89	11,26
Részben bejelentve, részben vállalkozóként és zsebbe fizetve	-9,97	25,01				
Teljesen legális, bejelentve	68,55	49,31				

KÖZEPES	Munkáltató	12,28	5,24	közintézmény	-52,35	35,10
				multinacionális vállalat	4,18	23,02
				magyar tulajdonú cég	21,75	24,22
				start-up	26,41	22,01
ALACSONY	Home office	5,79	4,76	nem lehetséges	-21,33	17,34
				bizonyos esetekben lehet	-3,67	10,02
				korlátlan lehetséges	25,00	21,76
	Munkaidő	4,06	3,83	fix	-15,74	16,23
				törzsidő, szabad kezdéssel és befejezéssel	-1,03	8,17
				teljesen rugalmas	16,77	15,48
	Extra programok	3,01	2,84	nincsenek	-8,06	7,59
				önfinanszírozással lehetséges	-8,00	7,56
				ingyenesek	16,05	15,14
	Képzések	2,98	2,35	nincsenek	-13,08	11,35
				esetenként	2,29	5,66
				tervezett tréning programok	10,80	8,00

A kutatás során kíváncsik voltunk arra is, hogy az egyes demográfiai tényezők mentén adódik-e eltérés az attribútumok fontosságát tekintve. Ehhez ANOVA vizsgálatokat végeztünk a következő jellemzők alapján: „nem” (férfi / nő) „életkor” (max. 23, 24, 24 év felett) „állandó lakóhely típusa” (Budapest / megyeszékhely / város / falu), „tanulmányi eredmény” (átlag alatti / átlagos / átlag feletti) és „származási régió” (Nyugat-Magyarország / Budapest / Kelet-Magyarország). Az életkor nem mutatott szignifikáns különbséget egyik attribútum esetében sem.

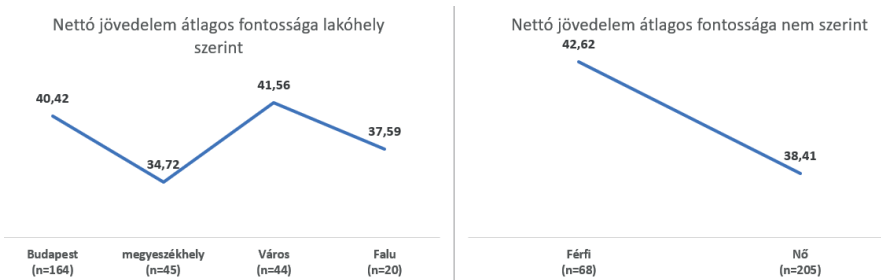
A fizetési mód több tényező mentén is mutatott eltéréseket. Az „állandó lakóhely típusa” (Levene $p=0,914$, $F p=0,001$) szignifikáns hatást mutat. A fizetési mód a megyeszékhelyről érkezők számára a legfontosabb ($M=21,53$), a városokból származók számára pedig kevésbé fontos ($M=13,29$). A „fizetési mód” a nők számára fontosabb (16,92), mint a férfiak számára (12,80) (Levene $p=0,897$, $F p=0,009$). A „tanulmányi eredmény” szintén befolyásolja a „fizetési mód” fontosságát (Levene $p=0,139$, $F p=0,020$). Minél jobb eredményt érnek el a hallgatók, annál fontosabbnak értékelik ezt az attribútumot, az átlag feletti átlagos fontosság 17,21, az átlagos hallgatók esetében 14,97, az átlag alattiaknál pedig 9,56. Az utolsó változó, amely befolyásolja ezt az attribútumot, a „származási régió” (Levene $p=0,588$, $F p=0,032$). A budapesti származású hallgatók a legkevésbé érzékenyek a „fizetési módra” ($M=14,64$), míg a kelet-magyarországi ($M=17,84$) és a nyugat-magyarországi ($M=18,89$) származású hallgatók jobban összpontosítanak erre az attribútumra.

13.9. ábra: Fizetési mód fontosságának eltérései



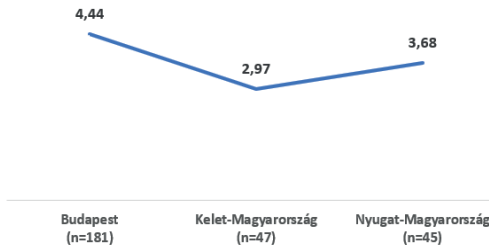
A „nettó jövedelmet” két változó befolyásolja. Az egyik az „állandó lakóhely típusa” (Levene $p=0,127$, $F p=0,008$). A jövedelem mértéke a legfontosabb azok számára, akiknek állandó székhelye városokban van ($M=41,56$), és kevésbé fontos azok számára, akiknek állandó lakcíme megyeszékhelyen van ($M=34,72$). A nettó jövedelem fontosságára szignifikáns hatással van a nem (Levene $p=0,110$, $F p=0,006$), mivel a férfiak számára fontosabb a nettó jövedelem ($M=42,62$), mint a nők számára ($38,41$).

13.10. ábra: A nettó jövedelem fontosságának eltérései



A „munkaidő” fontosságának esetében csak egy befolyásoló változó van, a „származási régió” (Levene $p=0,058$, $F p=0,048$). Ez a jellemző a Budapesten született és nevelkedett hallgatók számára a legjelentősebb ($4,44$), és a kelet-magyarországi származású hallgatók számára a kevésbé fontos ($2,97$).

Munkaidő átlagos fontossága származási régió szerint



Következtetések

A munkahelyválasztáshoz kapcsolódó attribútumok három csoportba sorolhatók a legfontosabbaktól a legkevésbé fontosig. Az egyetemi hallgatók preferenciái között a bér dominál. A közepesen fontos jellemzők a „lakóhelytől való távoltság”, a „fizetési mód” és a „munkáltató típusa”. E három jellemző közül csak az egyiket, a „fizetési módot” fejleszthetik a munkáltatók rövid távon – a másik két jellemző rövid távon inkább fix, azonban a munkába járást támogathatják. A legérzékenyebb jellemzőnek a „fizetési mód” tűnik, mivel a „nem”, a „származási régió”, az „állandó lakóhely típusa” és a „tanulmányi eredmény” is befolyásolja a hallgatók értékelését. A legkevésbé fontosak azok a jellemzők, amelyeket a munkáltató rövidebb távon fejleszthet, mivel ezek a rugalmasabbak, mint például a „képzés”, a „munkaidő”, a „home office” és az „extra programok”.

A munkaadóknak érdemes a versenyképes fizetési ajánlatokat a marketing mesterszakos hallgatóknak szánt álláshirdetéseik középpontjába helyezni. A jövedelem egyértelműen a legfontosabb szempont, így a magasabb fizetési csomagok vonzóbbá tehetik az állásajánlatokat. A tanulmányi eredmények alapján differenciált ajánlatokkal célozhatók meg a hallgatók. Az átlag feletti eredményekkel rendelkező hallgatók számára kiemelhető, hogy az ő teljesítményük alapján magasabb fizetési csomagot kínálnak. Bár az „otthontól való távolság” és a „home office” kevésbé fontosak, a rugalmasság lehetősége továbbra is vonzó lehet. A munkaadóknak érdemes lehet rugalmas munkaidő lehetőséget biztosítani, hogy vonzóbbá tegyék az ajánlatokat. Az állandó lakóhely típusa, a nem, a tanulmányi eredmény és a származási régió alapján célozott álláshirdetések hatékonyabbak lehetnek. Például a vidéki városokban élő hallgatóknak szóló hirdetésekben kiemelhetők a fizetési mód előnyei és a nettó jövedelem nagysága.

A MOL Bubi használata a koronavírus-járvány idején

Jelen esettanulmány célja a lineáris regressziós módszertan alkalmazásának bemutatása egy konkrét kutatási projekten keresztül. Az esettanulmány alapjául szolgáló kutatás és cikk a *Transportation Research Interdisciplinary Perspectives* című folyóiratban jelent meg 2022 szeptemberében (<https://doi.org/10.1016/j.trip.2022.100674>).

2020 elején egy újfajta koronavírus járvány, a Covid19 kezdett terjedni a világon és így Európában is. Az Egészségügyi Világszervezet 2020. március 11-én globális világjárvánnyá nyilvánította a Covid19-et. A kormányok azonnal reagáltak a betegség gyors terjedésére, és számos korlátozást vezettek be a társadalmi érintkezésre vonatkozóan. A távolságtartás, a karantén és a különféle lezárások kihatottak az utazási és ingázási szokásokra, valamint a különböző közlekedési módok használatára is.

Mivel a Covid19-járvány könnyen terjedt a társas érintkezésen keresztül, az emberek félni kezdtek a közösségi közlekedéstől, és inkább az egyéni utazási módokat részesítették előnyben. Számos város vezetett be intézkedéseket a kerékpározás népszerűsítésére és a kerékpáros infrastruktúra fejlesztésére. Ezek az intézkedések elősegítették a közösségi kerékpárrendszerek használatát is, ami a városi közösségi közlekedés egyszerű és olcsó alternatívája. A közösségi kerékpárrendszer egy alapvetően bérlet alapú szolgáltatás, amely lehetővé teszi a felhasználók számára, hogy az előre telepített gyűjtőállomások között kerékpárral utazzanak. Ezáltal tehát a felhasználók saját kerékpár nélkül is tudnak kerékpározni a városban, mindezt pedig jellemzően olcsón és könnyedén.

Magyarország sem volt kivétel a Covid19-járvány hatásai alól. Az első esetet 2020. március 4-én jelentették be, és néhány nappal később már nagyon szigorú intézkedésekről döntött a kormány. Az egyetemek március 12-től áttértek online oktatásra, az általános és középiskolák pedig március 16-tól követték a példát. A kutatás szempontjából releváns, hogy Karácsony Gergely, Budapest főpolgármestere március 16-tól a helyi közösségi kerékpárrendszer, a MOL Bubi menetdíját jelentősen csökkentette, mindössze 100 Ft lett a havibérlet ára.

A közösségi kerékpárrendszer használatára tehát két ellentétes erő hatott. Egyrészt, a lezárások és korlátozások jelentősen csökkentették a Budapesten közlekedők számát, az otthonról történő munkavégzés gyors elterjedésével pedig az ingázók száma is visszaesett. Ez önmagában csökkentette a közösségi kerékpárrendszerek használatát. Másrészt viszont a világjárvány idején jelentősen megnőtt a kerékpározás modális aránya, ami felfelé irányuló nyomást gyakorolt a teljes használatra. Mivel a közösségi kerékpárrendszerek könnyen használhatók a városokon belül, nem igényelnek jelentős beruházást a felhasználók részéről (nem szükséges hozzá saját kerékpárt vásárolni), és kombinálhatók más közlekedési módokkal is, így a használat még jelentősebb növekedése képzelhető el.

Esettanulmányunkban azt vizsgáljuk meg regressziós modellek segítségével, hogy a különböző tényezők hogyan hatottak a Bubi használatára 2020-ban.

Felhasznált adatok

A Bubi használati adatait a rendszer üzemeltetőjétől, a Budapesti Közlekedési Központtól (BKK) kaptuk meg. Az adatok a 2019-ben és 2020-ban megtett összes egyéni utazás adatát tartalmazzák (kezdőállomás, záróállomás, kezdő időpont, záró időpont, a felhasználó származási országa a telefonszáma előhívója alapján, jegy típusa, felhasználói azonosító). A nyers adatbázis összesen 771 459 sort tartalmazott, de ez némi tisztításra szorult még (egy percnél nem hosszabb utazás, hiányzó kezdő- vagy záróállomás esetén töröltük az adott sort), így az adatbázis 753 356 utazásra csökkent. Az utazások számát és átlagos időtartamát naponként és kezdőállomásonként összegeztük. Mivel a Bubi használati szokásai eltérők munkanapokon és hétvégén, ezért jelen esettanulmányban kizárólag a munkanapi utazásokat vizsgáljuk. Ezek elsősorban a munkába járáshoz kapcsolódnak, ennek megfelelően egy reggeli és egy délutáni csúcs figyelhető meg. Munkanapokon történt az utazások 77%-a 2019-ben és 74%-a 2020-ban.

Az egyszeri (jegyet váltó) felhasználók és a bérletet használók nem különböztek el egymástól, mivel az egyszeri felhasználók 2020-ban teljesen eltűntek (az összes utazás 0,9%-át tették ki abban az évben). Ez két okra vezethető vissza. Először is, 2019-ben a Bubi-felhasználók mintegy 10%-a volt külföldi, vélhetően turista, és szinte mindegyikük (97%-uk) egyszeri felhasználó volt. A Covid19-járvány miatt a külföldi turisták eltűntek Budapestről (2020-ban a felhasználók mindössze 0,6%-a volt külföldi), így csökkent az egyszeri felhasználók száma. Másrészt, a havibérlet árának csökkentése után (2020. március 16-tól) más típusú jegyeket és bérleteket egyáltalán nem vásároltak, lényegében mindenki havibérlettel használta a Bubit.

A Covid19-járvánnyal kapcsolatos kormányzati intézkedéseket az Oxfordi Egyetem által publikált COVID-19 Government Response Tracker (<https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker>) adatai segítségével számszerűsítettük. Az ún. Government Stringency Index (kormányzati szigorítások indexe) egy nullától százig terjedő skálán mért mutató. Azért esett erre a választásunk, mert ez a mutató az összes hivatalos korlátozást tükrözi, amelyek az emberek életére, mindennapi rutinjaira és viselkedésére hatással voltak.

Másrészt, a mobilitást a kormányzati korlátozásokon kívül más tényezők is befolyásolhatják (pl. a világjárványtól való félelem, az otthonról dolgozók aránya). Ez kihat a közlekedési módok iránti keresletre is. Az általános mobilitást egy ún. proxy változóval mértük, a budapesti városi közlekedés napi forgalmával. A BKK több út forgalmát monitorozza az utak mellé vagy fölé telepített járműszámláló kamerákkal. A kamerák folyamatosan ugyanazt az útszakaszt érzé-

kelik, amely időben nem változik. Mivel csak korlátozott számú kamera van, e forgalomszámok önmagukban nem relevánsak, de a változások jól tükrözhetik a mobilitás változását.

Ezek mellett a közösségi kerékpárrendszerek használatát meghatározó egyéb tényezőket is beépítettük a modellünkbe. Ezek közül az egyik legfontosabb az időjárás: a napi átlaghőmérséklet, a napi összes csapadékmennyiség és a napi átlagos szélsőbesség. Ezeket az adatokat Budapestre vonatkozóan a NASA Power Data Access (<https://power.larc.nasa.gov/data-access-viewer/>) adatbázisából gyűjtöttük le. Mivel a hőmérséklet hatása a közösségi kerékpárrendszerek használatára jellemzően nem lineáris (túl hidegben és túl melegben sem olyan kellemes kerékpározni), 5°C-os intervallumokat alkalmaztunk.

Továbbá a gyűjtőállomások kihasználtsága jelentősen függ a környezetüktől. A területhasználat és az épített környezet jelentősen befolyásolhatja az utazásszámokat is. Ez azonban rövid távon meglehetősen stabil, ezért időben nem változó hatásként kezelhető, ha az elemzés korlátozott időhorizontot vesz figyelembe. Ez jelen esetben különösen igaz, mivel a Bubi működési területe a város belső, már beépített részét fedi le; ezeken a területeken a fejlesztési projektek ritkábban lehetnek, bár van rájuk példa. Ezért mind a 160 állomást egy-egy terület-használati kategóriába soroltuk, és feltételeztük, hogy ez változatlan maradt az elemzett időhorizont alatt. Összesen hét területhasználati kategóriát különböztettünk meg, amelyek a következők voltak:

- Lakóövezet (46 állomás): az állomás környéke elsősorban lakóövezet;
- Kereskedelmi övezet (36 állomás): az állomás környéke elsősorban üzleti vagy kormányzati terület;
- Oktatási övezet (19 állomás): az állomás egy egyetemi kampusz mellett található;
- Parkos terület (24 állomás): az állomás környéke elsősorban pihenőterület közparkkal;
- Közlekedési csomópont (17 állomás): az állomás egy közlekedési csomópontban található;
- Egészségügyi övezet (4 állomás): az állomás egy kórház mellett található;
- Szórakozási övezet (14 állomás): az állomás kulturális vagy szórakoztató létesítmények (pl. múzeum, színház, koncertterem, szórakozóhely, strand) közelében található.

A jegyek és bérletek árai jelentős hatással vannak még a használatra. A vizsgált időszakban azonban csak egyetlen árcsökkenés történt (2020. március 16-tól kezdődően, ahogy korábban már bemutattuk). Mivel ez az árcsökkentés közvetlenül a járvány kezdetén történt (a helyzetre adott gyors válaszlépés részeként), nem lehet megmérni a Bubi használatára gyakorolt hatását. Ebben a tanulmányban azonban

különböző időhorizontokat alkalmazunk, és mivel az árcsökkenés közvetlenül a vizsgált időszak elején történt, ezért véleményünk szerint nem érvényteleníti az eredményeket. A felhasznált változók leíró statisztikáit a 13.2. táblázat tartalmazza.

13.2. táblázat A felhasznált változók leíró statisztikái

Változó	Megfigyelések száma	Átlag	Szórás	Min.	Max.
Bubi utazások száma (naponta, állomásonként)	68 833	8,00	8,21	0	89
Bubi utazások átlagos időtartama (perc)	63 876	14,48	13,82	1	1 205
Kormányzati szigorítások indexe	481	22,04	28,87	0	76,85
Napi városi forgalom (autók száma)	481	162 019	20 943	89 359	212 044
Napi átlaghőmérséklet (°C)	481	11,74	8,62	-6,57	27,84
Napi összes csapadék (mm)	481	1,93	4,17	0	42,28
Napi átlagos szélesebesség (m/s)	481	3,46	1,45	0,88	9,12

Módszertan: lineáris regressziós modell

A kormányzati korlátozások és a közösségi kerékpárrendszerek használata közötti ok-okozati kapcsolat azonosításához a megfigyelt utazási mintázatokat össze kell hasonlítani egy korlátozás nélküli alternatívával, a kettő közötti eltérés mutatja az intézkedések hatásait. Sajnos ez az ún. tényellentétes állapot nem áll rendelkezésre, hiszen nem tudjuk, hogyan alakult volna a Bubi-használat 2020-ban, ha nem lett volna koronavírus-járvány. Ezért ezt a tényellentétes állapotot nekünk kell megbecsülnünk, amelyet a világgjárványt megelőző év azonos időszakának használati adataival tettünk meg, azaz a 2019-es adatokat használtuk a 2020-as év Covid19-járványhoz kapcsolódó korlátozások nélküli tényellentétes állapotának. Ugyanakkor a két év közötti eltérésekre kontrollálni kell. Ezek két részre oszthatók. Először is, több, az utazások számát meghatározó magyarázó változó időben változik (pl. hőmérséklet, csapadék, szélesebesség, forgalom), és alapvetően eltérő értékeket vett fel 2019-ben és 2020-ban. Másodszor, az egyes gyűjtőállomások használati trendjei eltérők lehetnek a területhasználat, az épített környezet és a közlekedési hálózat fejlettsége közötti különbségek miatt. Ezt is figyelembe kell venni a modellben.

Az elemzés során a 2019-ben és 2020-ban az év adott napján, adott gyűjtőállomásról indított utazások számának különbségeit vettük. Ezt úgy számítottuk

ki, hogy a 2020-as adatokból kivontuk a 364 nappal (pontosan 52 héttel) korábbi adatokat. Ezt a különbséget magyaráztuk az időben változó magyarázó változókkal, illetve az állomásszintű trendekkel. A modell formálisan az alábbi volt:

$$\Delta y_{it} = \beta_0 + \beta_1 \Delta Stringency_t + \beta_2 \Delta \log(Traffic_t) + \Gamma \Delta X_t + c_i + u_{it}, \quad (13.1)$$

ahol y_{it} az adott i -edik Bubi-gyűjtőállomásról indított utazások száma a t -edik napon vagy azok átlagos hossza percben mérve, $Stringency_t$ a kormányzati szigorítások indexének értéke a t -edik napon Magyarországon, $Traffic_t$ Budapest forgalmi adatait tükrözi a t -edik napon, X_t tartalmazza a fentiekben vázolt, időjárással kapcsolatos magyarázó változókat, c_i az állomásszintű változás egyik évről a másikra (állomásszintű trend), míg u_{it} a véletlen hibát jelenti. A Δ a 2019 és 2020 közötti különbségre utal. Mivel 2019-ben nem voltak érvényben kormányzati korlátozások, így $\Delta Stringency_t = Stringency_t$ ebben az esetben. A modell becslése egyszerű lineáris regressziós módszertannal történt, ahol c_i az egyes állomások dummy változója volt.

Ennek a regressziónak az az előnye, hogy a kormányzati szigorítások, a mobilitás és az időjárási körülmények *ceteris paribus* hatásait is meg lehet becsülni. Az adatok azt mutatják, hogy a kormányzati korlátozások és a forgalom nem korrelálnak erősen egymással, a korrelációs együttható $-0,50$, ha a Covid19-járvány első időszakát vesszük figyelembe, de csak $-0,25$, ha a teljes 2020-as évet vizsgáljuk. Ez azt jelenti, hogy a korlátozások nem hatnak közvetlenül a forgalomra, a korreláció abszolút értékben nem túl magas, ezért nem okoz multikollinearitási problémákat. Ennek az lehet az oka, hogy például a társas összejövetelek tilalma azt jelentheti, hogy az emberek csak a legközelebbi rokonaikkal találkoznak, de ennek érdekében mégis utaznak. Ha ezt a korlátozást feloldják, elképzelhető, hogy az emberek ugyanannyit ingáznak, de más tevékenységeket végeznek, és más közlekedési módokat használnak. Ezzel együtt a közlekedési módok megválasztását befolyásolhatják a kormányzati szigorítások, és jelen elemzés célja éppen ennek kimutatása.

Ebben a regressziós modellben nem lehetséges az időben állandó magyarázó változók becslése, mivel ezeket a változókat az állomás dummy változók (c_i az (13.1) egyenletben) felveszi. Ezért a területhasználat nem vonható be ebbe a modellbe. A területhasználat Bubi-utazásokra gyakorolt hatásainak értékeléséhez az átlagos állomásszintű változásokat, tehát a c_i -ket a fent bemutatott hét területhasználati típus szerint átlagoltuk. Ezek az átlagok azt mutatják, hogy a különböző területhasználati típusokkal rendelkező állomások kihasználtsága hogyan változott a Covid19-járvány során.

A regresszióelemzés a Bubi csúcsidőszakára összpontosított, de ezen belül két időszakot különböztettünk meg: 1) a világjárvány első hullámát, és 2) az első

és a második hullám közötti időszakot. A szétválasztás lehetővé tette annak ellenőrzését, hogy megfigyelhető-e bármilyen hosszabb távú hatás. Ha a Bubi utazásszáma az első hullám alatt a szigorításokra adott válaszként megnőtt, de a pandémiás helyzet javulása után visszatért a járvány előtti szintre, akkor nem volt megfigyelhető semmilyen hosszabb távú hatás. Ha azonban az ingázók az első hullámban kipróbálták a Bubit, és hosszú távú felhasználókká váltak, akkor az intézkedések feloldása valószínűleg nem vagy csak kisebb mértékben befolyásolja a Bubi használatát (a mobilitásra való kontrollálás után), amely továbbra is magas szinten maradna. Ez különösen igaz lehet a munkanapi használatra, amely erősen kapcsolódik a munkába járáshoz. Ha az emberek napi ingázási szokásai a járvány első hulláma alatt megváltoztak, akkor a korlátozások megszűnése után megmaradhatnak az új szokások, tehát többet fogják használni a Bubit. Ez az a hosszabb távú hatás, amelyet meg szeretnénk becsülni.

Eredmények

Az előző részben bemutatott regressziós modell becslésének eredményeit a 13.3. és a 13.4. táblázatok mutatják.

A világjárvány első hullámában a kormányzati intézkedések nagymértékben befolyásolták a Bubi használatát munkanapokon. A kormányzati szigorítások indexének minden egyes értékkel való növekedése minden más változtatlansága mellett átlagosan 0,051-gyel növelte az állomásonként indított napi utazások számát az első hullám során. Figyelembe véve, hogy a kormányzati szigorítások indexének legmagasabb értéke 76,85 volt Magyarországon (13.2. táblázat), a használat maximális *ceteris paribus* növekedése körülbelül 3,9 utazás/állomás/nap volt (a kormányzati korlátozások nélküli esethez képest). Ez körülbelül a fele az állomások átlagos kihasználtságának. Ugyanakkor ez a masszív növekedés nem járt együtt az átlagos utazási időtartamnak a változásával, a 13.4. táblázat szerint 5%-on nem szignifikáns a változó.

A mobilitás szintén jelentősen befolyásolta a Bubi használatát. A városi forgalom egy százalékponttal nagyobb csökkenése minden más változtatlansága mellett átlagosan 0,057 utazás/állomás/nap csökkenést eredményezett a Bubi használatában. Az első hullám korai időszakában a forgalom csak fele akkora volt, mint egy évvel korábban; ez tehát masszív hatást mutat, közel 3 utazás/állomás/nap csökkenést (ami nagyjából 40%-os csökkenésnek felel meg az állomások átlagos kihasználtságához képest). Az első hullám során a mobilitás és a szigorítások ellentétes hatást fejtettek ki; a Bubi-használat teljes változása e kettő nettó hatásának eredménye.

A két hullám közötti időszak eredményei hasonlóak, de az értelmezésnél érdemes odafigyelni. A szigorítások indexének együtthatója nagyon jelentősen megnőtt. Mivel a korlátozásokat ebben az időszakban feloldották, a Bubi-hasz-

nálat a korlátozások megszüntetésével párhuzamosan csökkenni kezdett. Másrészt az adatok az általános mobilitásnak a Bubi-használatra gyakorolt hasonlóan pozitív hatását mutatják, mint az első hullám során. Ezek az eredmények azt jelzik, hogy a Bubi nem volt képes megtartani az új felhasználókat. Ahogy a kormányzati korlátozások enyhültek, az ingázók elfordultak a Bubitól. Ezért nem figyelhető meg egy hosszabb távú pozitív (növekvő) hatás az utazók számát illetően. Ezt tovább erősíti az a tény, hogy sem a szigorítások, sem a mobilitás nem befolyásolta a Bubi utazási időtartamát ebben az időszakban.

A teljes csúcsidőszakra vonatkozó eredmények hasonló következtetésekre engednek következtetni, nevezetesen arra, hogy mind a kormányzati szigorítások, mind a mobilitás pozitív hatással voltak a Bubi-utazások számára. Továbbá, amint a mobilitás növekszik, a Bubival megtett utak átlagos időtartama csökkenést mutat. A kisebb forgalom tehát hosszabb Bubi-utakkal jár együtt, bár a különbség csak néhány percet tesz ki.

Röviden érdemes még kitérni az időjárás változók értelmezésére is. A napi átlaghőmérséklet hatása valóban nem lineáris, a referenciacsoportként használt 15 és 20°C közötti hőmérséklet esetén a legnagyobb a használat, minden más esetben ennél szignifikánsan kisebb, hiszen a változók előjele negatív. Ez alátámasztja a kezdeti feltevésünket, hogy túlságosan hideg és túlságosan meleg időben sem kedvelik annyira az emberek a kerékpározást. A becsült paraméterek nagyságából látható, hogy minél nagyobb az eltérés az ideális 15–20°C-os értéktől, annál jelentősebb a használatra gyakorolt negatív hatás. Szintén összhangban van a várakozásokkal, hogy több csapadék és erősebb szellőkések esetén is szignifikánsan kisebb a Bubi-használat. Egy mm-rel több napi összes csapadék minden más változatlansága mellett például 0,347-tel csökkenti az adott napon a bérlések átlagos számát az egyes gyűjtőállomásokon, egyben 0,16 perccel csökkenti az átlagos utazási időt.

13.3. táblázat. A Bubival megtett napi utazások számát befolyásoló hatások becslési eredményei munkanapokon

Magyarázó változók	A teljes csúcsidőszak 2020-ban	A Covid19-járvány első hulláma	A Covid19-járvány első és második hulláma közötti időszak
	2020. március 1–augusztus 31.	2020. március 1–június 21.	2020. június 22–augusztus 31.
Kormányzati szigorítások indexe	0,060*** (0,005)	0,051*** (0,005)	0,268*** (0,037)
Napi városi forgalom százalékos változása	0,057*** (0,003)	0,057*** (0,005)	0,058*** (0,007)
Napi átlaghőmérséklet: -5–0°C	-7,081*** (0,689)	-6,997*** (0,687)	–
Napi átlaghőmérséklet: 0–5°C	-4,886*** (0,260)	-4,852*** (0,263)	–
Napi átlaghőmérséklet: 5–10°C	-3,334*** (0,223)	-3,301*** (0,219)	–
Napi átlaghőmérséklet: 10–15°C	-1,783*** (0,160)	-1,843*** (0,159)	–
Napi átlaghőmérséklet: 20–25°C	-0,800*** (0,141)	-1,031*** (0,292)	-0,021 (0,201)
Napi átlaghőmérséklet: > 25°C	-1,172*** (0,188)	-3,616*** (0,488)	-0,501** (0,218)
Napi összes csapadék	-0,347*** (0,012)	-0,388*** (0,015)	-0,296*** (0,019)
Napi átlagos szélesebesség	-0,238*** (0,035)	-0,223*** (0,040)	-0,047 (0,073)
N	16 728	9 808	6 920
R ²	0,237	0,292	0,254

Megjegyzések: A hőmérséklet esetén a referenciacsoport: 15–20°C. A regressziók tartalmaztak állomás dummy változókat, ezeket helytakarékossági okokból nem jelenítettük meg. Zárójelben a standard hibák. *p < 0,1; **p < 0,05; ***p < 0,01.

13.4. táblázat. Az átlagos Bubi-utak időtartamát befolyásoló tényezők becslési eredményei munkanapokon

Magyarázó változók	A teljes csúcsidezőszak 2020-ban	A Covid19-járvány első hulláma	A Covid19-járvány első és második hulláma közötti időszak
	2020. március 1–augusztus 31.	2020. március 1–június 21.	2020. június 22–augusztus 31.
Kormányzati szigorítások indexe	-0,003 (0,014)	-0,018 (0,016)	0,091 (0,106)
Napi városi forgalom százalékos változása	-0,037*** (0,008)	-0,035** (0,015)	-0,009 (0,021)
Napi átlaghőmérséklet: -5–0°C	-4,243** (1,960)	-4,383** (2,039)	–
Napi átlaghőmérséklet: 0–5°C	-3,683*** (0,738)	-3,775*** (0,782)	–
Napi átlaghőmérséklet: 5–10°C	-2,681*** (0,633)	-2,710*** (0,650)	–
Napi átlaghőmérséklet: 10–15°C	-1,020** (0,455)	-1,170** (0,472)	–
Napi átlaghőmérséklet: 20–25°C	-0,871** (0,402)	-1,850** (0,868)	0,412 (0,572)
Napi átlaghőmérséklet: > 25°C	-1,128** (0,535)	-0,578 (1,448)	-0,285 (0,620)
Napi összes csapadék	-0,160*** (0,034)	-0,217*** (0,045)	-0,099* (0,053)
Napi átlagos szélesség	-0,116 (0,100)	-0,086 (0,118)	0,040 (0,209)
N	16 728	9 808	6 920
R ²	0,027	0,032	0,045

Megjegyzések: A hőmérséklet esetén a referenciacsoport: 15–20°C. A regressziók tartalmaztak állomás dummy változókat, ezeket helytakarékossági okokból nem jelenítettük meg. Zárójelben a standard hibák. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$.

A területhasználat hatását az állomásszintű időbeli változások, tehát az állomás dummy változók elemzésével vizsgáljuk (13.5. táblázat). Fontos megjegyezni, hogy ezek más tényezőket is tartalmazhatnak, nem csak a területhasználatot. Megállapítható azonban, hogy a járvány első hulláma során a lakóterületeken és

a parkokban nőtt az utazások száma, míg a kereskedelmi területeken, az egyetem mellett, a szórakozóhelyeken és a közlekedési csomópontokban csökkent. A Bubi-használat a kórházakhoz közeli állomásokon is csökkent. Ez azzal magyarázható, hogy a kórházakban látogatási tilalom volt érvényben a Covid19-járvány idején, valamint számos nem sürgősségi műtétet elhalasztottak, mivel a kórházak a Covid19-járványhoz kapcsolódó esetekre összpontosítottak. A kórházba került betegek pedig nem kerékpárral mentek be az intézményekbe.

A világjárvány két hulláma között a közlekedési csomópontokban található gyűjtőállomásokon magasabb, a parkokban pedig alacsonyabb volt a Bubi-bérlések száma. Ez egy újabb jele annak, hogy az élet kezdett normalizálódni, és az ingázók ismét kombinálták a Bubit más közlekedési módokkal.

13.5. táblázat. Az egyes állomásokról indított utazások számának átlagos változása állomásszinten 2019 és 2020 között területhasználati típusok szerint (munkanapokon)

Területhasználat típusa	Állomások száma	Átlagos állomásszintű változások 2019 azonos időszakához képest		
		A teljes csúcsidezőszak 2020-ban	A Covid19-járvány első hulláma	A Covid19-járvány első és második hulláma közötti időszak
Lakóövezet	46	1,46	1,46	1,55
Kereskedelmi övezet	36	-0,47	-0,59	-0,44
Oktatási övezet	19	-1,80	-1,60	-1,34
Parkos terület	24	0,66	1,42	-0,49
Közlekedési csomópont	17	-0,59	-1,63	0,85
Egészségügyi övezet	4	-1,16	-0,45	-2,15
Szórakozási övezet	14	-1,01	-0,54	-1,78

Megjegyzések: A negatív számok relatív csökkenést, míg a pozitív számok relatív növekedést jeleznek az utazások számában.

Következtetések és közpolitikai javaslatok

A közösségi kerékpárrendszerek a városokon belüli közlekedés tiszta, megfizethető és egészséges alternatívái. Az ingázók ösztönzése a Bubi használatára hozzájárulhat az egészségük javulásához és a környezetszennyezés csökkentéséhez is. A Covid19-járvány jelentősen megnövelte a Bubi iránti érdeklődést. Egyrészt a Bubi biztonságos utazási lehetőség, amivel az közlekedők minimalizálni tudják

a másokkal való érintkezést. Másrészt olcsó és sokkal szélesebb közönség számára elérhető, mint a saját tulajdonú gépjármű. Budapest esetében ráadásul a Bubi-bérllet árát jelentősen csökkentették, ami lényegében a használat esetleges pénzügyi akadályait is megszüntette.

Eredményeink azt mutatják, hogy a Bubi használata valóban jelentősen megnőtt a vilávjárvány első hullámában Budapesten. A mobilitás munkanapokon elsősorban a munkába járáshoz kapcsolódik. Eredményeink alapján Budapesten is elképzelhető, hogy a közösségi közlekedést többen kerékpárra cserélték, mivel a Bubi használata jelentősen megnőtt, miközben az általános mobilitás csökkent.

Továbbá a Bubi-használaton belül is megfigyelhető volt egy elmozdulás. A lakóövezetekben vagy parkok közelében található állomások felhasználói száma nőtt, míg az egyetemekhez, kórházakhoz, szórakozóhelyekhez és kereskedelmi területekhez közeli állomások felhasználói száma csökkent. Ez vélhetően a korlátozások és az otthoni munkavégzés elterjedésének eredménye, de arra is utal, hogy a Bubit munkanapokon is használták szabad időszelokra is.

A vilávjárvány első és második hulláma között a Bubi kihasználtsága lassan visszatért a 2019-es szintre, bár még mindig magasabb volt, mint egy évvel azelőtt. Az autóforgalom ebben az időszakban visszatért a pandémia előtti szintre, és a megnövekedett forgalom a Bubi-használat növekedésével járt együtt. A kormányzati korlátozások feloldása azonban a Bubi-használat csökkenéséhez vezetett. Az eredmények azt mutatják, hogy amint az élet normalizálódott, a Bubi-használat a pandémia előtti szint felé tendált, bár a viteldíjakat nem emelték, amelyek tehát jelentősen a 2019-es árak alatt maradtak. Mindezek az eredmények azt mutatják, hogy a Bubi sajnos nem volt képes sok hosszú távú felhasználót bevonítani.

A vilávjárvány összességében egy természetes kísérletnek tekinthető, amely a Bubi használatának jelentős növekedéséhez vezetett. Számos ingázó legalább egyszer kipróbálta a Bubit. Sajnálatos módon ezen új felhasználók jelentős része nem vált hosszú távú felhasználóvá, hiszen a korlátozások megszüntetése negatívan befolyásolta a Bubi használatát, vagyis az ingázók jelentős része visszaváltott a Bubiról más közlekedési módokra. Ez egy általánosabb problémáról árulkozik, miszerint a Bubi nem vált az emberek mindennapi életének és utazási szokásainak részévé. Hosszú távon ez megkérdőjelezi e rendszerek életképességét, és csökkenti azokat a pozitív egészségügyi és környezeti előnyöket, amelyeket egy ilyen közösségi kerékpárrendszer nyújthat.

Az eredmények négy fontos szakpolitikai következményre világítanak rá. Először is, a Bubi (és más közösségi kerékpárrendszerek is) értékes utazási lehetőséget jelenthetnek egy vilávjárvány idején, amelyet valószínűleg számos ember fog igénybe venni. Mivel a jövőben gyakrabban fordulhatnak elő hasonló vilávjár-

ványok, a kapcsolódó infrastruktúra fejlesztése szükséges. Mivel a pandémiák a kereslet időbeli növekedéséhez vezethetnek, a rendszerüzemeltetőknek a közösségi kerékpárrendszerek rugalmasságának növelésére kell összpontosítaniuk. Ezt például úgy lehet elérni, hogy ideiglenes állomásokat állítanak fel, vagy lehetőség szerint bővítik a kerékpárok számát. Ezt a szakpolitikai ajánlást erősíti az a tény, hogy a motorizált járművekkel ellentétben a közösségi kerékpárrendszerek használata nem vezet további üvegházhatásúgáz-kibocsátáshoz; ezért a járvány terjedésének lassítása érdekében sokkal környezetbarátabb ezekre támaszkodni, mint a magán- vagy közösségi autókat előnyben részesíteni.

Másodszor, a lakóövezetekben és parkokban található gyűjtőállomások kihasználtsága sokkal nagyobb volt a világvárvány idején, mint a város más területein található gyűjtőállomásoké. Az előbb javasolt ideiglenes állomásokat tehát elsősorban ezeken a helyeken kell elhelyezni. Ugyanakkor a szórakoztató létesítmények, kereskedelmi területek, közlekedési csomópontok és egyetemek közelében található állomásokon a használat relatív csökkenése volt tapasztalható. A rendszernek tehát elég rugalmasnak kell lennie ahhoz, hogy ezt viszonylag rövid idő alatt kezelni tudja.

Harmadszor, fontos megjegyezni, hogy az árak az egész időszak alatt nagyon alacsonyak maradtak (100 Ft/hónap), ezért pénzügyi korlátok valószínűleg nem játszottak jelentős szerepet abban, hogy az emberek igénybe vették-e a Bubit. Ez megkérdőjelezi azt a nézetet, hogy az árak jelentős szerepet játszanak a közösségi kerékpárrendszerek használatában. El kell ismerni, hogy a túl magas árak elriaszthatják az utasokat, de az ár nullához közeli szintre csökkentése valószínűleg nem növeli a hosszú távú használók számát, különösen akkor nem, ha a rendszer nem nyújt kielégítő felhasználói élményt. Ez hasznos felismerés lehet az árképzési stratégia megtervezésekor.

Negyedszer, a világvárvány vagy hasonló természetes kísérletek rávilágíthatnak az ilyen rendszerek gyengeségeire is. A Bubi-használat és a jelenlegi rendszerekkel való elégedettség meghatározó tényezőinek megértése alapvető fontosságú egy jól működő rendszer megtervezéséhez. A felhasználók véleményének összegyűjtésére gyakran végeznek felméréseket, de nehéz begyűjteni azok véleményét, akik nem használják a szolgáltatást. Az ő bevonásuk azonban a rendszer potenciális növekedését jelenti. Egy olyan természetes kísérlet, mint a Covid19-járvány, a potenciális felhasználók sokkal bővebb körének preferenciáira és motivációira világíthat rá, mivel számos olyan ingázó is kipróbálta a szolgáltatást, aki egyébként nem tette volna. Az ő viselkedésük értékes információkkal szolgálhat a rendszer növekedési potenciálját illetően. A Bubi esetében a világvárvány megerősítette, hogy a rendszer növekedési potenciálja korlátozott, mivel az új felhasználók többsége nem vált hosszú távú felhasználóvá.

A Bubi üzemeltetője, a BKK a beérkezett panaszok és a saját maga által végzett fogyasztói felmérések alapján már tisztában volt azokkal a problémákkal, amelyek nagy valószínűséggel a jövőbeni fejlődés akadályai, nevezetesen, hogy a kerékpárokat nehéz tekerni, mivel a kerékpárvázak önsúlya nagy, a gumik pedig tömörek (nem felfújtak). A világvárvány világossá tette, hogy ezeket a problémákat meg kell oldani. Ennek következtében a Bubit végül 2020 végén ideiglenesen leállították Budapesten, és 2021 májusában egy új rendszer, a Bubi 2.0 indult el új, jobb minőségű és könnyebb kerékpárokkal és egy egyszerűbben kezelhető és felhasználóbarát applikációval. Mindez elengedhetetlen egy tisztább városi jövő elérése érdekében.

FELHASZNÁLT IRODALOM

- Backhaus, K., Erichson, B., Gensler, S., Weiber, R., Weiber, Th. (2018): *Multi-variate Analysemethoden – eine anwendungsorientierte Einführung*, 17. Auflage, Springer Gabler, Wiesbaden
- Békés, G., Kézdi, G. (2021): *Data Analysis for Business, Economics, and Policy*, Cambridge University Press, Cambridge
- Békés, G., Kézdi, G. (2024): *Adatelemzés*, Alinea Kiadó, Budapest
- Bryan O. (2002), Sawtooth Software, Inc.: *CVA User Manual*
- Gyulavári, T., Mitev, A. Z., Neulinger, Á., Neumann-Bódi, E., Simon, J., Szűcs, K., (2014): *A marketingkutatás alapjai*, Akadémiai Kiadó, Budapest
- Hajdú, O. (2005): *A diszkrét kiválasztási modell becslése Cox regresszióval*, Szigma, 36. évf. 3-4. sz. 113-128. o.
- Hunyadi, L. Mundruczó, K., Vita, L. (1996): *Statisztika*, Aula Kiadó, Budapest
- Kemény, I., Simon, J., Berezvai, Z., Kun, Zs. (2021): *Marketingkutatás kvantitatív módszerei – segédanyag SPSS program használatához*, Budapesti Corvinus Egyetem, Budapest
- Kenesei, Zs., Kolos, K. (2014): *Szolgáltatásmarketing és -menedzsment*, Alinea Kiadó, Budapest
- Ketskemény, L., Izsó, L., Könyves Tóth, E. (2011): *Bevezetés az IBM SPSS Statistics programrendszerbe*, Artéria Stúdió, Budapest
- Kuhfeld, W. F. (1997), SAS Institute, Inc.: *Efficient Experimental Designs Using Computerized Searches*
- Malhotra, N. K., Simon, J. (k.m.) (2009): *Marketingkutatás*, Akadémiai Kiadó, Budapest
- Sajtos, L., Mitev, A. (2007): *SPSS kutatási és adatelemzési kézikönyv*, Alinea Kiadó, Budapest
- Simon, J. (2006): *A klaszterelemzés alkalmazási lehetőségei a marketingkutatásban*, Statisztikai Szemle, 84. évf. 7. sz. 627-650. o.
- Szűcs, K., Lázár, E., Németh, P. (2023): *Marketingkutatás 4.0*, Akadémiai Kiadó, Budapest
- Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press, Cambridge
- Wooldridge, J. M. (2020): *Introductory Econometrics. A Modern Approach*, 7th ed. Cengage Learning, Mason

MELLÉKLET – A KÉRDŐÍV

A többszatsornás értékesítés és fenntarthatóság kapcsolatára vonatkozó fogyasztói magatartásról

Bemutatkozás – a kutatás célja röviden

Figyelemfelhívás az önkéntes kitöltésre

Szűrő kérdés

1. Az elmúlt egy évben vásároltál ruházati (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába), kozmetikai vagy szórakoztató elektronikai (pl.: mobiltelefon, kamera, e-book stb.) termékeket?
 - a) Igen
 - b) Nem → kérdőív vége

Általános vásárlási szokások 5 termékkategóriát tekintve: Ruházat, kozmetikumok, szórakoztató elektronika, könyv és ruházati kiegészítők

Q01 – Mely terméket/termékeket vásároltad az alábbiak közül az elmúlt egy évben? Több választ is megjelölhetsz!

6. Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)
7. Kozmetikum
8. Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book stb.)
9. Könyv
10. Ruházati kiegészítő (cipő, táska és egyéb kiegészítők)

Q02 – Hol szoktad vásárolni az alábbi termékkategóriákat (elmúlt egy évben)? Több válasz is lehetséges!

Q2-1 Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)

8. Hipermarketben (pl. Tesco, Auchan)
9. Szupermarketben (pl. Spar)
10. Diszkontban (pl. Aldi, Lidl)
11. Fizikai szaküzletben (pl. ruhabolt, könyvesbolt, drogéria)
12. Általános webáruházban (pl. EMAG, Alza)
13. Online szakáruház (pl. online könyvesbolt/ruhabolt/drogéria)
14. Egyéb helyen:

Q02_2 Kozmetikum

1. Hipermarketben (pl. Tesco, Auchan)
2. Szupermarketben (pl. Spar)

3. Diszkontban (pl. Aldi, Lidl)
4. Fizikai szaküzletben (pl. ruhabolt, könyvesbolt, drogéria)
5. Általános webáruházban (pl. EMAG, Alza)
6. Online szakáruház (pl. online könyvesbolt/ruhabolt/drogéria)
7. Egyéb helyen:

Q02-3 Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book)

1. Hipermarketben (pl. Tesco, Auchan)
2. Szupermarketben (pl. Spar)
3. Diszkontban (pl. Aldi, Lidl)
4. Fizikai szaküzletben (pl. ruhabolt, könyvesbolt, drogéria)
5. Általános webáruházban (pl. EMAG, Alza)
6. Online szakáruház (pl. online könyvesbolt/ruhabolt/drogéria)
7. Egyéb helyen:

Q02-4 Könyv

1. Hipermarketben (pl. Tesco, Auchan)
2. Szupermarketben (pl. Spar)
3. Diszkontban (pl. Aldi, Lidl)
4. Fizikai szaküzletben (pl. ruhabolt, könyvesbolt, drogéria)
5. Általános webáruházban (pl. EMAG, Alza)
6. Online szakáruház (pl. online könyvesbolt/ruhabolt/drogéria)
7. Egyéb helyen:

Q02-5 Ruházati kiegészítő (cipő, táska és egyéb kiegészítők)

1. Hipermarketben (pl. Tesco, Auchan)
2. Szupermarketben (pl. Spar)
3. Diszkontban (pl. Aldi, Lidl)
4. Fizikai szaküzletben (pl. ruhabolt, könyvesbolt, drogéria)
5. Általános webáruházban (pl. EMAG, Alza)
6. Online szakáruház (pl. online könyvesbolt/ruhabolt/drogéria)
7. Egyéb helyen:

Q03 – Válassz ki egyet az alábbi kategóriák közül, amelyet a legutóbb vásároltál! (A további kérdések erre a kategóriára fognak vonatkozni.)

1. Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)
2. Kozmetikum
3. Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book stb.)
4. Könyv
5. Ruházati kiegészítő (cipő, táska és egyéb kiegészítők)

Q04 – Gondolj az előző kérdésben megnevezett kategóriára! Mennyire jellemzőek rád a következő tevékenységek az előző kérdésben megjelölt kategória vásárlása során? Kérlek értékeld 1-től 7-ig skálán, ahol az 1 jelenti, hogy egyáltalán nem jellemző, a 7 pedig, hogy teljes mértékben jellemző.

5. Ha az interneten keresek, legtöbbször meg is veszem a kiválasztott terméket valamelyik webáruházban.
6. Sokszor az interneten keresem meg a terméket, amit meg akarok venni, de utána mindig elmegyek egy boltba és ott vásárlom meg.
7. Ha a boltban keresek, majd meglátok egy nekem tetsző terméket, ott helyben meg is veszem.
8. Ha a boltban keresek, majd meglátok egy nekem tetsző terméket, többnyire megkeresem az interneten és online veszem meg.

Okos vásárlónak lenni

Q05 – Mennyire törekszel a következőkre az általad megjelölt kategória vásárlása során? Értékeld 1-től 7-ig skálán, ahol az 1 jelenti, hogy egyáltalán nem törekszem, a 7 pedig, hogy teljes mértékben erre törekszem.

1. Vásárlásaimat a lehető legkényelmesebben intézzem.
2. Vásárlásaim ne okozzanak túl sok gondot, nyűgöt számomra.
3. A vásárlás érdekében ne kelljen extra erőfeszítéseket tennem.
4. A vásárlásaimat képes legyek minél gyorsabban megvalósítani.
5. Azokat a termékeket tudjam megvenni, amikre igazán vágyom.
6. A vásárlásaim tökéletesen illeszkedjenek az igényeimhez.
7. Jó minőségű termékeket tudjak megvásárolni.
8. Olcsóbban tudjam a termékeket megvásárolni, mint máshol
9. Jó áron tudjam a termékeket megvenni.
10. A vásárlással jól járjak.
11. A vásárlás után jól érzem magam.
12. A vásárlás után úgy érzem, hogy okos döntést hoztam.
13. Büszke legyek magamra, hogy jól vásároltam.

Q06 – Milyen gyakran szoktad az adott termék kategóriát online vásárolni (10 alkalomból hányszor)?

1. Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)
2. Kozmetikum
3. Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book stb.)
4. Könyv
5. Ruházati kiegészítő (cipő, táska és egyéb kiegészítők)

Q07 – Melyik webáruházban vásároltál már az alábbiak közül a fentebb megnevezett termékkategóriák esetében? Több válasz is adható

1. Alza
2. eMAG
3. Euronics
4. Media Markt webáruház
5. Mall.hu
6. Amazon
7. dm.hu
8. shop.rossmann.hu
9. CCC webshop/mobilapplikáció
10. ecipo
11. Deichmann webshop/applikáció
12. Egyéb, éspedig:.....

Q07_egyeb

Egyéb, éspedig:.....

Q08 – Hány alkalommal vásároltál online az elmúlt 1 hónap során az alábbi termékkategóriában?

Q10_1 Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)

Q10_2 Kozmetikum

Q10_3 Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book stb.)

Q10_4 Könyv

Q10_5 Ruházati kiegészítő (cipő, táska és egyéb kiegészítők)

Q09 – Az alábbi termékkategóriák közül, melyiket vásároltad online a legutóbbi alkalommal?

1. Ruházati termék (cipő, táska és kiegészítő nem tartozik ebbe a kategóriába)
2. Kozmetikum
3. Szórakoztató elektronika (pl.: mobiltelefon, kamera, e-book stb.)
4. Könyv
5. Ruházati kiegészítő (cipő, táska és egyéb kiegészítők)

Q10 – Mennyire jellemzőek rád az alábbi állítások, amikor az általad választott termékkategóriában ONLINE vásárolsz? Az előző kérdéskor választott kategória legutóbbi online vásárlására gondolj! Kérlek értékeld 1-től 7-ig skálán, ahol az 1 jelenti, hogy egyáltalán nem jellemző, a 7 pedig, hogy teljes mértékben jellemző.

1. A saját lakcímemre rendelem a termékeket.
2. A munkahelyem címére rendelem a termékeket.
3. Egyéb más címre rendelem a termékeket.

4. Csomagpontra kérem a termékeket.
5. Csomagautomatába kérem a termékeket.
6. Bolti átvételre kérem a termékeket

Q11 – Amennyiben nem kiválasztott címre kéred a rendelt terméket, legfeljebb mennyit vagy hajlandó utazni a vásárolt csomagért? Kérlek a választ egész számra kerekítve percben számítva add meg (pl.: 30 percet, 60 percet, 70 percet stb.).

Saját válasz

Q12 – Mennyire jellemzőek rád az alábbi állítások, amikor az általad választott termék kategóriában ONLINE vásárolsz? Kérlek értékeld 1-től 7-ig skálán, ahol az 1 jelenti, hogy egyáltalán nem jellemző, a 7 pedig, hogy teljes mértékben jellemző.

1. A kiszállítás módjának kiválasztásánál mindig az olcsóbb alternatívát választom.
2. A kiszállítás módjának kiválasztásánál mindig a leggyorsabb alternatívát választom.
3. A kiszállítás módjának kiválasztásánál mindig a legfenntarthatóbb (pl.: kevesebb legyen a szállítás következtében termelt károsanyag kibocsátás, kevesebb csomagolóanyagot használjanak, újrahasznosított csomagolóanyagot használjanak stb.) alternatívát választom.
4. A kiszállítás módjának kiválasztásánál mindig a kilométerben legközelebbi alternatívát választom.
5. A házhoz (vagy választott címre való) szállítás kiszámíthatatlansága miatt választok más alternatívákat.
6. Az aznapi vagy legfeljebb a másnapi kiszállítást preferálok.
7. Az ingyenes kiszállítási opciókat preferálok.
8. Pénzspórolás céljából inkább a boltba vagy csomagpontra kérem a megrendelt terméket.
9. Hajlandó vagyok tovább várni a termék kiszállítására, ha cserébe ingyenes a termék kiszállítása.
10. Szeretnék több információt kapni az online vásárolt termékek fenntartható (pl.: alacsonyabb károsanyag kibocsátás, elektromos járművel való kiszállítás) kiszállításáról.
11. Szeretnék több információt kapni az online vásárolt termékek fenntartható (pl.: kevesebb csomagolóanyag, újrahasznosított csomagolás stb.) csomagolásáról.
12. Szeretném csökkenteni a külföldről rendelt termékek mennyiségét.

13. Találkoztam már olyan online felülettel, amely biztosította a fenntartható (pl.: alacsonyabb károsanyag kibocsátás, kevesebb csomagolóanyag) kiszállítást.
14. Hajlandó lennék többet fizetni a fenntartható termékátvételért (pl.: szállítás elektromos járművel, teherbírásos biciklivel, tömegközlekedéssel beutazni az átvevő pontra stb.).
15. Hajlandó lennék többet fizetni a fenntartható kiszállítási csomagolásért (pl.: kevesebb csomagolás, környezetbarát csomagolóanyag, újrahasznosított csomagolóanyag stb.).
16. Hajlandó lennék többet fizetni a termék kiszállításáért/átvételéért, hogy ha a termék nem külföldről jön.

Visszaküldési preferencia

Q13 – Mennyire jellemzőek rád az alábbi állítások, amikor az általad választott termék kategóriában **ONLINE** vásárolsz? Kérlek értékeld 1-től 7-ig skálán, ahol az 1 jelenti, hogy egyáltalán nem jellemző, a 7 pedig, hogy teljes mértékben jellemző.

1. Számomra nagyon fontos a termék visszaküldési lehetősége abban az esetben, ha termék nem megfelelő a számomra.
2. Inkább a boltban veszem át a terméket, mert tudom, hogy ott biztosított a termék azonnali és ingyenes visszaküldése.
3. Hajlandó vagyok többet fizetni a termék kiszállításáért, ha cserébe fennáll az ingyenes visszaküldés lehetősége.
4. Hajlandó vagyok tovább várni a termék kiszállítására, ha cserébe fennáll az ingyenes visszaküldés lehetősége.
5. Általában egyáltalán nem érdekel, hogy milyen visszaküldési lehetőségek állnak a rendelkezésemre.

Fenntartható/környezettudatos omnichannel viselkedés

Q14 – Mennyire jellemzőek rád az alábbi állítások, amikor az általad választott termék kategóriában vásárolsz? Kérlek értékeld 1-től 7-ig skálán, ahol az 1 jelenti, hogy egyáltalán nem jellemző, a 7 pedig, hogy teljes mértékben jellemző.

1. Amennyire lehetséges, figyelembe veszem a környezeti hatásokat (pl.: kevesebb csomagolás, kevesebb károsanyag kibocsátás stb.) a hagyományos (bolti) vásárlásom során.
2. Amennyire lehetséges, figyelembe veszem a környezeti hatásokat (pl.: kevesebb csomagolás, kevesebb károsanyag kibocsátás stb.) az online vásárlásom során.
3. Számomra fontos, hogy a futárok teherautóikkal kevesebb kilométert tegyenek meg a csomagjaim kiszállítása közben.

4. Fontos számomra, hogy a futárok teherautóikkal a lakóhelyem környezetében kevesebb kilométert tegyenek meg a csomagjaim kiszállítása közben.
5. Csomagomat kész vagyok csomagpontban átvenni, ha a futároknak kevesebb kilométert megtennie ennek köszönhetően.
6. Hajlandó vagyok tovább várni a csomagjaimmal, ha a futárok így kevesebb kilométert tesznek meg.
7. Számomra fontos, hogy az általam egyszerre megrendelt termékeket egyben szállítsák ki.
8. Együttes kézbesítés esetén hajlandó vagyok tovább várni a csomagjaimmal.
9. Hajlandó vagyok többet fizetni a fenntartható elektromos járművel történő kiszállításért.
10. Hajlandó vagyok többet fizetni a fenntartható teherbiciklivel való kiszállításért.

Q15 – Hány forint méltányos véleményed szerint egy átlagos méretű csomag (max 5 kg) kiszállítási díjaként? (Ft)

Q16 – Amennyiben egy webshop ingyenes kiszállítást kínál egy rendelési összeghatár felett, úgy mekkora összeghatár az, amelyért még akár plusz tételeket is teszel a kosaradba az ingyenes szállítási lehetőségért? (Ft)

CON-JOINT kártyák Q17_01-től Q17_18-ig

	A opció	B opció
Ár	Ingyenes	Ingyenes hűségprogrammal
Szállítás teljesítésének ideje	Legalább 5 nap, de választott szállítási dátum	Legalább 3 nap, de választott szállítási dátum
Szállítás módja	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	Választott cím, hétköznapokon (9:00 és 18:00 között)
Visszaküldési lehetőség	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	Ingyenesen csomagküldőszekrényben (0-24 órában)

	A opció	B opció
Ár	Ingyenes kiszállítás 15.000 forint felett	Ingyenes
Szállítás teljesítésének ideje	Másnap	Legalább 3 nap, de választott szállítási dátum
Szállítás módja	Csomagautomata (0-24 órában)	Választott cím, hétköznapokon (9:00 és 18:00 között)
Visszaküldési lehetőség	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	990 Ft csomagküldőszekrényben (0-24)

	A opció	B opció
Ár	Ingyenes hűségprogrammal	Ingyenes kiszállítás 15.000 forint felett
Szállítás teljesítésének ideje	Legalább 3 nap, de választott szállítási dátum	Másnap
Szállítás módja	Csomagautomata (0-24 órában)	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)
Visszaküldési lehetőség	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	990 Ft csomagküldőszekrényben (0-24)

	A opció	B opció
Ár	Ingyenes hűségprogrammal	Ingyenes kiszállítás 15.000 forint felett
Szállítás teljesítésének ideje	Másnap	Harmadnap
Szállítás módja	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	Választott cím, kétórás időintervallumban
Visszaküldési lehetőség	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)

	A opció	B opció
Ár	990 forint	Ingyenes
Szállítás teljesítésének ideje	Legalább 5 nap, de választott szállítási dátum	Legalább 3 nap, de választott szállítási dátum
Szállítás módja	Választott cím, kétórás időintervallumban	Csomagautomata (0-24 órában)
Visszaküldési lehetőség	990 Ft csomagküldőszekrényben (0-24)	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)

	A opció	B opció
Ár	990 forint	Ingyenes kiszállítás 15.000 forint felett
Szállítás teljesítésének ideje	Harmadnap	Másnap
Szállítás módja	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	Csomagautomata (0-24 órában)
Visszaküldési lehetőség	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	Ingyenesen csomagküldőszekrényben (0-24 órában)

	A opció	B opció
Ár	Ingyenes kiszállítás 15.000 forint felett	Ingyenes hűségprogrammal
Szállítás teljesítésének ideje	Legalább 5 nap, de választott szállítási dátum	Harmadnap
Szállítás módja	Választott cím, hétköznapokon (9:00 és 18:00 között)	Csomagautomata (0-24 órában)
Visszaküldési lehetőség	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	990 Ft csomagküldőszekrényben (0-24)

	A opció	B opció
Ár	Ingyenes hűségprogrammal	990 forint
Szállítás teljesítésének ideje	Másnap	Legalább 5 nap, de választott szállítási dátum
Szállítás módja	Választott cím, kétórás időintervallumban	Csomagautomata (0-24 órában)
Visszaküldési lehetőség	Ingyenesen csomagküldőszekrényben (0-24 órában)	990 Ft csomagküldőszekrényben (0-24)

	A opció	B opció
Ár	990 forint	Ingyenes
Szállítás teljesítésének ideje	Harmadnap	Másnap
Szállítás módja	Választott cím, hétköznapokon (9:00 és 18:00 között)	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)
Visszaküldési lehetőség	Ingyenesen csomagküldőszekrényben (0-24 órában)	990 Ft csomagküldőszekrényben (0-24)

	A opció	B opció
Ár	Ingyenes kiszállítás 15.000 forint felett	990 forint
Szállítás teljesítésének ideje	Legalább 3 nap, de választott szállítási dátum	Másnap
Szállítás módja	Választott cím, kétórás időintervallumban	Választott cím, hétköznapokon (9:00 és 18:00 között)
Visszaküldési lehetőség	990 Ft csomagküldőszekrényben (0-24)	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)

	A opció	B opció
Ár	Ingyenes kiszállítás 15.000 forint felett	Ingyenes
Szállítás teljesítésének ideje	Legalább 3 nap, de választott szállítási dátum	Legalább 5 nap, de választott szállítási dátum
Szállítás módja	Választott cím, hétköznapokon (9:00 és 18:00 között)	Választott cím, kétórás időintervallumban
Visszaküldési lehetőség	990 Ft csomagküldőszekrényben (0-24)	Ingyenesen csomagküldőszekrényben (0-24 órában)

	A opció	B opció
Ár	Ingyenes	Ingyenes hűségprogrammal
Szállítás teljesítésének ideje	Legalább 3 nap, de választott szállítási dátum	Legalább 5 nap, de választott szállítási dátum
Szállítás módja	Választott cím, kétórás időintervallumban	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)
Visszaküldési lehetőség	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)	Ingyenesen csomagküldőszekrényben (0-24 órában)

	A opció	B opció
Ár	Ingyenes	Ingyenes hűségprogrammal
Szállítás teljesítésének ideje	Harmadnap	Legalább 3 nap, de választott szállítási dátum
Szállítás módja	Csomagautomata (0-24 órában)	Választott cím, kétórás időintervallumban
Visszaküldési lehetőség	Ingyenesen csomagküldőszekrényben (0-24 órában)	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)

	A opció	B opció
Ár	990 forint	Ingyenes kiszállítás 15.000 forint felett
Szállítás teljesítésének ideje	Legalább 3 nap, de választott szállítási dátum	Harmadnap
Szállítás módja	Csomagautomata (0-24 órában)	Választott cím, hétköznapokon (9:00 és 18:00 között)
Visszaküldési lehetőség	Ingyenesen csomagküldőszekrényben (0-24 órában)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)

	A opció	B opció
Ár	Ingyenes hűségprogrammal	990 forint
Szállítás teljesítésének ideje	Harmadnap	Másnap
Szállítás módja	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)	Választott cím, kétórás időintervallumban
Visszaküldési lehetőség	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)

	A opció	B opció
Ár	Ingyenes hűségprogrammal	990 forint
Szállítás teljesítésének ideje	Legalább 5 nap, de választott szállítási dátum	Legalább 3 nap, de választott szállítási dátum
Szállítás módja	Csomagautomata (0-24 órában)	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)
Visszaküldési lehetőség	990 Ft csomagküldőszekrényben (0-24 órában)	Ingyenesen csomagküldőszekrényben (0-24 órában)

	A opció	B opció
Ár	990 forint	Ingyenes
Szállítás teljesítésének ideje	Harmadnap	Legalább 5 nap, de választott szállítási dátum
Szállítás módja	Választott cím, kétórás időintervallumban	Választott cím, hétköznapokon (9:00 és 18:00 között)
Visszaküldési lehetőség	Ingyenes hűségprogrammal, a kereskedelmi lánc üzletében (nyitvatartási időben)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)

	A opció	B opció
Ár	Ingyenes	Ingyenes kiszállítás 15.000 forint felett
Szállítás teljesítésének ideje	Másnap	Legalább 5 nap, de választott szállítási dátum
Szállítás módja	Választott cím, hétköznapokon (9:00 és 18:00 között)	Kereskedelmi lánc üzlete vagy csomagpont (nyitvatartási időben)
Visszaküldési lehetőség	990 Ft csomagküldőszekrényben (0-24)	Ingyenesen, a kereskedelmi lánc üzletében vagy csomagponton (nyitvatartási időben)

Fenntartható/Környezettudatos vásárlás

Q18 – Mennyire jellemzők rád az alábbi állítások, amikor vásárolsz? Kérlek értékelj 1-től 7-ig skálán, ahol az 1 jelenti, hogy egyáltalán nem jellemző, a 7 pedig, hogy teljes mértékben jellemző.

1. Minden esetben megnézem a termékek előállításának körülményeit.
2. Kozmetikumok esetén, minden esetben megnézem a termékek összetételét.
3. Nem vásárolok fast fashion termékeket.
4. Csomagolásmentesen vásárolok.
5. Nem vásárolok olyan kozmetikumokat, amelyeket állatokon teszteltek.
6. Nem kérek kiszállítást, inkább a boltban veszem át a termékeket, hogy ezzel is csökkentsem a károsanyag kibocsátást.
7. Ha van választási lehetőség, mindig azt a terméket választom, amely a legkisebb mértékben járul hozzá a környezeti ártalmakhoz.

8. Vannak olyan termékek, amelyeket környezetvédelmi okokból lecseréltem.
9. Ha megértem, hogy egyes termékek milyen potenciális károkat okozhatnak a környezetben, akkor nem vásárolok meg ezeket a termékeket.
10. Nem veszek olyan háztartási termékeket, amelyek károsak a környezetre.
11. Lehetőség szerint újrahasznosítható vagy újrahasznosítható edénybe csomagolt termékeket vásárolok.
12. Mindent megteszek, hogy újrahasznosított papírból készült papírtermékeket (vécépapír, zsebkendő stb.) vásárolok.
13. Nem veszek meg egy terméket, ha tudom, hogy az árusító cég társadalmilag felelőtlen.
14. Nem vásárolok termékeket olyan cégektől, amelyekről tudom, hogy munkavégzést, gyermekmunkát vagy más rossz munkakörülményeket használnak.
15. Hajlandó vagyok többet fizetni a környezetbarát termékekért akkor is, ha van olcsóbb alternatíva.
16. Hajlandó vagyok többet fizetni a társadalmilag felelős termékekért akkor is, ha van olcsóbb alternatíva.

Q19 – Találkoztál-e már a vásárlásaid során olyan termékkel, amelyen valamilyen formában feltüntették a kézhez vétel (kiszállítás, bolti átvétel, csomagponti/csomagautomatás átvétel stb.) környezetbarát jellegét?

1. Igen, viszonylag gyakran.
2. Ritkán.
3. Nem, még soha.

Általános fenntartható/környezettudatos magatartás

Q20 – Mennyire jellemzőek rád az alábbi állítások? Kérlek értékelj 1-től 7-ig skálán, ahol az 1 jelenti, hogy egyáltalán nem jellemző, a 7 pedig, hogy teljes mértékben jellemző.

1. Számomra fontosak a fenntarthatósági (pl.: környezetvédelem, társadalmi felelősségvállalás, tisztességes munkakörülmények biztosítása) szempontok.
2. Véleményem szerint az emberek fogyasztási (vásárlási) szokásai szerepet játszanak a fenntarthatósági (pl.: klímaváltozás, felmelegedés tisztességtelen munkakörülmények, elszegényedés) problémák fokozódásában.
3. Szelektíven gyűjtöm a szemetet.
4. Turkálóban vásárolok.
5. Vegetáriánus vagy vegán életmódot folytatok.

6. Igyekszem a közvetlen környezetem (pl.: szülők, család, barátok stb.) körében is népszerűsíteni a fenntarthatósági (pl.: környezetvédelem, társadalmi felelősségvállalás, tisztességes munkakörülmények biztosítása) szempontokat.
7. Energiatakarékosan főzöm/készítem el az ételeimet.
8. Nem dobom ki a maradékot, a következő étkezésnél megeszem.
9. Harapnivalókat, italokat eldobható csomagolásban vásárolok (elvitelre, gyorsétterem, kávé elvitelre stb.).
10. Előfordul, hogy kidobom az élelmiszereket.
11. Rendszeresen eszem tejtermékeket (vaj, sajt, joghurt stb.).
12. Elajándékozom vagy elcserélem a nem kívánt ruhadarabokat, amelyeket már nem hordok.
13. Ahelyett, hogy új ruhadarabot vásárolnék egy különleges alkalomra, inkább kölcsönkérek valamit.
14. Magam készítek ruhadarabokat (pl. varrás, kötés).
15. Kidobom azokat a ruhadarabokat, amiket már nem hordok.

Demográfia

D1 – Nemed?

1. Férfi
2. Nő

D2 – Melyik évben születted?

D3 – Hol van az állandó lakhelyed?

1. Budapesten
2. Nem Budapesten

D4 – Hol laksz jelenleg Budapesten?

1. Albérletben lakom
2. Saját vagy szülő/rokonok tulajdonában levő ingatlanban lakom
3. Kollégiumban lakom
4. Naponta bejárok Budapestre

D4_egyéb

Egyéb, éspedig:.....

D5 – Dolgozol jelenleg egyetemi tanulmányaid mellett?

- 1 Igen, főállásban
- 2 Igen, részmunkaidőben

- 3 Igen, alkalmi munkákat vállalok
- 4 Nem dolgozom

D5_egyéb

Dolgozol jelenleg egyetemi tanulmányaid mellett?

Egyéb, éspedig:

D6_1 – Mekkora összeget költesz átlagosan havonta az alábbi tételekre? – Lakhatás (Ft/hó)

D6_2 – Mekkora összeget költesz átlagosan havonta az alábbi tételekre? – Étkezés, élelmiszerek (Ft/hó)

D6_3 – Mekkora összeget költesz átlagosan havonta az alábbi tételekre? – Utazás (Ft/hó)

D6_4 Mekkora összeget költesz átlagosan havonta az alábbi tételekre? – Szórakozás – (Ft/hó)

A társadalomtudományi, a közgazdaságtani és a gazdálkodástudományi kutatásokban, ezen belül is hangsúlyosan a marketingkutatásban jelentős szerepe van az adatok kvantitatív jellegű elemzésének. A felhasznált adatok gyakran primer megkérdezéssel készült kutatásokból származnak.

A könyv célja a megkérdezéssel készült kutatások adatainak elemzésére leggyakrabban használt kvantitatív vizsgálati módszerek bemutatása, a statisztikai módszerek lépéseinek gyakorlatorientált ismertetése, valamint az egyik leggyakrabban használt statisztikai elemző szoftver, az SPSS használatának részletes, videóanyaggal segített leírása.

A könyv a megkérdezéssel készült vizsgálatok előkészítésével, a kérdőív-szerkesztés módszertani aspektusával, az egy- és kétváltozós elemzési módszerekkel, az összefüggésvizsgálattal, a többváltozós struktúrafeltáró módszerekkel és a preferenciavizsgálatok módszerével, vagyis a conjoint elemzéssel foglalkozik. A módszerek tárgyalását és a szoftver alkalmazásának lépéseit egy, didaktikai és elemzési célú adatbázis segítségével mutatja be, gyakorlati példákon keresztül.

A könyv szerzői a Budapesti Corvinus Egyetem oktatói, akik jelentős tapasztalattal rendelkeznek a marketingkutatás oktatásában, a könyvet a marketingkutatáson túlmenően mind a gazdálkodástudományi, közgazdasági és társadalomtudományi kutatásokban, mind a felsőoktatási programokban való felhasználásra ajánlják.