

Kvantitatív elemzési módszerek

Írta: Balogh Péter

Tartalomjegyzék

1. Bevezetés	2
2. A kvantitatív kutatási módszerek csoportosítása	4
3. A hipotézisvizsgálatok és jelentőségük az elemzésekben	6
<i>3.1. A hipotézis vizsgálat menete</i>	6
3.1.1 A szakmai kérdés megfogalmazása, statisztikai formája.....	6
3.1.2 A szignifikanciaszint meghatározása	6
3.1.3 A próbastatisztika helyes megválasztása, számítása.....	6
3.1.4. A kritikus érték meghatározása, a statisztikai döntés H_0 elfogadásáról.....	7
3.1.5. Szakmai következtetés	7
<i>3.2. Paraméteres próbák</i>	7
3.2.1 Egymintás t-próba	8
3.2.2 Két független mintás t-próba	10
3.2.3 Párosított mintás t-próba	11
3.2.4 Egy és több szempontos varianciaelemzések	12
4. Két és többváltozós lineáris regresszió számítása	17
<i>4.1. Változók közötti összefüggések elemzése</i>	17
5. A logisztikus regresszió számítása	20
<i>5.1. Az alapmodell</i>	20
<i>5.2. Modell építés, eredmények értékelése</i>	22
6. A faktoranalízis számítása	24
7. Klaszterelemzési technikák	35
8. Korrespondancia-analízis	41
9. A diszkriminancia analízis bemutatása	45
Felhasznált irodalom	56
Mellékletek	58

1. Bevezetés

A marketingkutatók eredményeként nem csak információkat kapunk a fogyasztók viselkedéséről illetve a vállalkozások tevékenységeiről, hanem ezek alapján olyan következtetések levonása is lehetséges lesz, amelyet előzetesen – ezen adatok ismeretek nélkül – nem tudtunk volna megalapozni (MALHOTRA, 2010). Napjainkra az információstechnológia elterjedésével – számos statisztikai programcsomag segítségével – rövid idő alatt el tudjuk végezni az összegyűjtött adataink elemzését, felgyorsítva ezzel a kutatási folyamatot. Ugyanakkor a technológia még nem képes helyettesíteni a szakmai tudást, amivel az eredményeinket szakszerűen értelmezni tudjuk.

A kutatási folyamat 4 fő lépése (KOTLER – ARMSTRONG, 2020) közül ebben a könyvben a 3. azaz az **Adatelemzés** rész kerül bemutatásra.

A statisztikai adatelemző módszerek helyes alkalmazásának feltétele a megszerzett információk szakszerű értelmezése, amihez szükséges a marketingkutatói és statisztikai alapfogalmak pontos ismerete is. Az eltérő statisztikai programcsomagok (pl. IBM SPSS) lehetővé teszi számunkra a jelenségek gyors és sokoldalú vizsgálatát, de ha az általunk begyűjtött információk nem fedik le teljesen a valóságot – esetleg tévesek – abban az esetben az ezekből kiszámított eredmények és a levonható következtetések hamisak lesznek (MALHOTRA – SIMON, 2016; KOTLER – KELLER, 2017).

Az információ jellegzetessége szerint beszélhetünk kvantitatív, mennyiségi és kvalitatív, minőségi jellegű kutatásról. A kvantitatív (mennyiségi) módszerekhez a számszerű és számszerűsíthető információk, illetve ezek megszerzésének módszerei tartoznak. A kvantitatív mennyiségi kutatási eljárások során, reprezentatív minta alapján az egész sokaságra általánosítható eredményekhez juthatunk statisztikai elemzések segítségével.

A kvantitatív kutatás az adatok jellege szerint mennyiségi, a minta elemszáma szerint nagy elemű (reprezentatív) és a célja szerint az alapsokaságra általánosítható, számszerűsíthető eredményeket vár el. A kutatási eszköz általában a standardizált kérdőív, az eredmények feldolgozása matematikai, statisztikai eljárások alkalmazásával történik, ugyanakkor a kiszámított eredmények hasznosításával üzleti döntési javaslatokat, alternatívákat állíthatunk fel (TAMUSné, 2009).

Leggyakrabban kérdőíveket használunk ahhoz, hogy egy adatbázist létrehozzunk. A kérdőíves felmérésben mind a három féle elem (feltáró, leíró és ok-okozati) is jelen lehet (SAJTOS – MITEV, 2007).

A kérdőív különböző kérdéseit külön-külön és egymással összevetve is értékelhetjük. Ez azt jelenti, hogy az alkalmazott elemzési technika függ attól, hogy mennyi kérdést (változót) viszonyítunk egymáshoz.

Egy kérdőívben nem lehet – vagy nem érdemes – egyetlen direkt kérdéssel mérni olyan összetett dolgokat, mint a környezeti attitűd, környezetvédelemmel kapcsolatos viselkedés, lakóhelyhez vagy egy bizonyos termékhez való ragaszkodás. A kérdőív kérdéseire kapott válaszokat a marketing mérésnek kezeli, és elfogadhatónak azokat a méréseket tekinti, amelyek érvényesek és megbízhatóak. Megbízhatónak akkor tekinthető egy mérés, ha megismételve ugyanazt az eredményt kapjuk. Ezért megállapíthatjuk, hogy ha egy összetett dolgot nem egyetlen, hanem több kérdéssel mérünk, fokozzuk (*fokozhatjuk*) a mérés

érvényességét (SZÉKELYI – BARNA, 2002). Azonban a több kérdéssel mérő kérdések megbízhatósága nem minden esetben ennyire egyértelmű.

A kvantitatív elemzési módszerek részletes bemutatása a könyv további fejezeteiben kerül ismertetésre. Az elemzési módszerek csoportosítását a 2., a hipotézisvizsgálatokat a 3. a két- és többváltozós regresszió számítását a 4., a diszkriminanciaanalízist az 5., míg a faktor- és klaszteranalízist a 6. és 7. fejezet mutatja be.

2. A kvantitatív kutatási módszerek csoportosítása

Ahogy az a Kvalitatív piackutatási módszerek és a piackutatás legújabb módszerei című e-learning tananyagban részletesen olvasható, „*míg a kvalitatív kutatás eredményei nem számszerűsíthetők, nem lehet belőlük általános következtetéseket levonni, a kvantitatív (mennyiségi) kutatás eredményei számszerűsíthetők, és megfelelő mintavétel esetén általánosíthatók a vizsgált sokaságra.*” A kutatás kezdetekor a szekunder információk jó kiindulópontot jelentenek. De ezek önmagukban nem mindig biztosítják a kutató illetve vállalatvezető számára a szükséges ismereteket. Ezért szükséges a primer adatok összegyűjtése és elemzése is (KOTLER – KELLER, 2017; MALHOTRA, 2010).

A kvalitatív és a kvantitatív elemzési technikák alapjai a „*Kvalitatív piackutatási módszerek és a piackutatás legújabb módszerei*” és a „*Kvantitatív piackutatási módszerek alapjai*” című e-learning tananyagokban részletesen kifejtésre kerültek, ezért itt csak a különböző kvantitatív kutatási módszerek csoportosítását ismertetjük.

A kvantitatív módszerek között megkülönböztethetünk függőségi és kölcsönös függőségi kapcsolatokat (SAJTOS – MITEV, 2007). Az ezek közötti különbséget az jelenti, hogy az egyik esetben meg tudjuk különböztetni a függő és független változókat, a másik esetben pedig nem tudunk ilyen jellegű megállapítást tenni. A **függőségi kapcsolat** esetében fontos az, hogy a függő és független változók milyen skálán mértek, s ez befolyásolja azt, hogy milyen jellegű elemzést tudunk alkalmazni, amelyek a következők lehetnek:

Nem paraméteres próbák:

keresztábraelemzés Khi-négyzet próbával, Fisher-féle egzakt teszttel,
Mann-Whitney próba,
Kruskal-Wallis próba,

Paraméteres próbák:

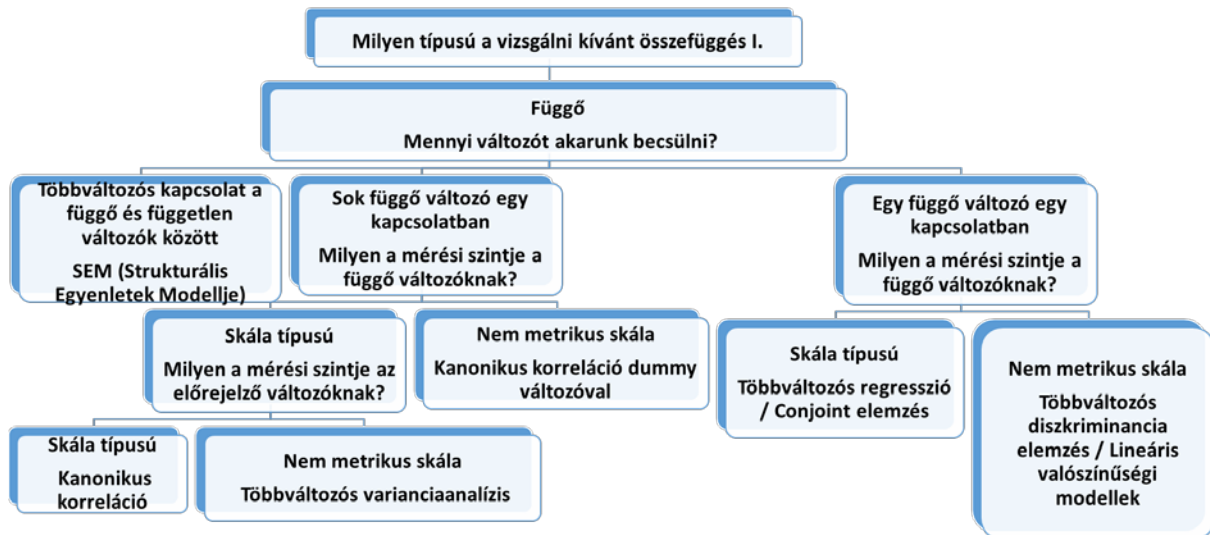
egy mintás t-próba,
két független mintás t-próbák,
párosított mintás t-próba,
egy és több szempontos varianciaelemzés,
korreláció- és regresszióelemzés,
logisztikus regresszió elemzés,
diszkriminanciaelemzés.

A **kölcsönös függőségi kapcsolatok** között a faktor- (változókat vizsgálunk) és klaszterelemzés (eseteket vizsgálunk).

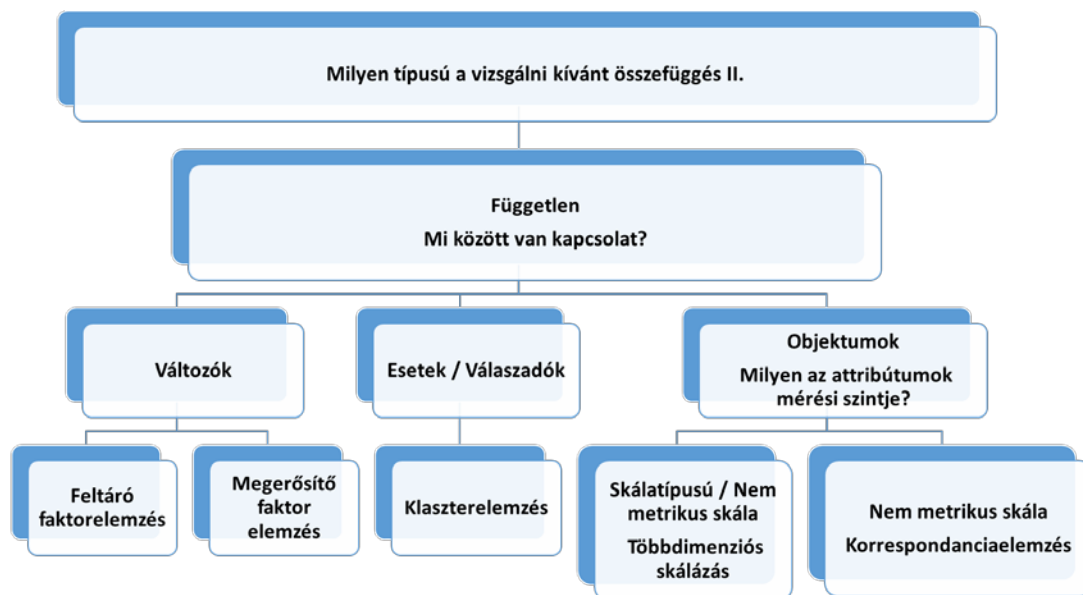
A következő ábrákon bemutatjuk, hogy milyen kérdések alapján lehet eldönteni, hogy melyik többváltozós elemzést alkalmazzuk (HAIR et al., 2014).

- Valamilyen elmélet alapján elkülöníthetők-e a változóink függő és független változókra?
- Ha elkülöníthetők a változók, akkor mennyi a függő változók száma egy elemzésben?
- Milyen mérési szintűek a függő és független változók?

A döntést az előzőekben ismertetett kérdésekre adott válaszok alapján hozhatjuk meg.



1. ábra: A többváltozós elemzési technikák kiválasztásának szempontjai I.
 Forrás: HAIR et al., (2014) alapján saját módosítás



2. ábra: A többváltozós elemzési technikák kiválasztásának szempontjai II.
 Forrás: HAIR et al., (2014) alapján saját módosítás

A mérési skálák egyik típusát a nem metrikus skálák jelentik, melyek egy tulajdonság meglétét vagy hiányát mutatják. A nem metrikus skálák két típusa a nominális (névleges) és az ordinális (sorrendi) skála. A mérési skálák másik típusát a metrikus skálák jelentik, melyek egyik típusa az intervallum (különbségi), másik típusa pedig az arányskála.

3. A hipotézisvizsgálatok és jelentőségük az elemzésekben

A hipotézisvizsgálatnak a marketingkutatáson belüli használata nagyon elterjedt és széles körűen alkalmazott eljárás. Alkalmazása arra irányul, hogy egy vagy több sokaságra vonatkozó olyan feltevések – ún. hipotézisek – helyességét vizsgálja, ellenőrizze mintavételi eredményekre támaszkodva, melyek fennállásában nem vagyunk biztosak. A hipotézisek a vizsgált sokaság eloszlására vagy az adott eloszlás egy vagy több paraméterére vonatkozhatnak. A hipotézisek helyességének ellenőrzésére különféle tesztek, próbákat használunk (VITA, 2011). A terjedelmi korlátok miatt most csak röviden teszünk említést az elméleti alapokról. További részletes információ található a különböző statisztikai tankönyvekben (pl. HUNYADI, 2001; RAMANATHAN, 2003; HUZSVAI, 2012; HUZSVAI – BALOGH, 2015).

3.1. A hipotézis vizsgálat menete

A hipotézisvizsgálatra a gyakorlatban minden olyan esetben szükség van, amikor valamely sokaság, illetve eloszlás jellemzőivel kapcsolatban bizonyos feltevéseink vagy elvárásaink vannak, s azok teljesülését nem teljes körű adatfelvételtől, hanem csak a sokaságból vett mintából nyert információkra támaszkodva tudjuk vizsgálni (VITA, 2011). A próbák fontos szerepet játszanak a különféle marketingkutatás segítségével elemzett jelenségek leírására törekvő statisztikai modellek építése és használata során.

3.1.1 A szakmai kérdés megfogalmazása, statisztikai formája

Minden hipotézisvizsgálat két egymásnak ellentmondó feltevés: egy H_0 -lal jelölt nullhipotézis és egy H_1 -gyel jelölt ellenhipotézis – más néven alternatív hipotézis – megfogalmazásával kezdődik. A két hipotézisnek olyannak kell lennie, hogy azok kizárják egymást és bármelyik alapján meg tudjuk válaszolni a bennünket érdeklő kérdést.

3.1.2 A szignifikanciaszint meghatározása

Az $1 - \alpha$ értéket (pl. $1 - \alpha = 0,95$) a próba megbízhatósági szintjének, az α (pl. $\alpha = 0,05$) értéket pedig szignifikanciaszintnek nevezzük. Ezeket az értékeket százalékos formában szokták közölni.

3.1.3 A próbastatisztika helyes megválasztása, számítása

Ennek során kiszámítjuk a próbafüggvényt, és meghatározzuk a próbastatisztika értékét. Ez a statisztikai teszt kiválasztását és alkalmazását jelenti. Lényeges információ, hogy csak az alkalmazhatósági feltételeknek megfelelő teszte(ke)t alkalmazhatjuk. A próbafüggvény kiszámított értéke az elméleti eloszlás értéke. Ez normáloszlás esetén a z-érték, míg t-eloszlás esetében a t-érték. Ezeknek az értékeknek az ismeretében megadható egy p valószínűség, amely megmutatja, hogy milyen valószínűséggel vehet fel a próbafüggvény a kiszámítottal azonos vagy nagyobb értéket, ha a H_0 igaz. Ha ez a valószínűség kisebb, mint a szignifikancia-szint,

akkor elutasíthatjuk a nullhipotézist, mert a hiba elkövetésének a valószínűsége kisebb, mint az előre választott megengedhető maximális érték (HUZSVAI, 2012).

3.1.4. A kritikus érték meghatározása, a statisztikai döntés H_0 elfogadásáról

Ha a próbafüggvénynek a minta adataiból számított értéke a visszautasítási tartományba esik, akkor elvetjük H_0 -t, ellenkező esetben megtartjuk azt az adott szignifikanciaszinten.

3.1.5. Szakmai következtetés

A hipotézisvizsgálat kapott eredményeinek ismeretében szakmailag kell meghoznunk a döntést arról, hogy milyen következtetést tudunk levonni.

1. táblázat

Néhány gyakran használt skála a marketingben

A mérés tárgya	Kategoróriák				
	nagyon rossz	rossz	közepes	jó	nagyon jó
<i>Attitűd</i>	nagyon rossz	rossz	közepes	jó	nagyon jó
<i>Fontosság</i>	egyáltalán nem fontos	nem fontos	közömbös	fontos	nagyon fontos
<i>Elégedettség</i>	nagyon elégedetlen	elégedetlen	közepesen elégedett	elégedett	nagyon elégedett
<i>Vásárlási szándék</i>	biztosan nem veszi meg	valószínűleg nem veszi meg	lehet, hogy megveszi, lehet, hogy nem	valószínűleg megveszi	biztosan megveszi
<i>Vásárlási gyakoriság</i>	soha	ritkán	néha	gyakran	nagyon gyakran

Forrás: MALHOTRA – SIMON (2016) alapján saját módosítás

3.2. Paraméteres próbák

Ha az eloszlás jellege ismert, és a nullhipotézisünk az eloszlás valamely paraméterére vonatkozik, akkor beszélünk paraméteres próbáról. Az ilyen jellegű vizsgálatoknál a próbák alkalmazása nominális és ordinális mérési szintű változók esetében nem ajánlott. A középértékek összehasonlítására leggyakrabban alkalmazott próbák az egy és két mintás z-próba és t-próba attól függően, hogy az alapsokaság szórása ismerte-e.

Fel kell hívni az olvasó figyelmét arra, hogy az IBM SPSS Statistics használatnak alapjai a „Kvantitatív piackutatási módszerek alapjai” című e-learning tananyagban már részletesen kifejtésre kerültek, ezért ebben az e-learning anyagban – a terjedelmi korlátok miatt – az SPSS program használatának ismertetése csak az adott elemzési módszer piackutatásban leggyakrabban előforduló funkcióinak bemutatására fog korlátozódni.

Ebben az anyagban a vizsgálataink során egy közös mintaadatbázist (*asvanyviz.sav*) fogunk használni 36 változóval, melyen keresztül bemutatjuk a különböző marketingkutatásban alkalmazott elemzési technikák kivitelezését és az eredmények értelmezéseit (3. ábra). A tisztított adatbázisban eredetileg 33 változó volt (plusz 3 új változót már mi állítottunk elő), míg a kérdőívet helyesen kitöltők száma 906 fő volt. A vizsgálat során számos kérdésre kellett a résztvevőknek válaszolnia az ásványvízzel kapcsolatos fogyasztói szokásokkal kapcsolatban.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role	
1	Sorszam	Numeric	8	0		None	None	8	Right	Nominal	Input
2	K1_1	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a körm.	{1, Egyáltal...}	None	18	Right	Scale	Input
3	K1_2	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással	{1, Egyáltal...}	None	20	Right	Scale	Input
4	K1_3	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni a	{1, Egyáltal...}	None	20	Right	Scale	Input
5	K1_4	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet f	{1, Egyáltal...}	None	20	Right	Scale	Input
6	K1_5	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a	{1, Egyáltal...}	None	13	Right	Scale	Input
7	K2_1	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy r	{1, Egyáltal...}	None	16	Right	Scale	Input
8	K2_2	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék v	{1, Egyáltal...}	None	14	Right	Scale	Input
9	K2_3	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben	{1, Egyáltal...}	None	15	Right	Scale	Input
10	K3_1	Numeric	8	0	Értékelje, hogy milyen fontos a következő környezeti szempont: Csökkenteni az éghajlat változást	{1, Nem font...}	None	20	Right	Ordinal	Input
11	K3_2	Numeric	8	0	Értékelje, hogy milyen fontos a következő környezeti szempont: Több tevékenység a természetvédel	{1, Nem font...}	None	14	Right	Ordinal	Input
12	K3_3	Numeric	8	0	Értékelje, hogy milyen fontos a következő környezeti szempont: A szemét és a háztartási hulladék c	{1, Nem font...}	None	15	Right	Ordinal	Input
13	K4	Numeric	8	0	A használt vizes palackokat visszazinné-e a vásárlás helyszínére?	{1, Nem}	None	5	Right	Ordinal	Input
14	K5_1	Numeric	8	0	Figyelembe veszi-e a MÁRKÁT, amikor palackozott vizet vásárol?	{0, Nem}	None	5	Right	Nominal	Input
15	K5_2	Numeric	8	0	Figyelembe veszi-e a MERETET, amikor palackozott vizet vásárol?	{0, Nem}	None	5	Right	Nominal	Input
16	K5_3	Numeric	8	0	Figyelembe veszi-e a FORMÁT, amikor palackozott vizet vásárol?	{0, Nem}	None	5	Right	Nominal	Input
17	K5_4	Numeric	8	0	Figyelembe veszi-e a SÜLYT, amikor palackozott vizet vásárol?	{0, Nem}	None	5	Right	Nominal	Input
18	K5_5	Numeric	8	0	Figyelembe veszi-e a CSOMAGOLÁS DESIGNT, amikor palackozott vizet vásárol?	{0, Nem}	None	5	Right	Nominal	Input
19	K5_6	Numeric	8	0	Figyelembe veszi-e a TERMÉK VÉDELMEZET, amikor palackozott vizet vásárol?	{0, Nem}	None	5	Right	Nominal	Input
20	K5_7	Numeric	8	0	Figyelembe veszi-e az ANYAG MINŐSÉGÉT, amikor palackozott vizet vásárol?	{0, Nem}	None	5	Right	Nominal	Input
21	K5_8	Numeric	8	0	Figyelembe veszi-e a ZÖLD CSOMAGOLÁST, amikor palackozott vizet vásárol?	{0, Nem}	None	5	Right	Nominal	Input
22	K5_9	Numeric	8	0	Figyelembe veszi-e az ÁRAT, amikor palackozott vizet vásárol?	{0, Nem}	None	5	Right	Nominal	Input
23	K6_1	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: Követem a környezetbarát témákat	{1, Soha}	None	5	Right	Scale	Input
24	K6_2	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: Kényelmetlenséget is vállaltam azért, hogy	{1, Soha}	None	11	Right	Scale	Input
25	K6_3	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: Vásárlás közben a saját bevásárló táskám	{1, Soha}	None	5	Right	Scale	Input
26	K6_4	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: A háztartásomban újrahasznosított a dolg	{1, Soha}	None	5	Right	Scale	Input
27	K6_5	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: A vásárlásaim során olyan termékeket vála	{1, Soha}	None	5	Right	Scale	Input
28	K6_6	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: A nem környezetbarát viselkedése/írma	{1, Soha}	None	5	Right	Scale	Input
29	K6_7	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: Adományozok a környezetvédelemmel kap	{1, Soha}	None	5	Right	Scale	Input
30	K7	Numeric	8	0	Nemek	{0, Nő}	None	11	Right	Nominal	Input
31	K8	Numeric	8	0	Életkor kategóriák	{1, 25-34}	None	12	Right	Ordinal	Input
32	K9	Numeric	8	0	Legmagasabb iskolai végzettség	{1, Maxima}	None	18	Right	Ordinal	Input
33	K10	Numeric	8	0	Havi jövedelem (ezer Ft)	{0, Nem vál...}	0	20	Right	Ordinal	Input
34	K4_két_kategória	Numeric	8	0	A használt vizes palackokat visszazinné-e a vásárlás helyszínére?	{0, Nem / T...}	None	20	Right	Nominal	Input
35	K1_összesen	Numeric	8	0	Környezetbarát attitűd összesen	None	None	14	Right	Scale	Input
36	K6_összesen	Numeric	8	0	Környezetbarát viselkedés összesen	None	None	14	Right	Scale	Input

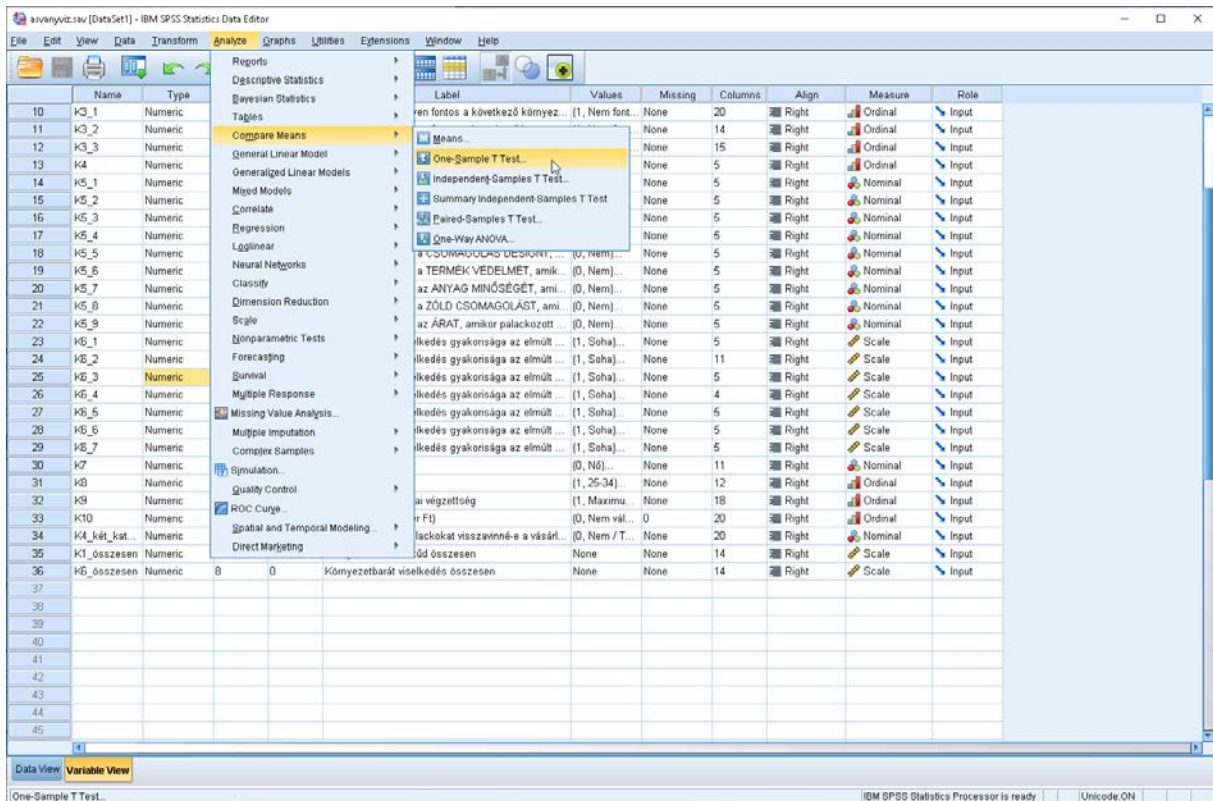
3. ábra: Az SPSS adatbázis bemutatása (*asvanyviz.sav*)

A továbbiakban bemutatjuk azokat az elemzési lehetőségeket, amelyekkel számos szakmai kérdésekre kapjuk meg a válaszokat.

3.2.1 Egymintás t-próba

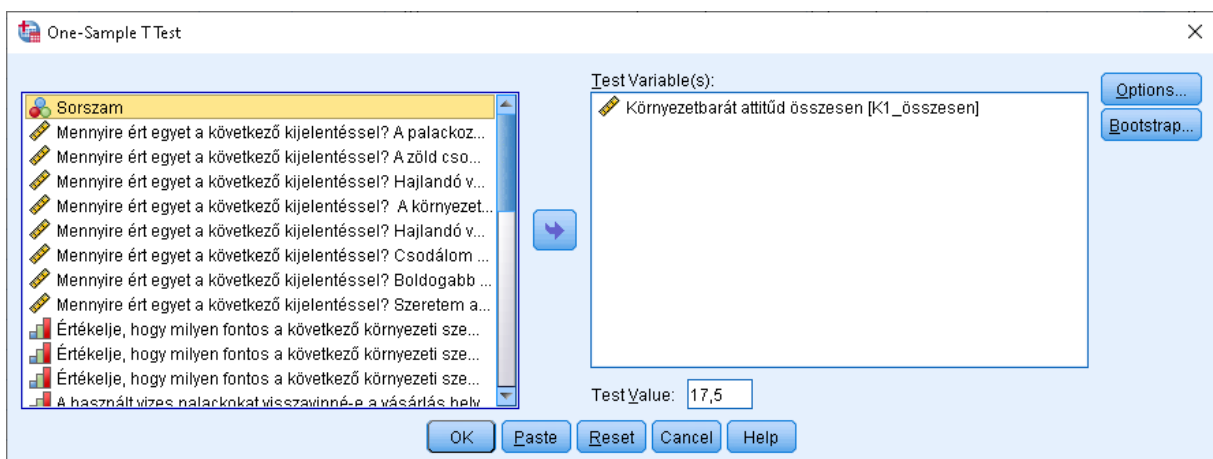
Kérdés: *Vizsgáljuk meg, hogy a válaszadók un. környezet iránti attitűdjei együttesen milyenek és az átlaguk megegyezik-e egy már lefolytatott másik kutatás átlagával (17,5-del)! Ehhez előzetesen a „Mennyire ért egyet a következő kijelentéssel?” kérdések közül az első ötöt (K1_1 – K1_5) összegeztük. Az így létrehozott új változó neve a „Környezetbarát attitűd összesen” (K1_összesen) volt.*

Ha választ akarunk adni a fenti kérdésre, ahhoz az egy mintás t-próbát kell kiszámítanunk. A számítást az SPSS ANALYZE / COMPARE MEANS / ONE-SAMPLE T TEST... menüpontjával tudjuk elvégezni (4. ábra).



4. ábra: Az egy mintás t-próba indítása

A párbeszédablakban a bal oldali változólistából vigyük át az általunk vizsgálni kívánt változót (K1_összesen). Ezt követően a „Test Value” cellába írjuk be az általunk már ismert 17,5-es átlagértéket (5. ábra). Meg kell említeni, hogy egyszerre több változót is elemezhetünk az SPSS segítségével, ha ugyan azt az átlagértéket szeretnénk mindegyik esetében tesztelni. Ha készen vagyunk, kattintsunk az „OK”-ra. Az eredmények az SPSS Viewer (vagy output-) ablakban jelennek meg.



5. ábra: Az egy mintás t-próba beállításai

Az outputablakban megjelenő két táblázatot a 6. ábra mutatja. Az első táblázat „One-Sample Statistics” a változó leíró statisztikai jellemzőit mutatja. Az első oszlopban a változó un. hosszú neve (label: Környezetbarát attitűd összesen) szerepel. Ezt követi az átlag („Mean”), szórás

(„Std. Deviation”) és az átlag hibája („Std. Error Mean”). Ezekből leolvashatjuk, hogy a válaszadók átlaga 17,63 volt, amit a program a második táblázatban fog összehasonlítani az általunk megadott értékkel (17,5). A második táblázat „One-Sample Test” tartalmazza az egy mintás t-próba eredményeit. A számított t-érték („t”) kicsi (1,041), a szabadság fok („df”) 905 és a szignifikancia-szint („Sig. (2-tailed)”) 0,298, ami nagyobb, mint 0,05, így azt mondhatjuk, hogy a mintánk átlaga (17,63) és az általunk megadott előzetes érték („Test Value = 17.5”) közötti különbség nem szignifikáns. *Figyeljünk arra, hogy az SPSS program tizedes vessző helyett tizedes pontot jelez ki!* A két érték közötti különbséget a („Mean Difference”) jelzi, de három tizedes pontossággal (0,126 ~ 0,13). Az output tábla két utolsó oszlopa – az általunk megadott szinten (95%) – a különbség konfidenciaintervallum alsó és felső határát („95% Confidence Interval of the Difference”) is megmutatja nekünk (-0,11 – 0,36). Ha ez a tartomány tartalmazza a nullát, az azt jelenti, hogy a két átlag közötti különbség lehet nulla, azaz a mintánk átlaga nem különbözik a tesztelt értéktől.

The screenshot shows the SPSS One-Sample Test results for the variable 'Környezetbarát attitűd összesen'. The output is divided into two main sections: 'One-Sample Statistics' and 'One-Sample Test'.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Környezetbarát attitűd összesen	906	17,63	3,639	,121

One-Sample Test

Test Value = 17.5

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Környezetbarát attitűd összesen	1,041	905	,298	,126	-,11	,36

The status bar at the bottom of the window indicates: 'Log is visible', 'IBM SPSS Statistics Processor is ready', and 'Unicode:ON | H: 2.75, W: 33.52 cm'.

6. ábra: Az egy mintás t-próba eredményei

3.2.2 Két független mintás t-próba

A kétmintás t-teszttel megvizsgálhatjuk, hogy származhat-e a két független megfigyelés, minta azonos középértékű populációból? Azonosnak tekinthető-e a két populáció középértéke, amelyekből a minták származnak?

A középértékek összehasonlítására szolgáló statisztikai próbák eltérőek attól függően, hogy az alappopulációk szórása egyenlőnek tekinthető-e. Amennyiben a szórások megegyeznek az alábbi próbastatisztikát használjuk, az eloszlás t-eloszlású, $DF = n_1 + n_2 - 2$ szabadságfokkal. A nevezőben az s_p a két minta összevont varianciájának (pooled variance) négyzetgyökét jelenti, melyet a két minta összevont szórásának nevezünk.

Alkalmazhatósági feltételek:

- Két független minta,

- Normális eloszlású sokaságok,
- A szórások ismeretlenek, de azonosak

Az ismeretlen közös szórást a mintákból számított szórásnégyzetekből becsülhetjük az alábbi képlet szerint.

A két populáció középértéke, amelyekből a minták származnak, abban az esetben tekinthetők azonosnak, ha:

$$|t| \leq t^*$$

A próbastatisztika kritikus t -értékét kétoldali alternatív hipotézis esetén 2 -nél, egyoldali alternatív hipotézis esetén, kell a táblázatból meghatározni. Ha a két populáció ismeretlen szórásnégyzete korábbi ismeretek, ill. a mintákból számított szórásnégyzetek alapján nem tekinthető azonosnak, akkor a t -próba helyett a Welch-próbát kell alkalmazni, mely igen hasonló a t -próbaéhoz, a különbség a szabadságfokok meghatározásában van.

A t -teszt alkalmazásakor előre tudni kell, hogy a két csoport szórása megegyezik-e, tehát tesztelni kell a csoportok szórását (F-próba). Amennyiben a szórások egyenlők, akkor a vizsgálatba vont összes csoportból kell a varianciát becsülni (pooled variance). A próba valószínűségi változója t -eloszlású, így a középértékek különbségének szignifikanciája a kritikus t -érték alapján állapítható meg.

Amennyiben a két csoport szórása szignifikánsan különbözik, ilyenkor a két összehasonlítandó csoport varianciáját súlyozni kell a variancia becsléséhez (separate variancia). A próba valószínűségi változója ebben az esetben nem t -eloszlású, ezért a szabadságfokokat Bonferroni módszerével korrigálni kell, és ezt kell használni a középértékek különbözőségének elbírálásakor (a szabadságfokok korrekciója (Bonferroni teszttel)).

3.2.3 Párosított mintás t -próba

Párosított t -próbát akkor használunk, ha a két minta elemei páronként összefüggnek, pl. ugyanazon egyeden két különböző időpontban mérünk egy tulajdonságot, vagy valamilyen csoportképző tulajdonság alapján párokat tudunk képezni.

A két minta középértékének azonossága helyett a párosított minták d különbségének (előjeles) várható értékére fogalmazzuk meg a H_0 hipotézist.

Az előző eljárásokhoz hasonlóan itt is z -, ill. t -próbát alkalmazhatunk attól függően, hogy ismert-e a d különbségek eloszlása és szórása, illetve mekkora a minta elemszáma?

Alkalmazhatósági feltételek:

- A d különbségek eloszlása normális
- d ismeretlen (a mintából számított)

A párosított t -próba statisztikája.

A próba t-eloszlású, $DF=n-1$ szabadságfokú. A képletben sd a párosított minták különbségének szórása, amelyet a minta alapján becsülünk.

A t-próba ereje

A statisztikai próba ereje a korábban definiáltak szerint: a valódi d különbség kimutatásának valószínűsége. Ezt $1-\beta$ -val jelöltük. Annál erősebb egy statisztikai próba, minél nagyobb valószínűséggel mutatja ki a valódi hatást. A t-próbánál a t-érték valójában egy standardizált hatás (Es, standardised effect), amelynél két csoport átlagának különbségét osztjuk a különbség várható értékének szórásával.

A standardizált hatás nagysága alapján:

- kicsi 0,2
- közepes 0,5
- nagy hatás 0,8 feletti

Amennyiben a két csoport várható értéke megegyezik, a különbségük várható értéke nulla körül mozog. Tehát, ha H_0 igaz, t várható értéke nulla.

A d valódi különbség létezésekor a hatás kimutatásának valószínűsége függ:

- a minta elemszámától
- a d nagyságától
- a szórástól
- az elsőfajú hiba nagyságától, azaz a szignifikancia-szinttől
- a t-próba típusától (egymintás, kétmintás, párosított)
- alternatív hipotézistől (egyoldali vagy kétoldali)

A fenti tényezők zömét az analízis megkezdése előtt tudjuk beállítani. Ilyen a minta elemszáma, szignifikancia-szint, t-próba típusa, alternatív hipotézis. A szórás a vizsgált jelenség tulajdonsága, ezt csak megbecsülni lehet. A valódi különbség nagyságáról csak előzetes információink lehetnek, pl. korábbi szakirodalmi adatok alapján.

Vizsgálataink során nagyon fontos előre tudni, hogy adott különbséget mekkora valószínűséggel lehet kimutatni. Ekkor tervezzük meg a kísérletet, felmérést, a minimális mintaelemszámot. Ebben a fázisban kell eldönteni, hogy egyáltalán érdemes-e hozzáfogni a vizsgálathoz. Ehhez először a másodfajú hiba nagyságát kell meghatározni. Hogyan? Meg kell határozni a d középértékű t-eloszlásnál egy adott értéknél kisebb értékek előfordulási valószínűségét. Mit jelent az adott érték? A kritikus t-értéket. Ha a kritikus t-értéknél kisebb számított t-értéket kapunk, a nullhipotézist kell elfogadni akkor is, ha a d különbség valóban létezik. Azt mondhatjuk, hogy a nullhipotézis erősebb.

3.2.4 Egy és több szempontos varianciaelemzések

A variancia-analízis a t-próba kiterjesztése kettőnél több minta esetére. Tehát három vagy több mintával rendelkezünk. Mindegyik minta egy csoportképző ismérv egy-egy szintjét reprezentálja. Pl. különböző kefir márkák. A márkákon belül egy skála típusú változó várható

értékét vizsgálhatjuk. Megvizsgálhatjuk, hogy a különböző kefirek várható eladási árai, zsírtartalmai, fehérjetartalmai, stb. megegyeznek-e. Ez azt jelenti, hogy a skála típusú változó nem függ a nominális, csoportképző változótól. A függőváltozó a variancia-analízis modellben mindig valamilyen skála típusú, a független változó(k) nominális mérési szintűek. Amennyiben az árak, stb. nem egyeznek meg a különböző kefireknél, akkor összefüggés van közöttük, és a márkákkal részben magyarázhatjuk a különbségeket. A magyarázat a függő változó teljes heterogenitásának két részre bontását jelenti. A teljes heterogenitás egyik része az, amelynek „okai” a független változók, a másik heterogenitás-rész pedig az, amelynek „okait” az egyéb, általunk nem vizsgált tényezők tartalmazzák. Ez utóbbit sokszor a véletlen hatásaként is emlegetik.

A heterogenitás mérésére korábban többféle mérőszámot ismertettünk, ismétlésként a legfontosabbak az alábbiak:

- (1) terjedelem (range); a legnagyobb és legkisebb érték közötti távolság
- (2) átlagos eltérés; ;
- (3) szórás; ;
- (4) variancia- vagy szórásnégyzet; .

Alapfogalmak

Nézzük át azokat az alapfogalmakat, amelyeket a variancia-analízis során használunk.

- a) Faktor: Faktornak nevezzük a vizsgálatba bevont független változókat, pl. különböző kezeléseket, tényezőket, ilyen a kefir márka. Kísérletekben inkább kezeléseknak hívjuk.
- b) Faktor szint: A faktor értékkészletének az eleme, mely beállítása mellett vizsgálhatjuk meg a függő változónkat. A kezelése szintjei, pl. kefir márkán belül Danone, Milli, Müller, stb. Kísérletben pl. műtrágyaadagok.
- c) Kvalitatív és kvantitatív faktorok: Ha a faktorszintek nem numerikusak vagy intervallum skálájúak, akkor kvalitatív, ellenkező esetben kvantitatív faktorokról beszélünk.
- d) Cellák: Egyfaktoros modellekben a cellák megfelelnek a faktorok szintjeinek, többfaktoros esetben a figyelembe vett faktorok szintjeiből előálló kombinációk a cellák. Pl. amikor a 2 faktor műtrágyaadagok és öntözési módok, akkor a cellák a (műtrágyaadagok, öntözési módok) összes lehetséges kombinációjából állnak.
- e) Interakció: Két független változó kapcsolatában akkor áll fenn interakció (kölsönhatás), ha változó hatása függ az változó szintjétől és fordítva.
- f) Egyszempontos variancia-analízis: Variancia-analízis, ahol csak egy faktor van. Egyutasnak is nevezik
- g) Többszempontos variancia-analízis: Variancia-analízis, ahol kettő vagy több faktor van.
- h) Egyváltozós variancia-analízis: amelyben csak egy függő változót vizsgálunk.
- i) Többváltozós variancia-analízis: amelyben kettő vagy több függő változót elemzünk.

A lineáris modell

Alkossuk meg az egytényezős variancia-analízis matematikai modelljét. Egy kísérletben k számú kezeléssel vagy populációval és r számú ismétléssel rendelkezünk. Az adataink száma tehát $n=k*r$. Minden mért adat y_{ij} felbontható három összetevőre, amelyek: a kísérlet főátlaga (m), a kezeléshatás (A_i), és a meg nem magyarázott rész, a maradék (e_{ij}). A maradéktagokat hibának is nevezik (error).

Az egytényezős lineáris modell:

Valójában a kezeléshatás az m és A_i összege. Ez a kettő adja a lineáris modellel becsült értéket, azaz a modellezett értéket. Az A_i a kísérlet főátlagtól vett eltérést jelenti (kezeléshatás-főátlag). A korábban tanultak szerint az alapadatok átlagtól vett eltéréseinek összege nulla. Ez a lineáris modellre is igaz, a kezeléshatások összege nulla, vagyis a kezelések szimmetrikusak a főátlagra. Az e_{ij} maradéktagok tulajdonságai nagyon fontosak, amelyek egyben megegyeznek a variancia-analízis alkalmazhatósági feltételeivel, melyeket a következőkben ismertetünk.

A variancia-analízis alkalmazásának feltételei

Az alkalmazhatósági feltételek a maradéktagokra vonatkoznak:

(a) Az egyes kezelésekhöz tartozó maradékoknak függetleneknek kell lenniük a blokk, a kezeléshatástól és a függő változótól. Ezt leginkább a kísérleti elrendezéssel, randomizálással biztosíthatjuk. A függetlenség azt jelenti, hogy a maradékok nagyságát nem befolyásolhatja a kezelés. Amennyiben hatással van rá, akkor ez keveredhet a kezeléshatással, és torz becslést kapunk, helytelen becsült értékeket fogunk előállítani.

(b) Az maradék normális eloszlású, nulla várható értékű valószínűségi változó. Attól, hogy egy normál-eloszlású mintához egy konstans értéket hozzáadunk, vagy abból levonunk, az eloszlás és a minta szórása nem változik. A normalitást korábban ismertetett módszerek valamelyikével ellenőrizhetjük. (Megjegyezzük, hogy a matematikai-statisztikai kézikönyvek az ANOVA-t robusztus eljárásnak tekintik, s azt állítják, hogy a függő változónak nem kell normális eloszlásúnak lennie). Ha matematikailag korrekt módon akarjuk az ANOVA-t használni, akkor a függő változót normális eloszlásúvá transzformálhatjuk. Azért kell normális eloszlásúnak lennie, mert a hatások megítélésekor a normál-eloszlás tulajdonságait használjuk fel, az eloszlás nevezetes értékeit.

(c) A maradékok szórásnégyzetei a kezeléskombinációkon belül azonosak, azaz homoszkedasztikus a modell. (Az SPSS programban ezt a homogenitást a Levene teszt alapján tesztelhetjük.)

A variancia-analízis alkalmazásának lépései

1. A variancia-analízis modell felállítása
2. Szignifikancia-szint megválasztása
3. A variancia-analízis kiszámítása, az F-próba
4. A modell érvényességének ellenőrzése
5. Amennyiben az F-próba szignifikáns, középértékek többszörös összehasonlítása

A középértékre vonatkozó hipotézisek a következők:

azoknak a populációknak a középértékei, amelyekből a minták származnak azonosak.

létezik legalább egy olyan középérték pár, ahol a középértékek nem tekinthetők azonosnak, legalább egyszer.

Az analízis megkezdése előtt ábrázolni kell az alapadatokat. Olyan ábrát érdemes készíteni, amelyben a várhatóérték mellett a középérték hibáját (se) is ábrázoljuk. Erre azért van szükség, mert ha csak az átlagokat tüntetjük fel az y -tengely léptékétől függően nagyon kicsi

különbségeket is fel lehet nagyítani, és a jelentős különbségeket is el lehet tüntetni. A standardizált hatások, amit az angol szakirodalomban „standard effect” néven emlegetnek, nem más mint a kezeléshatás osztva a szórással, ingadozással. Ez azt mutatja, hogy a kezeléshatás, hogyan aránylik a szóráshoz, azaz a véletlen ingadozáshoz.

Az így meghatározott standard hatások nagyságát a korábban ismertetett módon ítélni lehetjük meg.

1. A variancia-analízis modell felállítása

A módszer alapgondolata szerint a modellben a mérési, megfigyelési értékeket összegként képzeljük el. Az n megfigyelés mindegyikére a korábban ismertetett modellegyenlet írható fel, amelynek alapján a mintaelemekeken mért, ill. megfigyelt y_{ij} értékek felbonthatók a modell által meghatározott részre és a hibára. A modell által meghatározott rész a szisztematikus hatásokat tartalmazza, a hibakomponens pedig a véletlen hatást jelenti.

A variancia-analízis legegyszerűbb modelljében a vizsgálatban szereplő k számú populációból egyszerűen r elemű véletlen mintát veszünk, majd a mintánkénti középértékeket hasonlítjuk össze, ezt nevezzük egyszempontos variancia-analízisnek (kísérlet esetén teljesen véletlen elrendezésnek). Az elrendezés modellegyenlete, ahol X_{ij} az i -edik minta j -edik eleme; μ a kísérlet vagy minta főátlaga; α_i az i -edik mintához tartozó populáció hatása (növelheti vagy csökkentheti a főátlagot); ϵ_{ij} véletlen hatás. Ebben a modellben a modell által meghatározott rész, csak az i -edik mintához tartozó populáció várható értékét tartalmazza, tehát szisztematikus különbséget csak a populációk várható értékei között tételezhetünk fel. A véletlen okozta hatásokat a hibakomponens tartalmazza. Amennyiben teljesülnek a variancia-analízis alkalmazásának feltételei, akkor $\sum \alpha_i$ összege nulla, és ϵ_{ij} normális eloszlású nulla várhatóértékű sokaság, és független a blokk és kezeléshatástól, valamint a modell homoszkedasztikus.

2. Szignifikancia-szint megválasztása

A szignifikancia-szint nagyságát leggyakrabban 5%-nak választják. Ez az érték szerepel legtöbb statisztikai programban is kezdeti értéként. Amennyiben túl szigorúnak ítélik ezt, választhatunk 10%-os szintet is. Ebben az esetben a kezelés okozta valódi hatások kimutatásának nagyobb a valószínűsége. Természetesen az elsőfajú hiba ilyenkor 5-ről 10%-ra nő. A szignifikancia-szintet választhatjuk 1 vagy 0,1%-nak is. Ezek már nagyon szigorú feltételek, alig követünk el elsőfajú hibát, de annál nagyobb a valószínűsége a másodfajú hibának. Elméletileg bármilyen szignifikancia-szintet választhatunk, ha szakmailag meg tudjuk indokolni. Amennyiben eldöntöttük az elsőfajú hiba nagyságát, meg tudjuk határozni a kritikus F -értéket. A kritikus F -érték az a legnagyobb érték, amelyet a véletlen ingadozás mellett kaphatunk. Ennél kisebb érték esetén a H_0 -t kell elfogadni.

3. A variancia-analízis kiszámítása, az F-próba

Az Excel táblázatkezelő eredménytáblázatában az alábbi fogalmakkal találkozunk (a számításhoz szükséges beállításokat a példában ismertetjük):

Tényezők: a variancia okai

Eltérés-négyzetösszegek (SS)

Csoportok között: kezeléshatások eltérés-négyzetösszegei * r .

Csoporton belüli: kezeléseken belül az eltérés-négyzetösszegek összege

Összes: alapadatok eltérés-négyzetösszegei

Szabadságfokok:
Csoportok között: $k-1$
Csoporton belül: $n-k$
Összesen: $n-1$

Varianciák: eltérés-négyzetösszegek osztva a szabadságfokokkal.

F-próba

Az F-eloszlás sűrűségfüggvényét mutatja az ábra. Az x-tengelyen az F-értékek, az y-tengelyen a valószínűségek láthatók. A függőleges kék vonal mutatja a kritikus F-értéket. Ezt a szignifikancia-szint és a két szabadságfok ismeretében tudjuk meghatározni. Korábban említettük, ha ennél kisebb a számított F, akkor a nullhipotézist kell elfogadni. Az ábrán a függőleges piros vonal a számított F-értéket mutatja. Ez jóval nagyobb, mint a kritikus, ezért már nem tekinthető a véletlen ingadozás hatásának, a nullhipotézist vissza kell utasítani. Mi az elsőfajú hiba elkövetésének valószínűsége ebben az esetben? Az 5,84 értéknél nagyobb értékek előfordulási valószínűsége. Példánkban ez nem éri el a 0,1%-t sem, ezért nyugodtan elvethetjük a nullhipotézist.

Mikor szignifikáns az F-próba? Amennyiben szakmailag teljesen korrektek akarunk lenni, akkor azt kell válaszolni, ha létezik legalább egy szignifikáns kontraszt a csoportok között. A kontraszt egy lineáris összehasonlító függvény. A függvény együtthatóinak összege nulla.

4. Két és többváltozós lineáris regresszió számítása

4.1. Változók közötti összefüggések elemzése

Amennyiben két metrikus változó közötti kapcsolatot szeretnénk megvizsgálni, akkor korrelációs számítást kell végezni. A lineáris (vagy Pearson-féle) korrelációs együttható -1 és +1 közötti értéket vehet fel, abszolút értéke a kapcsolat szorosságát, előjele a kapcsolat irányát mutatja. A táblázat összefoglalja a korrelációs együttható értékelését.

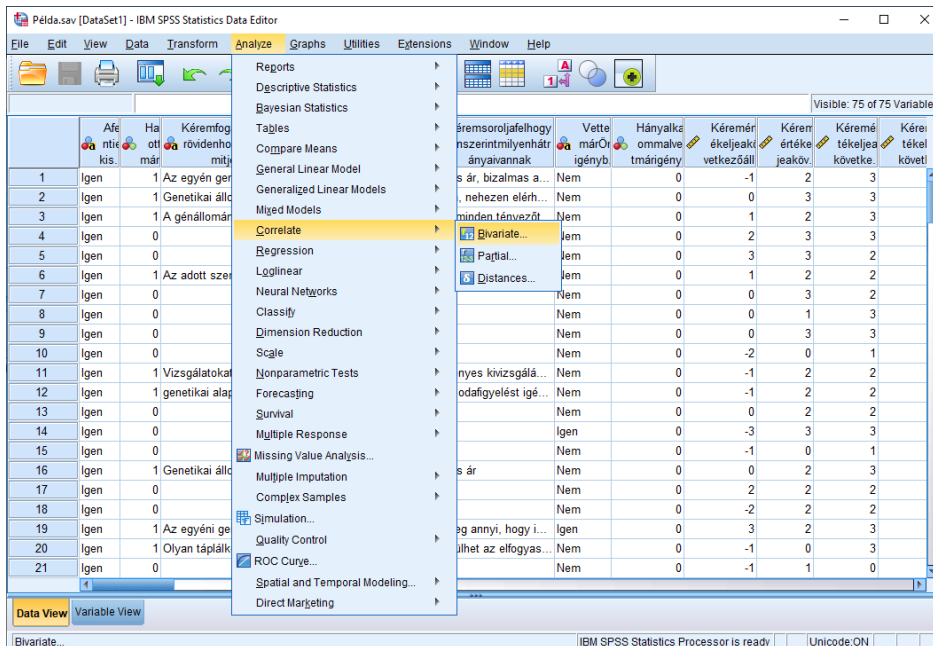
2. táblázat

A korrelációs együttható lehetséges értékei

<i>Korreláció (r) értéke</i>	<i>Kapcsolat iránya és erőssége</i>
$r=1$	tökéletes pozitív kapcsolat (tökéletes együttjárás)
$0,7 \leq r < 1$	erős pozitív kapcsolat
$0,2 \leq r < 0,7$	közepes pozitív kapcsolat
$0 < r < 0,2$	gyenge pozitív kapcsolat
$r=0$	nincs lineáris kapcsolat
$-0,2 < r < 0$	gyenge negatív kapcsolat
$-0,7 < r \leq -0,2$	közepes negatív kapcsolat
$-1 < r < -0,7$	erős negatív kapcsolat
$r=-1$	tökéletes negatív kapcsolat (teljesen ellentétes mozgás)

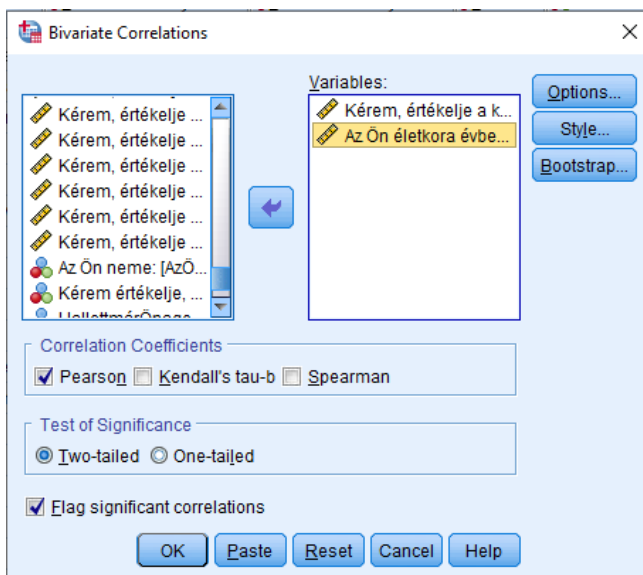
Forrás: SAJTOS – MITEV (2007)

A példát folytatva vizsgáljuk meg, hogy van-e összefüggés a válaszadók kora és a között, hogy mennyire tartják valószínűnek, hogy igénybe fogják venni a személyre szabott táplálkozási tanácsadást. A korrelációs számítást az SPSS ANALYZE / CORRELATE / BIVARIATE menüpontjával tudjuk elvégezni.



7. ábra. Korrelációs számítás indítása

A párbeszédablakban a már szokásos módon a bal oldali változólistából vigyük át a két szóban forgó változót. Hagyjuk bejelölve a „Pearson Correlations Coefficient”-et, a kétoldali szignifikanciatesztet („Two-tailed Test of Significance”) és a „Flag significant correlations” (jelölje meg a szignifikáns korrelációkat) lehetőséget. A példában csak két változó közötti korrelációt vizsgáljuk, de egy lépésben több változót is be lehet emelni az elemzésbe. Ha készen vagyunk, kattintsunk az „OK”-ra.



8. ábra. A korrelációs számítás beállításai

Az outputablakban megjelenő táblázatot (korrelációs mátrix) az ábra mutatja. A táblázat átlójában 1-es korrelációk szerepelnek, hiszen ott a változók önmagukkal vett korrelációja van, ami definíció szerint 1. A táblázat (és a korreláció) szimmetrikus, így az átló alatti vagy feletti értékeket kell csak nézni. Mivel jelen példában csak két változót vizsgáltunk, egyetlen

korrelációs értékünk van, a $-0,042$, ami egy nagyon gyenge negatív kapcsolatra utal. Azonban a korrelációs együttható szignifikancia-szintje („Sig. (2-tailed)”) $0,603$, ami nagyobb, mint $0,05$, így azt mondhatjuk, hogy a két változó közötti korreláció nem szignifikáns. Ha szignifikáns lett volna, akkor a korrelációs együttható után csillagokkal jelezte volna a program a szignifikáns kapcsolatot (*: 5%-on szignifikáns, azaz $0,05$ -nél kisebb a szignifikancia-szint; **: 1%-on szignifikáns, azaz $0,01$ -nél kisebb a szignifikancia-szint). A korreláció értékelésénél vegyük figyelembe, hogy az együttható csak a két változó együttjárásának mértékét mutatja, ok-okozati kapcsolatot nem.

Correlations

Correlations			
		Kérem, értékelje a következő állításokat a GENETIKAI ALAPÚ SZEMÉLYRE SZABOTT TÁPLÁLKOZÁSI TANÁCSADÁSSAL kapcsolatban! (-3=nagyon valószínűtlen, 3=nagyon valószínű) [A jövőben biztosan igénybe fogom venni ezt a szolgáltatást.]	Az Ön életkora évben (pl.: 27):
→ Kérem, értékelje a következő állításokat a GENETIKAI ALAPÚ SZEMÉLYRE SZABOTT TÁPLÁLKOZÁSI TANÁCSADÁSSAL kapcsolatban! (-3=nagyon valószínűtlen, 3=nagyon valószínű) [A jövőben biztosan igénybe fogom venni ezt a szolgáltatást.]	Pearson Correlation	1	-.042
	Sig. (2-tailed)		.603
	N	155	155
Az Ön életkora évben (pl.: 27):	Pearson Correlation	-.042	1
	Sig. (2-tailed)	.603	
	N	155	155

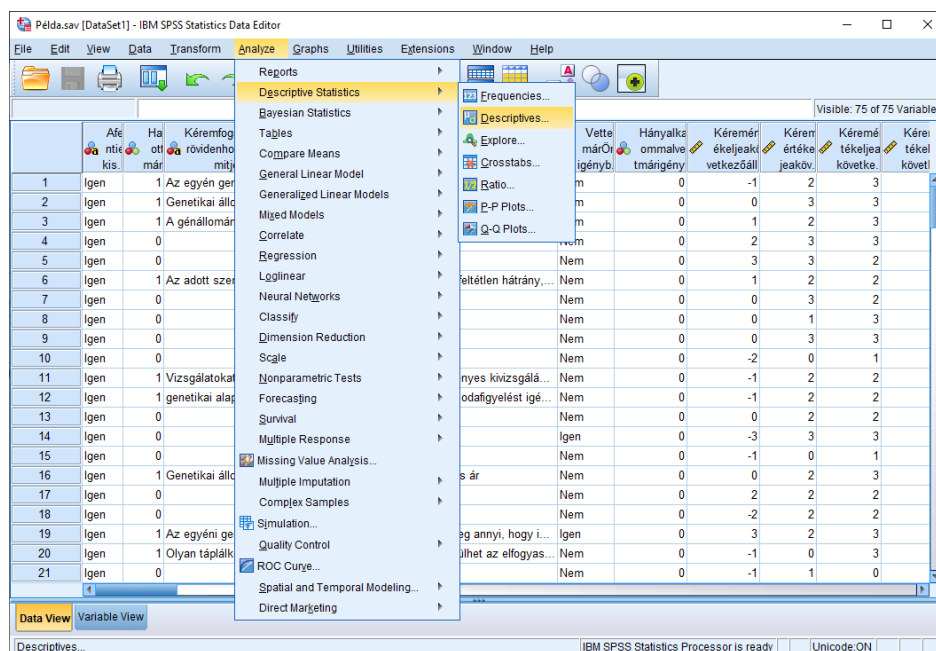
9. ábra. A korrelációs mátrix

5. A logisztikus regresszió számítása

5.1. Az alapmodell

A logisztikus regresszió a leginkább hasonló eljárás a diszkriminancia-analízishoz, tekintve, hogy a kategorikus függő változó magyarázatára használható. A logisztikus regresszió abban az esetben választandó eljárás, ha nem észszerű feltételezni a független változók normális eloszlását, amely a diszkriminancia-analízis szükséges feltétele. Emellett a diszkriminancia-analízisnek feltétele a homoszkedaszticitás és szükséges, hogy a magas mérési szintű független változók egy értelmes magyarázó-struktúrába legyenek rendezhetőek. Ezekre a feltételekre a logisztikus regresszió nem érzékeny. Külön problémát jelenthet diszkriminancia-analízis esetében a diszkrimináló függvények nehéz értelmezhetősége, míg a logisztikus regresszió esetében bonyolultabb statisztikai konstrukciókkal kell szembenéznie a felhasználóknak.

Az SPSS-ben a fenti mutatószámok egy részét az ANALYZE / DESCRIPTIVE STATISTICS / DESCRIPTIVES menüpontban találjuk meg.



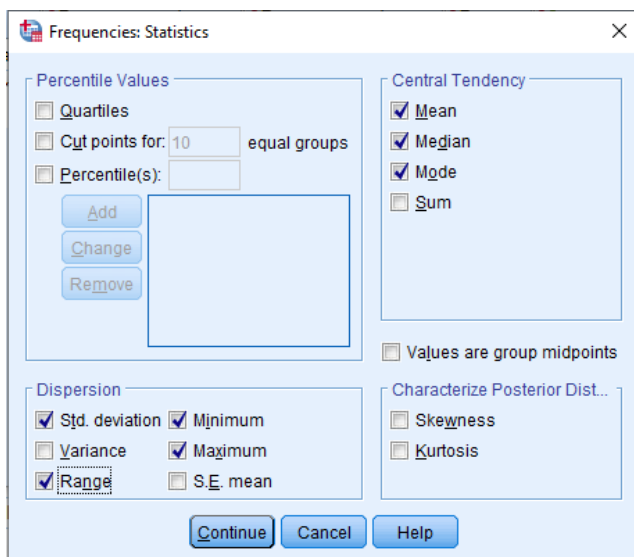
10. ábra. Leíró statisztikák lekérése

A megnyíló új párbeszédablak bal oldaláról kell kiválasztani az(oka)t a változó(ka)t, amely(ek)re vonatkozóan le akarjuk kérni a statisztikákat (egyszerre több változóra is le lehet kérni), majd a kiválasztott változó(ka)t áttesszük a jobbra mutató nyíl segítségével a jobb oldali „Variable(s):” boxba. Például a korábbi példát folytatva arra vagyunk kíváncsiak, hogy átlagosan milyen valószínűséggel vennének igénybe a megkérdezettek genetikai alapú személyre szabott táplálkozási szolgáltatást egy -3 (nagyon valószínűtlen) és +3 (nagyon valószínű) skálán mérve, mi volt a minimum és maximum érték, amit adtak, illetve mi volt az adott értékek terjedelme. Ezért a szolgáltatás igénybe vételére irányuló kérdés változóját áttesszük a boxba, majd az „Options” gombra kattintunk.

Az eredmények az SPSS Viewer (vagy output-) ablakban jelennek meg (35. ábra). Az ablak bal oldali sávjában fa struktúrában látjuk az elvégzett műveleteket; ha már sok elemzést hajtottunk végre, akkor ebben gyorsabban, könnyebben tudunk navigálni. Egy kis piros nyíl jelzi, hogy épp mi látszik a jobb oldalon. Ha utólag rájövünk, hogy valamelyik lekérdezésre még sincsen szükségünk, az adott elemzést a fa struktúrában is tudjuk törölni a megfelelő sorok törlésével. Az ablak nagyobbik részét az output felület teszi ki, itt jelennek meg a táblázatok, grafikonok, illetve az elvégzett műveletek. Itt mindegyik grafikonba, táblázatba bele tudunk lépni két kattintással, és tudunk onnan másolni (például Excel táblázatba), átírni, átformázni (például a grafikonokat), a jobb egérgomb által kínált opciókkal. Szintén van lehetőség átírni a címsorokat, hogy később könnyebben azonosítani tudjuk, hogy éppen melyik elemzésről van szó.

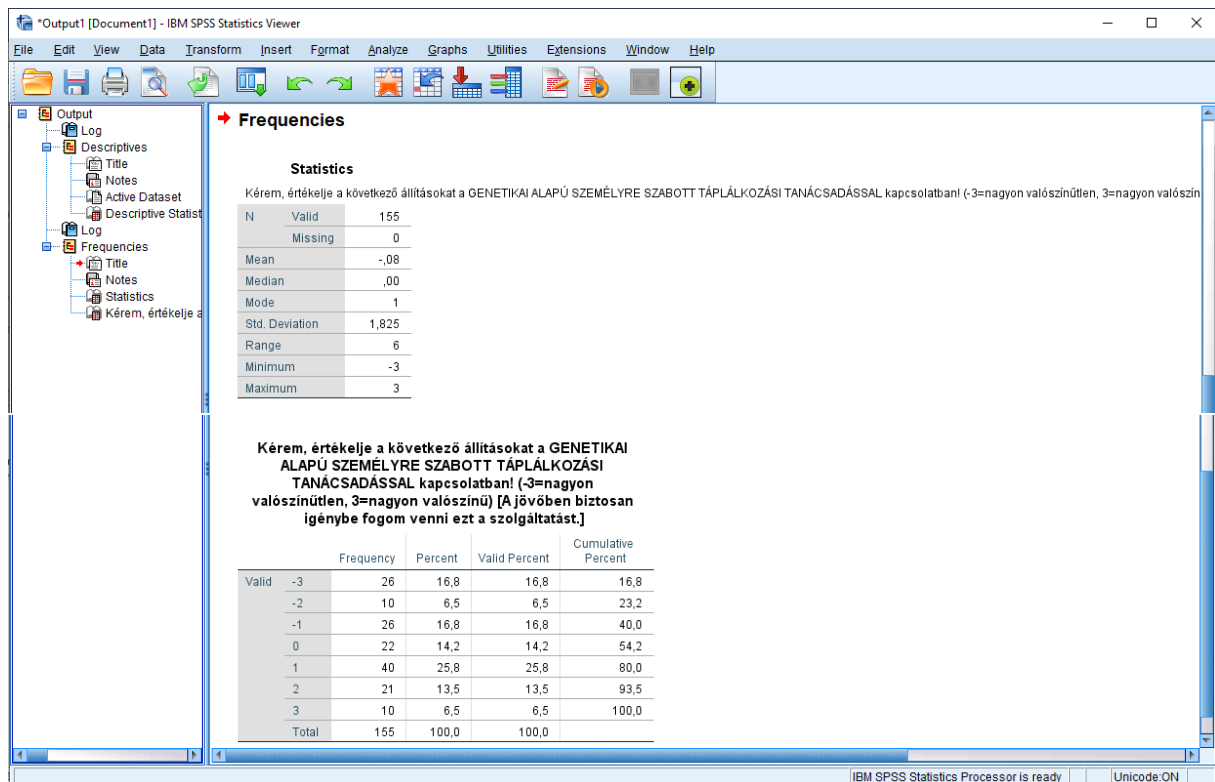
A megnyíló párbeszédablakban az előzőekhez hasonlóan a bal oldali változólistából kiválasztjuk a megfelelő változót, és a nyíl segítségével áttesszük a „Varibale(s):” boxba. A változólista alatt a „Display frequency tables” opció alapértelmezetten ki van választva, ezt így is hagyjuk, hiszen a gyakoriságokat ez fogja megadni.

A „Statistics” gombbal egy következő ablakhoz jutunk, ahol a korábbi leíró statisztikákat be tudjuk jelölni („Mean”, „Std. deviation”, „Range”, „Minimum”, „Maximum”), továbbá a mediánt („Median”) és a móduzt („Mode”) is. Kattintsunk a „Continue”-ra, majd az előző ablakban az „OK”-ra.



11. ábra. Leíró statisztikák lekérése

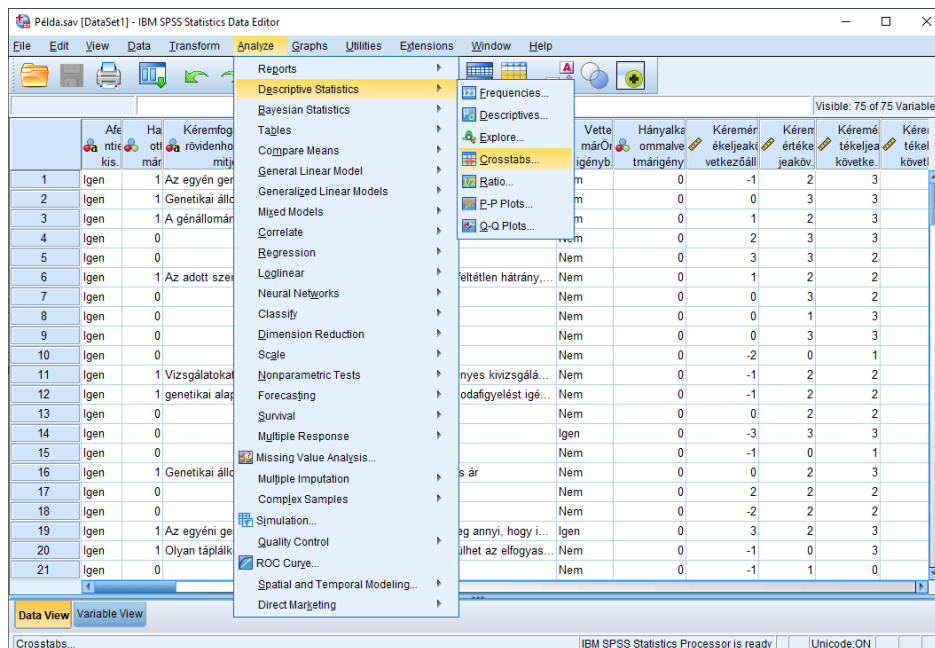
Az outputablakban két táblázatot kaptunk. Az első táblázatból ugyanazokat az értékeket látjuk, mint korábban, illetve a lekért mediánt (0) és móduzt is (1). Továbbá az is látszik, hogy nem volt hiányzó érték (Missing value). A második táblázat a válaszlehetőségek gyakoriságait („Frequency”) mutatja, valamint a válaszok százalékos megoszlását (Percent). Ebből is látszik az, hogy a leggyakoribb érték, azaz a móduz az 1 volt.



12. ábra. Leíró statisztikák

5.2. Modell építés, eredmények értékelése

Ha két nominális vagy ordinális (esetleg kategorizált metrikus) skálán mért változó közötti kapcsolatot szeretnénk vizsgálni, akkor keresztábra-elemzést kell végeznünk. Például a korábbi online felmérés adatai alapján kíváncsiak vagyunk arra, hogy van-e eltérés abban, hogy a válaszadók hallottak-e már a genetikai alapú személyre szabott táplálkozásról (nominális változó) korcsoportok szerint (kategorizált metrikus változó). A keresztábra-elemzés az SPSS-ben az ANALYZE / DESCRIPTIVE STATISTICS / CROSSTABS menüpontban érhető el.



13. ábra. Keresztábra-elemzés indítása

A párbeszédablakban először a sorokban („Row(s):”) szereplő változót kell a bal oldali listából átvinni a nyíl segítségével, majd az oszlopokban („Column(s):”) szereplőt.

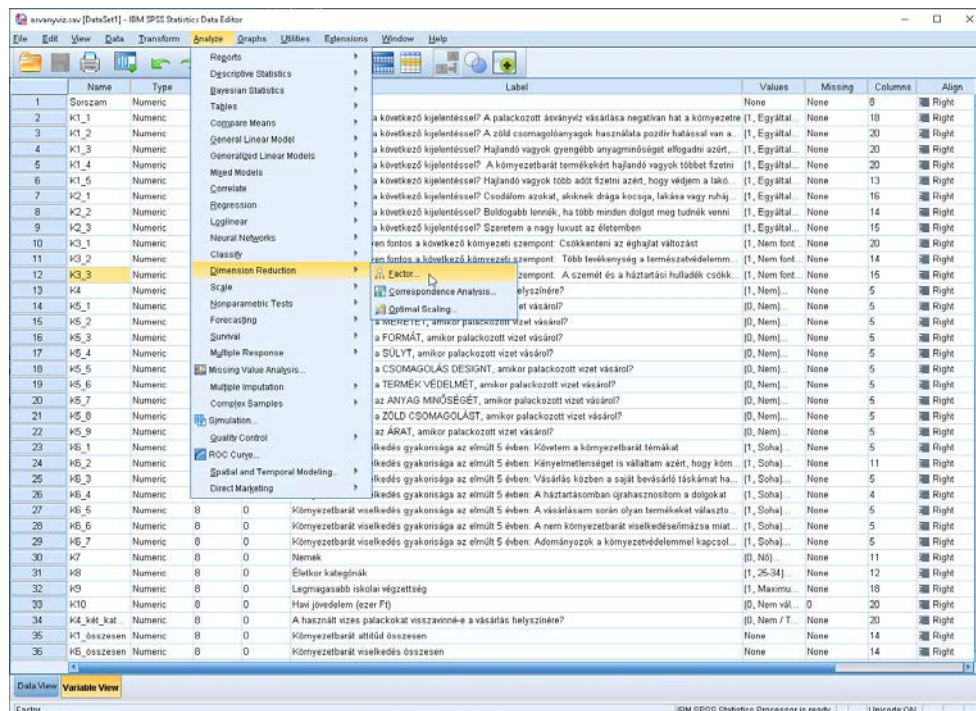
A „Statistics” opciót választva egy új párbeszédablakban kiválaszthatjuk, hogy milyen statisztikákat kérünk le az összefüggés megítéléséhez. A kapcsolat meglétét a chí-négyzet statisztika mutatja („Chi-square”), majd ha bebizonyosodik, hogy van kapcsolat, a kapcsolat erősségére is tudunk következtetni számos mutatószám alapján, melyek közül a keresztábra mérete, szimmetrikussága, illetve a változó mérési skálája alapján tudunk választani (bővebben lásd: SAJTOS – MITEV, 2007). (SAGE, 2019). Ha készen vagyunk, kattintsunk a „Continue”-ra.

6. A faktoranalízis számítása

A faktorelemzés során az általunk feltett kérdések mögött valamilyen látens struktúrát feltételezünk/keresünk. A faktorelemzés logikája az, hogy lehet néhány látens (nem mérhető / közvetlenül nem megfigyelhető) változó és ezek a látens elemek bizonyos kérdésekkel megragadhatók (vagy megpróbáljuk kifejezni ezeket a feltett kérdéseink segítségével). A gyakorlatban sokszor a kutatók a vizsgálni kívánt jelenségekhez kapcsolódó kérdéseket fogalmazznak meg és az elemzés során derül ki, hogy ezek a látens változók kimutathatóak-e (SZÉKELYI – BARNA, 2002). A nagyszámú sztochasztikusan összefüggő eredeti változó helyett, kisszámú un. faktorváltozót képezünk, melyek segítségével az adataink értelmezése és további elemzése leegyszerűsödik, mivel csökkenteni tudjuk a kiinduló változóink számát (ÁCS, 2009). A faktoranalízis egy struktúra feltáró módszer, mivel a függő és független változók nem előre meghatározottak és így a módszer fő célja a változók közötti összefüggések feltárása (SAJTOS – MITEV, 2007). Megállapítható, hogy a kialakított faktorok között minimális a korreláció nagysága. A faktorokkal további elemzéseket végezhetünk, helyettesítve velük az eredeti (nagyszámú) kérdéseket.

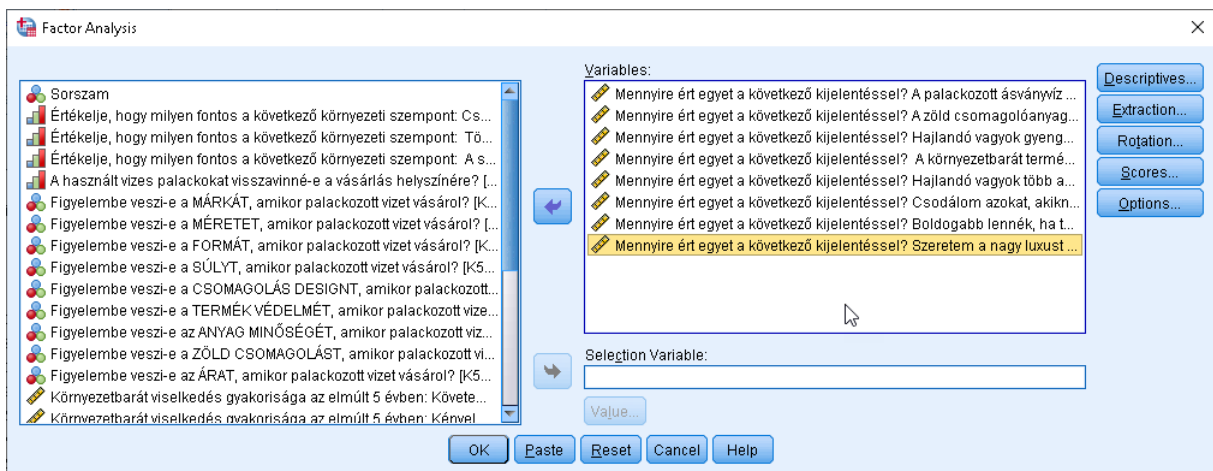
Kérdés: Vizsgáljuk meg – 8 kérdés (K1_1 – K2_3 változók) alapján – faktorelemzés segítségével, hogy a válaszadóknak a környezet iránti és a gazdagság iránti attitűdjei („Mennyire ért egyet a következő kijelentéssel?.....”) mögött van-e valamilyen közös látens struktúra!

A számítást az SPSS ANALYZE / DIMENSION REDUCTION / FACTOR... menüpontjával tudjuk elvégezni (. ábra).



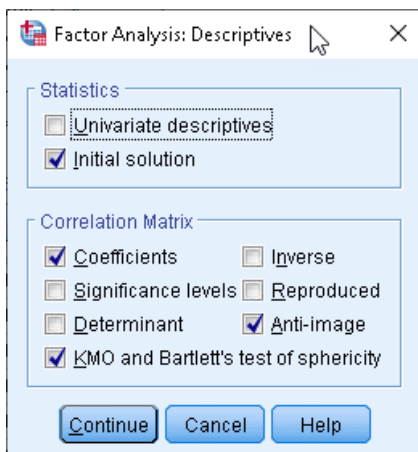
14. ábra. A faktorelemzés indítása

Ezután a megjelenő párbeszédablakban a bal oldali változólistából vigyük át az általunk vizsgálni kívánt 8 változót (K1_1 – K2_3) a jobb oldali változó („Variables”) listába.



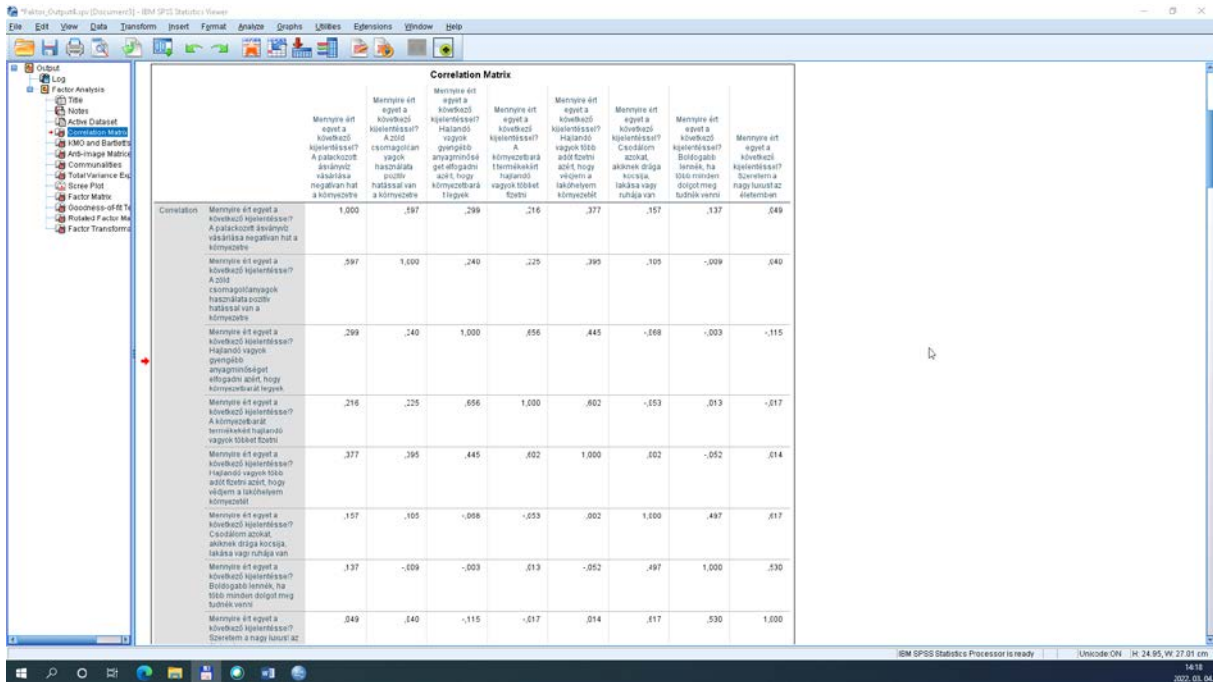
15. ábra: A faktorelemzés beállításai

A „Descriptives” opciót választva egy új párbeszédablakban (. ábra) lehet megvizsgálni, hogy a bevont változók/kérdések alkalmasak-e a faktoranalízis lefuttatására.



16. ábra: A „Descriptive” menüpont lehetséges beállítása

A „Statistics” menürészben egyváltozós leíró statisztikákat („Univariate descriptives”) kérhetünk a vizsgálatba vont változóinkról az alapbeállításon („Initial solution”) felül. A korrelációs mátrix („Correlation Matrix - Coefficients”) kiszámítható a programmal, mivel a változók közötti korreláció bizonyos szintje alapvető ahhoz, hogy faktorelemzést készíthessünk. Ezeket a korrelációs mátrixban szereplő kiszámított értékeket mutatja a . ábra.



17. ábra: A faktorelemzésbe vont változók korrelációs mátrixa

Ez alapján megállapíthatjuk, hogy van kapcsolat a változók között és kijelenthetjük, hogy a kiválasztott változók alkalmasak a faktorelemzés elvégzésére.

A „Descriptives” párbeszédablakban (. ábra) a másik fontos előfeltétel teszteléséhez az „Anti-image” doboz kipipálása (kijelölése) szükséges. Mivel a változók szórásnégyzete felbontható két részre – a megmagyarázott és a nem megmagyarázott szórásnégyzetre –, amit az „Anti-image” kovariancia és korrelációs mátrixok mutatnak. Az „Anti-image” kovarianciamátrix átlótól különböző értékei a varianciának azt a részét fejezik ki, amely a többi változótól független. Ezért érdemes megvizsgálni ezeket az értékeket, és ha több mint háromnegyedik 0,09-nél kisebb az arra utal, hogy van mögöttes kapcsolat a vizsgálatba vont változóink között SAJTOS – MITEV, 2007). Az „Anti-image” korrelációs mátrix főátlójában szereplő értékei (0-1 közötti tartományban mozognak) az un. MSA „Measure of Sampling Adequacy” értékek, amelyek azt mutatják meg, hogy egy változó milyen szoros kapcsolatban van a többi – faktorelemzésben szereplő – változóval. Azokat a változókat, amelyeknek az MSA értéke 0,5 alatti, ki kell zárni a további elemzésből, mivel nem illeszkednek megfelelően a faktorszerkezetbe. A mi elemzésünkben a változóink MSA értékei 0,624 és 0,743 között voltak (. ábra).

életemben									
Anti-image Correlation	Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre	,640 ^a	-,520	-,162	,102	-,176	-,087	-,170	,093
	Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	-,520	,663 ^a	-,008	,002	-,176	-,054	,128	-,041
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyek	-,162	-,008	,694 ^a	-,550	-,023	-,003	-,043	,136
	Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	,102	,002	-,550	,624 ^a	-,463	,071	-,063	-,055
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	-,176	-,176	-,023	-,463	,743 ^a	-,001	,129	-,073
	Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van	-,087	-,054	-,003	,071	-,001	,690 ^a	-,243	-,476
	Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni	-,170	,128	-,043	-,063	,129	-,243	,692 ^a	-,344
	Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben	,093	-,041	,136	-,055	-,073	-,476	-,344	,640 ^a

a. Measures of Sampling Adequacy(MSA)

18. ábra: A faktorelemzésbe vont változók „Anti-image Correlation” mátrixa

A . ábrán bemutatjuk, hogy a „Descriptives” opció választása után be lehet állítani a párbeszédablakban a „KMO and Bartlett’s test of sphericity” teszteket is. Ennek eredményeként ki tudjuk számítani az MSA értékek átlagát az összes változóra, aminek a Kaiser-Meyer-Olkin kritérium (KMO) a neve. 0,5 alatti érték esetén adataink alkalmatlanok a faktoranalízis elvégzésére. 0,9 feletti érték esetén tökéletesnek tekinthetők az adataink az elemzésre. A Bartlett-teszt nullhipotézise az, hogy a változóink nem korrelálnak egymással. Ami azt jelenti, hogy a vizsgálat során a korrelációs mátrix főátlón kívüli elemei nem térnek el szignifikánsan a nullától. Célunk, hogy ezt a nullhipotézist elvessük.

A következő ábrán (. ábra) a „KMO and Bartlett’s Test” eredményeit tüntettük fel, amely alapján megállapítható, hogy a kiszámított KMO érték 0,672. Ez azt jelenti, hogy a változóink közepesen alkalmasak a faktoranalízisre. A Bartlett-próba nullhipotézisét is elvethetjük, mivel a teszt szignifikancia értéke („Sig.”) kisebb, mint 0,05, Ez azt jelenti, hogy a változóink alkalmasak a faktorelemzésre, mivel van korreláció közöttük.

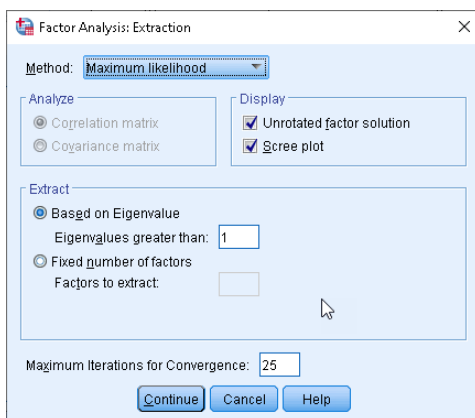
KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,672
Bartlett's Test of Sphericity	Approx. Chi-Square	2435,466
	df	28
	Sig.	,000

19. ábra: A KMO kritérium és a Bartlett-próba értékei

A következő lépésben a lehetséges faktorok számának meghatározását kell elvégeznünk és ehhez több eljárás közül választhatunk. Az „Extraction” gomb kiválasztása után lehetséges az eltérő faktorelemzési módszerek közül kiválasztanunk a számunkra legmegfelelőbbet (részletek SAJTOS – MITEV, 2007. 253-254. o.). A példánk esetében – mögöttes faktorstruktúrát feltételezve – a „Maximum likelihood” eljárást választottuk ki (. ábra). Ez az eljárás a megfigyelt korrelációs mátrixból indul ki és olyan becslést ad, amely ezt a korrelációs mátrixot a legnagyobb valószínűség mellett alakíthatta ki (*normális eloszlást feltételez az alkalmazása!*) Ha a célunk az lett volna, hogy a sokaságban lévő legkevesebb információt veszítsük el, akkor érdemes lett volna a „Principal components” módszert kiválasztani. Mivel a főkomponens-elemzés a teljes varianciát figyelembe veszi.

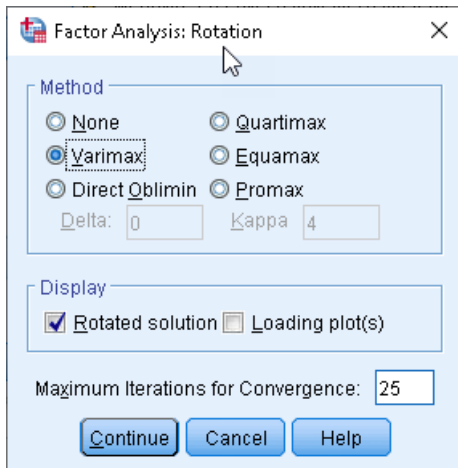
Az „Extract” dobozban a faktorok számát lehet meghatározni. Ha a sajátérték „Eigenvalue” alapján akarunk dönteni (*Kaiser-kritérium*), akkor az alapbeállítást használjuk („Based on Eigenvalue”). Ebben az esetben a program abból indul ki, hogy egy faktornak több információt kell jelentenie, mint egy eredeti változó. Ha ismerjük a lehetséges faktorok számát választhatjuk a másik beállítást is („Fixed number of factors”). Ekkor elég a négyzetben megadnunk a kívánt faktorok számát.



20. ábra: A faktorelemzés módszerének kiválasztása és a lehetséges faktorok számának meghatározása

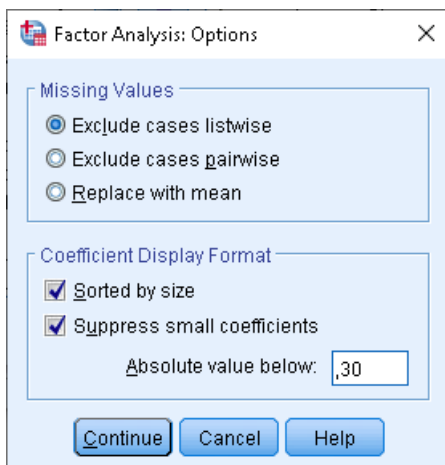
A faktorok számának meghatározásához további lehetőséget jelent a „Scree plot” doboz kipipálása. A „*Könyökszabály*” ábra a sajátértékeket („Eigenvalue”) fogja megmutatni a lehetséges faktorok száma szerint. Ez alapján vizuálisan is könnyen eldönthetjük, hogy mennyi legyen a faktorok száma, mivel ahol a vonal meredeksége csökken és kezd ellaposodni, az adja meg a kutató számára a javasolt faktorszámot.

Ezután a „Rotation” almenüt (. ábra) választhatjuk ki, amely lehetővé teszi számunkra, hogy a kiszámított faktorok tengelyeit elforgatva könnyebben értelmezhető eredményt kapjunk. A forgatás során a program a megmagyarázott varianciát egyenletesebben fogja elosztani a faktorok között. Javasolt a „Varimax” módszer kiválasztása, mivel így – az eredetihez képest – stabilabban lesznek szétválasztva – és jobban is magyarázhatók – a faktorok. Ez egy ortogonális forgatási eljárás és az így elkülönített faktorok egymással nem korrelálnak. A „Display” dobozban elég a „Rotated solution” feliratot választani, mivel most nem szükséges az elforgatott eredmények térbeli „Loading plot(s)” ábrázolása.



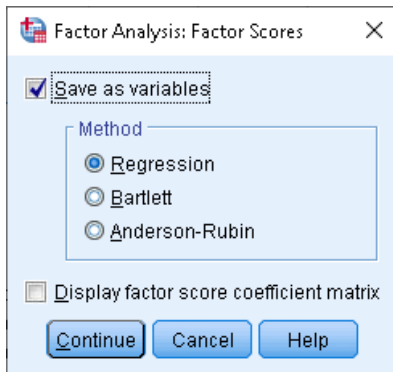
21. ábra: A faktorok elforgatásának („Rotation”) beállítása

Az „Options” almenü kiválasztásával meghatározhatjuk a közölt faktorok értelmezhetőségének kritériumait (. ábra). A „Sorted by size” doboz kijelölésével a faktorsúly–mátrixban („Rotated Factor Matrix”) csökkenő sorrendben lesznek súlyok szerint a változóink. A „Suppress small coefficients” dobozt kiválasztva csak a megadott faktorsúly értéknél (0,3) nagyobb értékeket tünteti fel a program a két faktorsúly–mátrixban.



22. ábra: A kialakított faktorok faktorsúly–mátrixának beállításai

Ha az eredményeinket el szeretnénk menteni a „Scores” gomb kijelölésével kell folytatnunk. A „Save as variables” dobozt kipipálva a „Regression” módszer aktív lesz (. ábra). Így a kialakított faktorszámnak megfelelő számú faktorváltozót hoz létre az SPSS az adatbázisunkban. Minden megfigyelt személyhez faktoronként külön értékek fognak tartozni. Ezeket az új változókat a későbbi elemzésekben (pl. regressziószámítása vagy klaszterezés) fel fogjuk tudni használni.



23. ábra: A kialakított faktorok elmentése

Ha végeztünk minden beállítással, az „OK” gomb megnyomásával tudjuk elindítani a faktorelemzést. Ezután az outputablakban számos táblázat fog megjelenni, melyek közül eddig már az első hármat elemeztük. A következő ábrán (. ábra) a faktorelemzésben szereplő változók un. kommunalitás értékeit láthatjuk. A „Maximum likelihood” faktorelemzési eljárás esetében az „Initial” oszlopban a többszörös korrelációs együtthatók négyzetei (R^2) szerepelnek. Ezek az értékek mutatják meg azt, hogy egy változó szórását/varianciáját a többi elemzésbe vont változó mekkora mértékben magyarázza. Ha valamelyik kérdés esetében nagyon kicsi értéket kapunk, azt a kérdést a későbbiekben – nagy valószínűséggel – ki kell zárunk a faktorelemzésből.

Fel kell hívnunk a figyelmet arra, hogy ha az „Extraction / Method” menüben (. ábra) módszerként a főkomponens elemzés („Principal components”) módszert választottuk volna ki, minden változó esetében a kommunalitás „Initial” értéke 1 lenne.

Az „Extraction” oszlopban a végső kommunalitások szerepelnek, és ez azt mutatja meg, hogy a kialakított 3 faktor az egyes változók szórásának/varianciájának hány százalékát magyarázza meg. Ha valamelyik változó esetében 0,25 alatti érték szerepel, annak a változónak nincs elég magyarázó ereje és ezt ki kell hagyni a további faktorelemzésből. Az adatokat vizsgálva megállapíthatjuk, hogy mind a 8 elemzésbe vont változó megfelelő kommunalitás értékkel rendelkezik ahhoz, hogy a faktoranalízis további vizsgálataiban szerepeljen.

Communalities^a

	Initial	Extraction
Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre	,428	,666
Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	,401	,542
Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyek	,471	,473
Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	,562	,999
Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	,458	,474
Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van	,438	,590
Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni	,362	,428
Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben	,466	,665

Extraction Method: Maximum Likelihood.

24. ábra: A változók kommunalitásának táblázata

A következő ábrán a teljes és a faktorok által magyarázott variancia értékei láthatóak a kezdeti, a faktoranalízis utáni és a forogtatást követő értékek feltüntetésével. Mivel a főkomponens és a faktorelemzés az SPSS-ben közös pontból indul ki, ezért a maximum likelihood eljárásnál az első három oszlop („Initial Eigenvalues”) azt a kiindulási helyzetet mutatja, mintha egy főkomponenselemzést hajtottunk volna végre (SZÉKELYI – BARNA, 2002). Ez alapján megállapítható, hogy a 8 bevont változóból képzett 8 főkomponens a teljes variancia 100%-át magyarázza és ebből az első három főkomponens sajátértéke lenne 1 felett (az általuk megmagyarázott variancia 73,77% lenne). Most mi a többi oszlop tartalmát fogjuk elemezni. Az „Extraction Sums of Squared Loadings” részben lévő adatok a faktorok által reprezentált információ tartalmát (a „Total” oszlopokban van a sajátérték feltüntetve) mutatják. A 8 változó által képviselt lehetséges összes információ 8 egység, amiből az első faktor 1,896, a második 1,753 és a harmadik 1,188 egységnyi jelent. A „% of Variance” oszlopban az egyes faktorok által magyarázott variancia nagyságát olvashatjuk le (23,7%; 21,9% és 14,9%) csökkenő sorrendben. A „Cumulative %” oszlop összesítve mutatja az egyes faktorok összeadása utáni magyarázott varianciákat (pl. 23,7 + 21,9 = 45,6). Minket az érdekel, hogy a kialakított faktorok által magyarázott rész összesített % értéke 60% felett legyen. Ez azt jelenti, hogy sikerült olyan faktorstruktúrát találnunk, ami „*elég nagy részét*” magyarázza az eredeti 8 változó által képviselt összes információnak. A legtöbb esetben az így kapott eredeti faktorstruktúra nem igazán jól magyarázható szakmailag, ezért érdemes elforgatnunk a faktorok tengelyeit (Ennek a módszerét állítottuk be a „Rotation” almenüben „Varimax” eljárásnéven). Az ábrán megfigyelhetjük a „Rotation Sums of Squared Loadings” részben, hogy a forogtatás

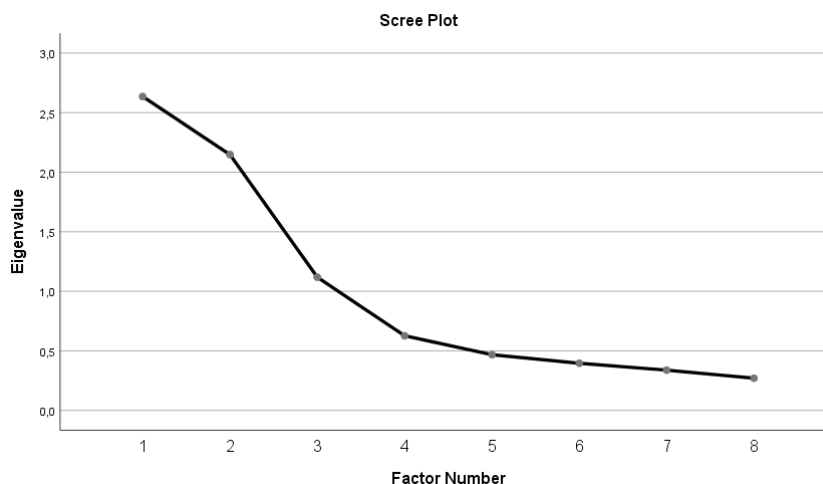
természetesen nincs hatással az elforgatott faktorok által megmagyarázott összes variancia nagyságára (60,456% maradt). Ismert az is, hogy a forgatás nem változtatja meg a modell illeszkedést és az egyes változók kommunalitásait sem. Arra törekszik, hogy egyenletesebben ossza el a magyarázott varianciát a faktorok között. A varimax eljárással az egy változóhoz tartozó faktorsúlyok négyzetösszegeit maximalizáljuk. Ennek eredménye az, hogy – a forgatás előtti állapothoz képest – minden változót megpróbál még inkább egy faktorhoz hozzárendelni, amivel növeli a faktorok magyarázhatóságát.

Factor	Total Variance Explained								
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2,636	32,946	32,946	1,896	23,696	23,696	1,835	22,937	22,937
2	2,148	26,846	59,792	1,753	21,910	45,605	1,681	21,009	43,946
3	1,118	13,979	73,770	1,188	14,851	60,456	1,321	16,510	60,456
4	,627	7,832	81,602						
5	,468	5,844	87,447						
6	,396	4,947	92,394						
7	,338	4,225	96,619						
8	,270	3,381	100,000						

Extraction Method: Maximum Likelihood.

25. ábra: A teljes és a faktorok által magyarázott variancia értékei

A faktoranalízis beállításainál már említésre került, hogy a lehetséges faktorok számáról grafikus ábra („Scree-plot” vagy „Könyök-ábra”) segítségével is dönthetünk. A . ábráról könnyen leolvasható, hogy az ajánlott faktorok száma 3, mivel ez után már ellaposodik az ábránk. A „könyök-szabály” alapján ezért kijelenthető, hogy az ajánlott faktorszám maximálisan 3 lehet, mivel a negyedik faktor esetében a sajátérték már nagyon kicsi lenne (0,627).



26. ábra: A javasolható faktorok számát mutató „Könyök-ábra”

Az output file-ban a kiszámított eredmények között a következő táblázatok a forgatás nélküli („Factor Matrix”) és a forgatás utáni („Rotated Factor Matrix”) faktorsúlyokat bemutató táblázatok. Ha főkomponens elemzést végzünk ezek a táblázatok „Component Matrix” illetve „Rotated Component Matrix” néven szerepelnek az outputban.

A „Factor Matrix”-ban – előjellel ellátva – a faktorsúlyok vannak feltüntetve, amelyek azt jelzik, hogy az egyes változók milyen mértékben (milyen súllyal) vesznek részt a különböző faktorok kialakításában. A faktorsúly az eredeti változó és a faktor közötti korreláció szorosságát mutatja, aminek négyzetes értéke kifejezi, hogy a faktor a változó varianciájának hány százalékát magyarázza. Érdeemes megemlíteni, hogy a nagyon kicsi (kisebb, mint 0,3) faktorsúly azt jelenti, hogy a változó nem kapcsolható össze az adott faktoralal. Ha egyik faktor esetében sem éri el az adott változó faktorsúlya ezt a kritikus értéket, abban az esetben a változót ki kell zárni a további faktorelemzésből. Ha pedig több faktoralon is nagyobb érték szerepel, mint 0,3 akkor a legnagyobb értéket kell összevetni a többi érték kétszeresével. Az egyes faktorokban feltüntetett értékek alapján megpróbálhatjuk a különböző faktorokat elnevezni. Felhívjuk a figyelmet arra, hogy ez nem minden esetben egyszerű. Általában a forgatás segít abban, hogy tisztábban lássuk a változóink és a faktorok közötti kapcsolatokat, ezzel megkönnyíti a kutató helyzetét abban, hogy jelentést tudjon adni az egyes elforgatott faktoroknak. A következő ábrán csak a 0,3 feletti faktorsúlyokat láthatjuk, mivel ezt az értéket állítottuk be az „Options” almenü „Suppress small coefficients” dobozában (. ábra). Megfigyelhetjük, hogy az egyes faktorokon belül a változók súlyai csökkenő sorrendbe rendezettek, mivel előzetesen az „Options” menüben ezt a parancsot „Sorted by size” is beállítottuk. A faktorsúlyok abszolút értékben mutatják a változók értelmezhetőségét a faktorok viszonylatában. Példánkban az első faktornál a legjelentősebb hatása a „környezetbarát termékekért való magasabb fizetési hajlandóságnak” volt és ezt követte a másik két („anyagminőségre” és az „adófizetésre” vonatkozó) változó. A következő faktort a „szeretem a nagy luxust....”, a „csodálom azokat, akiknek drága kocsija,” és a „boldogabb lennék, ha több minden dolgot....” változók alakították ki. A harmadik faktorba csak két kérdés a „palackozott ásványvíz vásárlása negatívan hat a környezetre” és a „zöld csomagolóanyagok használata pozitív...” tartozott.

Rotated Factor Matrix^a

	Factor		
	1	2	3
Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	,998		
Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyenek	,648		
Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	,586		
Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben		,815	
Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van		,756	
Mennyire ért egyet a következő kijelentéssel? Bológabb lennék, ha több minden dolgot meg tudnék venni		,653	
Mennyire ért egyet a következő kijelentéssel? A palackozott ásvíz vásárlása negatívan hat a környezetre			,790
Mennyire ért egyet a következő kijelentéssel? Azöld csomagolóanyagok használata pozitív hatással van a környezetre			,710

Extraction Method: Maximum Likelihood.
 Rotation Method: Varimax with Kaiser Normalization.
 a. Rotation converged in 5 iterations.

27. ábra: A rotálás utáni faktorsúlymátrix

Mivel előzetesen a „Scores” gomb kijelölésével a „Save as variables” dobozt kipipáltuk az SPSS az eredményeinket (a 3 új faktort) az adatbázisban elmentjük, azaz 3 új változót (FAC1_1 – FAC3_1) hozott létre. A további elemzések megkönnyítésére a „Data Editor”-ban a „Variable view” ablakban az új változók „Label” celláiba lehet beírni az elnevezéseket.

A rotált faktorsúlymátrix alapján megpróbálhatjuk elnevezni is ezt a három új változót:

- Környezetvédelem iránti fizetési hajlandóság (1. faktor)
- Kényelem és pénz orientáltság (2. faktor)
- Környezet iránti pozitív attitűd (3. faktor)

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
25	K6_3	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: Vásárlás közben a saját bevásárló tászkámat ha...	{1, Soha}...	None	5	Right
26	K6_4	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: A háztartásomban újrahasznosított dolgokat	{1, Soha}...	None	4	Right
27	K6_5	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: A vásárlásaim során olyan termékeket választot...	{1, Soha}...	None	5	Right
28	K6_6	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: A nem környezetbarát viselkedése/írásmása miat...	{1, Soha}...	None	5	Right
29	K6_7	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: Adományozok a környezetvédelemmel kapcsol...	{1, Soha}...	None	5	Right
30	K7	Numeric	8	0	Nemek	{0, N6}...	None	11	Right
31	K8	Numeric	8	0	Életkor kategóriák	{1, 25-34}...	None	12	Right
32	K9	Numeric	8	0	Legmagasabb iskolai végzettség	{1, Maximu...}	None	18	Right
33	K10	Numeric	8	0	Havi jövedelem (ezer Ft)	{0, Nem vál...}	0	20	Right
34	K4_két_kat...	Numeric	8	0	A használt vízes palackokat visszavinné-e a vásárlás helyszínére?	{0, Nem / T...}	None	20	Right
35	K1_összesen	Numeric	8	0	Környezetbarát attitűd összesen	None	None	14	Right
36	K6_összesen	Numeric	8	0	Környezetbarát viselkedés összesen	None	None	14	Right
37	FAC1_1	Numeric	11	5	környezetvédelem iránti fizetési hajlandóság	None	None	13	Right
38	FAC2_1	Numeric	11	5	kényelem és pénz orientáltság	None	None	13	Right
39	FAC3_1	Numeric	11	5	környezet iránti pozitív attitűd	None	None	13	Right
40									
41									

28. ábra: Az új faktorok és elnevezéseik az adatbázisban

7. Klaszterelemzési technikák

A marketingkutatás egyik leggyakoribb célja a fogyasztók, a megkérdezettek szegmentációja, vagyis olyan csoportok képzése, amelyek belül a válaszadók viszonylag homogének, ugyanakkor egymástól jól elkülönülők (SIMON, 2006). A leggyakrabban használt csoportképző ismérvek a fogyasztókat jellemző szocio-demográfiai jellemzők, de lehetséges a választott termékek tulajdonságai alapján is szegmentálni a fogyasztókat.

Általában a klaszterelemzés fő célja, hogy a megfigyelési egységeket viszonylag homogén csoportokba sorolja a kiválasztott változók alapján úgy, hogy az adott csoportba tartozó megfigyelési egységek hasonlítsanak egymásra, de különbözzenek más csoportok tagjaitól (FÜSTÖS et al., 2004; HAJDU, 2003). Ebben az anyagban a skálatípusú adatok elemzésére alkalmas klaszterezési eljárást mutatjuk be.

A klaszterelemzés menete:

- A probléma megfogalmazása
- A távolság mérték kiválasztása
- A klasztermódszer kiválasztása
- Döntés a klaszterek számáról
- A klaszterek értelmezése és jellemzése
- A klaszterelemzés érvényességének ellenőrzése

A probléma megfogalmazása során kiválasztjuk a csoportképzés alapjául szolgáló változókat. A nem megfelelő változó bevonása ronthat a bevonás nélküli jó csoportosításon. A változók kiválasztása történhet korábbi kutatások alapján, elméleti megfontolások vagy a kutató saját döntése, intuíciója alapján (BACKHAUS et al., 2003).

A távolság mérték kiválasztása során az egységek közötti hasonlóságot azok közötti távolsággal mérjük. Különböző távolságmértékek használata (pl. Euklideszi, Csebisev, Manhattan, Pearson) eltérő megoldásokhoz vezethet, így célszerű különböző mértékeket használni, úgy elvégezni az elemzést, majd az eredményeket összehasonlítani.

A klasztermódszer kiválasztása esetében is több lehetőség adódik, mivel a klaszterezési eljárások lehetnek hierarchikusak és nem hierarchikusak. A hierarchikus módszereket két csoportra bonthatjuk, mint agglomeratív (összevonó) és divízív (felosztó) eljárások. A klaszterek közötti távolság képzésének eljárásai a következők: egyszerű lánc módszer, teljes lánc módszer, centroid módszer, medián módszer, csoportátlag módszer és ward módszer.

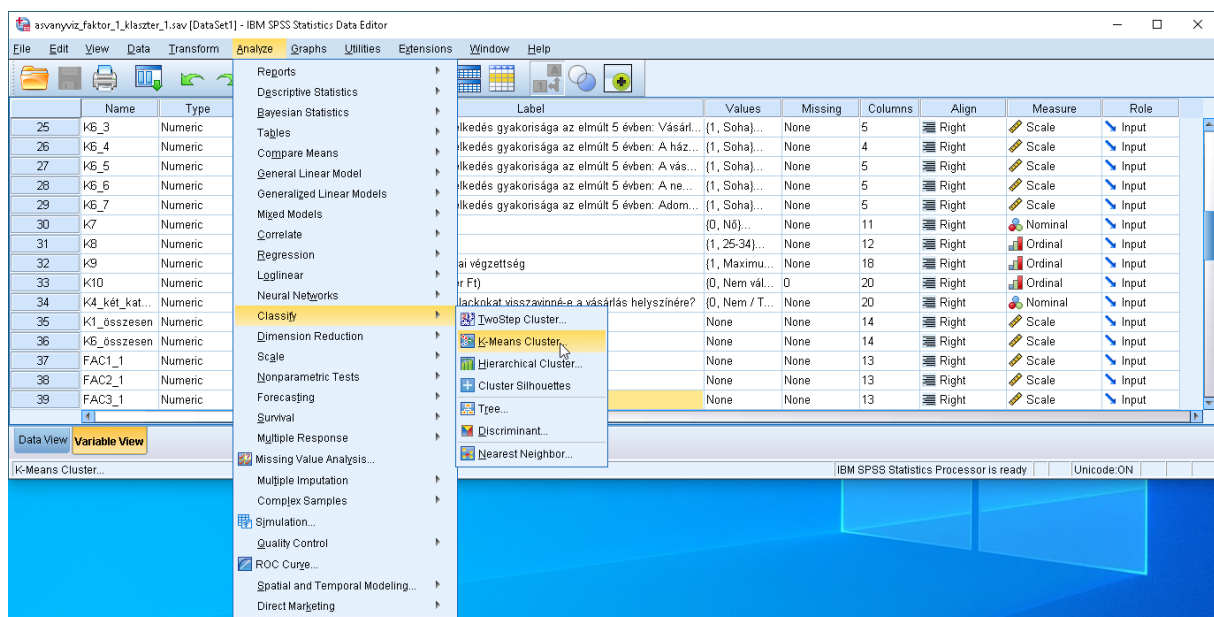
A nem hierarchikus klaszterképzés eredményeként diszjunkt klasztereket állítunk elő. A klaszterek száma az egyes módszerek során alakul ki, más módszereknél előre meg kell adjuk ezt az értéket. A hierarchikus és nem hierarchikus módszerek között az az egyik lényeges különbség, hogy míg a hierarchikus eljárások esetén, ha két elem egy csoportba kerül, akkor a továbbiakban már együtt is marad, addig a nem hierarchikus eljárások esetében lehet, hogy később külön csoportba kerülnek át.

A klaszterelemzés érvényességének ellenőrzése során számos különböző lépést tehetünk. Ezek a következők lehetnek:

- Más távolság mértéket alkalmazunk és az így kapott eredményeket összehasonlítjuk.
- Különböző klasztereljárásokkal dolgozunk.
- Az adatainkat véletlenszerűen két almintára bontjuk, és mindkettőre elvégezzük az elemzést.
- Véletlenszerűen elhagyunk változókat, és csökkentett változószámmal végezzük el újra az elemzést.
- Nem hierarchikus elemzéseknél futtassuk az elemzést az esetek különböző sorrendjével, mígnem stabilizálódik a megoldás.

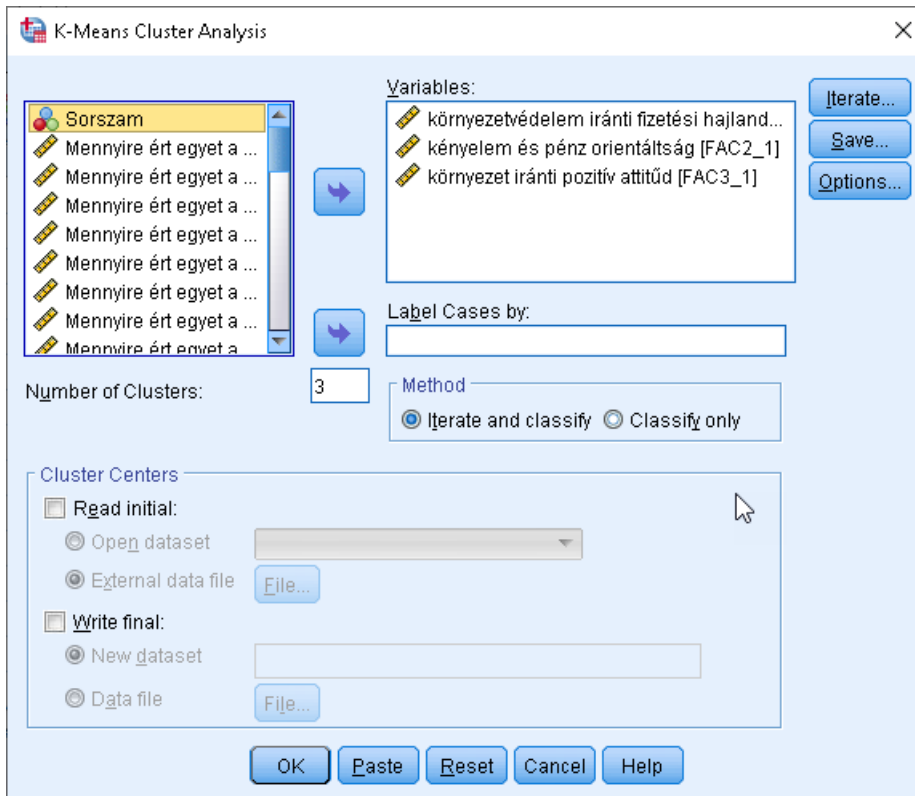
Kérdés: Vizsgáljuk meg – a környezetvédelemmel kapcsolatos 3 faktorváltozó (FAC1_1 – FAC3_1 változók) alapján – klaszterelemzés segítségével, hogy a válaszadók milyen egymástól jól elkülönülő csoportokba sorolhatók!

Az általunk bemutatott példa esetében a nem hierarchikus klaszterképzést fogjuk alkalmazni, mivel előzetes szakirodalmi adatok alapján általában három klaszterbe sorolhatók az ásványvizet vásárlók. A számítást az SPSS ANALYZE / CLASSIFY / K-MEANS CLUSTER... menüpontjával tudjuk elvégezni (. ábra).



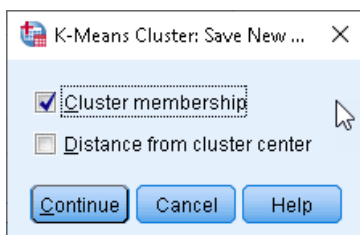
29. ábra: A klaszteranalízis kiválasztása

A K-közép klaszteranalízis kiválasztása után meg kell adnunk azokat a változókat, amelyek alapján a válaszadókat külön csoportokba szeretnénk rendezni (. ábra). Ezért a baloldali ablakban ki kell jelölnünk a három faktorváltozót (FAC1_1 – FAC3_1) és át kell mozgatnunk a jobboldali változó (Variables) ablakba. A szükséges csoportok számát a „Number of Clusters:” dobozban állíthatjuk be. Mivel előzetesen volt információnk a lehetséges csoportok számáról, ezért 3-at írjuk be értéként.



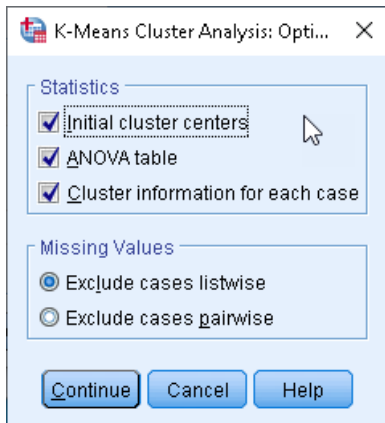
30. ábra: A 3 faktorváltozó kiválasztása és a klaszterszám megadása

Ezt követően lehetőségünk lesz arra, hogy a „Save” menü kiválasztásával beállítsuk azt, hogy a kiszámított klasztereket elakarjuk-e menteni („Cluster membership”) azért, hogy később elemzéseket futtassunk a kialakított klaszterekbe sorolt válaszadókkal.



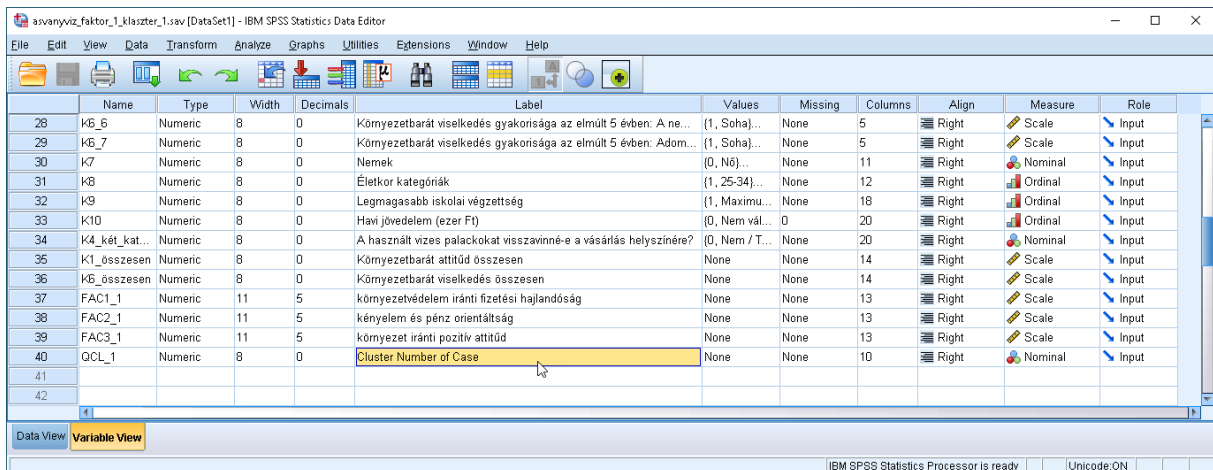
31. ábra: A klaszterváltozó létrehozása az adatbázisban

Az „Options” menüben be lehet állítani további eredményeket, ha bekattintjuk az alábbi cellákat.



32. ábra: Az Option almenü beállításai

Lefuttatva a fenti beállításokkal a klaszterezést, az SPSS változói között megjelenik egy új klaszterváltozó is.



33. ábra: Az új klaszterek egyedi értékeit tartalmazó klaszterváltozó

A kiindulási klaszterértékeket nagyon jól nyomon lehet követni a következő ábrán feltüntetett értékek alapján.

	Initial Cluster Centers		
	Cluster		
	1	2	3
környezetvédelem iránti fizetési hajlandóság	-1,71696	-2,94564	1,47346
kényelem és pénz orientáltság	-1,91199	1,46550	1,99534
környezet iránti pozitív attitűd	-,82473	1,53577	,24664

34. ábra: A kiindulási klaszterek közötti távolságok értékei

A program 8 iterációban alakította ki a klasztereket.

Iteration History^a

Change in Cluster Centers

Iteration	1	2	3
1	1,956	1,915	1,813
2	,239	,371	,140
3	,282	,271	,129
4	,286	,258	,103
5	,095	,072	,014
6	,026	,023	,000
7	,000	,017	,019
8	,000	,000	,000

35. ábra: Az iterációnként változások eredményei

Meg lehet vizsgálni egyenként azt, hogy ki melyik klaszterbe tartozik és milyen távol van a hozzá leghasonlóbbtól.

Cluster Membership		
Case Number	Cluster	Distance
1	2	3,144
2	2	3,144
3	2	3,144
4	2	3,144
5	2	3,144
6	2	3,144
7	2	2,596
8	2	2,596
9	2	2,596
10	2	2,922
11	2	2,922
12	2	2,922
13	2	2,640
14	2	2,640

36. ábra: A klaszterszám megadása és a közöttük lévő távolságok

A végső klaszterközéppontokat érdemes elemezni szakmai következtetések levonása miatt is.

	Final Cluster Centers		
	Cluster		
	1	2	3
▶ környezetvédelem iránti fizetési hajlandóság	,53263	-1,08863	,66831
kényelem és pénz orientáltság	-,88709	,12897	,75047
környezet iránti pozitív attitűd	,05718	,00750	-,06587

37. ábra: A 3 faktorváltozó központjainak klaszterenkénti értékei

A program számszerűsíteni tudja a végső klaszter-középpontok közötti távolságokat is.

Distances between Final Cluster Centers

Cluster	1	2	3
1		1,914	1,648
2	1,914		1,865
3	1,648	1,865	

38. ábra: A klaszter-középpontok közötti távolság megadása

Varianciaanalízissel össze tudjuk hasonlítani a különböző csoportátlagok közötti különbségeket. Az alábbi ábrából leolvasható, hogy az utolsó változó (környezet iránti pozitív attitűd) megítélése nem szignifikáns azaz a válaszadók kb. egységesen ítélik ezt meg.

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
környezetvédelem iránti fizetési hajlandóság	297,351	2	,342	903	869,201	,000
kényelem és pénz orientáltság	199,909	2	,364	903	549,839	,000
környezet iránti pozitív attitűd	1,119	2	,759	903	1,474	,230

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

39. ábra: Variacionálízis eredményei

Fontos információ az is, hogy ez egyes klaszterekben hány válaszadó van, mivel nekik hasonló ízlésük és várhatóan hasonló döntéseik lesznek.

Number of Cases in each Cluster

Cluster	1	293,000
	2	322,000
	3	291,000
Valid		906,000
Missing		,000

40. ábra: A 3 klaszter elemszámainak megadása

8. Korrespondancia-analízis

A korrespondencia analízis egy olyan exploratív technika, mely az asszociációs kapcsolatok vizuális elemzése érdekében egy gyakorisági tábla adatait grafikus ábrává konvertálja. A tábla sorai mint pontok az oszlopok terében, az oszlopai pedig mint pontok a sorok terében kerülnek ábrázolásra egy redukált egy-, két-, vagy háromdimenziós altérben. A pontfelhők helyzetének az elemzése révén nyílik lehetőség a sorok és az oszlopok közötti asszociációk feltárására. Sor vagy oszlop kategóriák kombinációját is reprezentálhatja.

Egyszerű korrespondencia analízisről (CA) beszélünk, ha a sorok v.s. oszlopok pontfelhőt ábrázoljuk. Ez a helyzet például, mikor csak két változó alkotja a gyakorisági táblát. Ezzel szemben többszörös – multiple - korrespondencia analízist (MCA) alkalmazunk, ha kettőnél több változó szerepel elemzésünkben és valamennyi változó valamennyi kategóriáját önálló pontként ábrázoljuk (HAJDU, 2011).

Technikailag az MCA nem más, mint egy speciális gyakorisági táblán, az indikátor mátrixon végzett CA. Az indikátor mátrixban a megfigyelési egységek (háztartások, személyek) képezik a sorokat, és valamennyi változó valamennyi kategóriája egy-egy önálló oszlopot alkot: a megfelelő belső cella gyakorisága 1, ha az illető megfigyelés az illető oszlophoz, mint kategóriához tartozik, a gyakoriság egyébként zéró. Matematikailag a korrespondencia analízis az asszociáció Pearson-féle χ^2 mértékét bontja komponensekre hasonló módon, mint azt a főkomponens analízis a totális varianciával teszi. Az eljárás a sorokat (oszlopokat) a megoszlásaikból képzett, redukált dimenziójú térben, mesterséges CA koordináták alapján ábrázolja. Itt a tengelyeket úgy definiáljuk, hogy rendre csökkenő százalékos mértékben (sorrendben) járuljanak hozzá a χ^2 statisztikához. Mikor az első, vagy az első kettő mesterséges tengely a teljes asszociáció meghatározó (80-90% körüli vagy több) hányadát magyarázza, a gyakorisági tábla síkbeli ábrája a pontfelhő eredeti konfigurációját hűen, kevés információvesztéssel tükrözi.

Az asszociáció mértéke egy n számú megfigyelést csoportosító kontingencia táblában adott. A CA a korrespondencia mátrixot elemzi, melynek általános eleme az i sor és a j oszlop együttes $p_{ij} = f_{ij}/n$ relatív gyakoriság ahol f_{ij} az előfordulási gyakoriság.

Az s_i sorösszesenek és az o_j oszlopösszesenek szintén az n mintaméret százalékában kerülnek kifejezésre. A relatív gyakoriságok feltételes, adott soron belüli sorozatát sorszerkezetnek, adott oszlopon belüli sorozatát pedig oszlopszerkezetnek nevezzük. A sorszerkezeteket az oszlopok által kifeszített J -dimenziós tér, míg az oszlopszerkezeteket a sorok által kifeszített I -dimenziós tér pontjaiként kezeljük, két pontfelhőt definiálva. A peremmegoszlásokat, mint a megfelelő tengelyekhez rendelt súlyokat az s és az o vektorokba foglaljuk.

A sorszerkezetek és az oszlopszerkezetek a korrespondencia mátrix elemeihez a következő módon kapcsolódnak.

Megfigyelésenként összegezve mindkét oldalát, a tábla alapján látszik, hogy az s_i és o_j peremek a belső oszlop- és sorszerkezetek súlyozott átlagai, súlyként a peremek másik körét alkalmazva. Így az o_j perem egyben a sorprofilok centroidja, míg az s_i perem az oszlopprofilok centroidja. Az asszociáció teljes hiányát a sorok és az oszlopok között tehát úgy is definiálhatjuk, hogy

adott pontfelhő valamennyi pontja egybeesik egymással, és így a saját centroidjával is. Az asszociáció természetére tehát a pontfelhők szóródásából következtethetünk. A szóródás mérésére a Pearson- χ^2 statisztikát alkalmazzuk, melynek megszokott megnevezése a CA terminológiában: totális inercia. Ahol s_{ij} az $[i,j]$ cella függetlenség esetén várható relatív gyakorisága és a standardizált korrespondencia gyakoriság. Zérótól különböző g_{ij} előjele pozitív, vagy negatív asszociációt jelez az i sor és a j oszlop között. Az azonosság alapján a teljes inercia akár a sorfelhő, akár az oszlopfelhő pontjainak súlyozott, többváltozós varianciája. ahol $s_{ijc} = s_{ij} - o_j$ és $o_{ijc} = o_{ij} - s_i$ a centrált sor- és oszlopprofilok, melyek centroidja mindig az origó. A korrespondencia analízis a pontfelhőket mesterséges CA koordináták alapján ábrázolja. A koordináták értékét úgy határozzuk meg, hogy minden egyes egymást követő ($k=1,2,\dots,K$) koordináta-tengely egyre inkább csökkenő hányadban járul hozzá a teljes inerciához. Mikor például az első két tengely az inercia döntő hányadát magyarázza, akkor a pontfelhők konfigurációja a CA koordináták síkbeli ábrázolásában is valóságos marad. A centrált sorszerkezetek ábrázolására x , a centrált oszlopprofilok ábrázolására pedig y koordinátákat számítunk, melyeket az $X(I,K)$ és az $Y(J,K)$ mátrixokba foglalunk. A CA tengelyek lehetséges maximális száma $K = \min\{I-1, J-1\}$ mivel a relatív gyakoriságok összege adott profilon belül 1.

Fontos mozzanat, hogy a sorkoordináta a standardizált oszlopprofilok súlyozott átlaga, míg az oszlopprofilok a standardizált sorkoordináták súlyozott átlaga, súlyként rendre a sor-, illetve oszlopprofilokat alkalmazva. Az oszlopok és a sorok egymásba való átvitele lényegében a koordináták duális skálázását jelenti. Az x_{ik} és y_{jk} akkor van közel egymáshoz, mikor a j oszlop nagy s_{ij} súllyal szerepel az i sorprofilban vagy az i sor szerepel nagy o_{ij} súllyal a j oszlopprofilban. Ebben az esetben egy nagy sorkoordináta a k tengelyen szükségszerűen szintén nagy oszlopprofilját eredményez ugyanezen a tengelyen. Közös koordináta rendszerben ábrázolva a sorok és az oszlopok pontfelhőjét tehát, azon sorok és oszlopok kerülnek várhatóan közel egymáshoz, amelyek között szoros az asszociáció mértéke. Ez teszi lehetővé a kapcsolatok feltárását, mert pontok közötti távolságot csak pontfelhőn belül értelmezünk, pontfelhők között nem. A pontfelhők közötti korrespondenciát tehát a duális skálázás elve alapján ítéljük meg.

Az eljárás a Data Reduction főmenüben a Correspondence Analysis almenüjéből végezhető el, ahol először a sorváltozókat, majd az oszlopprofilokat kell megadni.

```
CORRESPONDENCE TABLE=K3_1(1 3) BY TSC_5674(1 3)
/DIMENSIONS=2
/MEASURE=CHISQ
/STANDARDIZE=RCMEAN
/NORMALIZATION=SYMMETRICAL
/PRINT=TABLE RPOINTS CPOINTS
/PLOT=NDIM(1,MAX) BIPLLOT(20).
```

41. ábra: A parancskódok bemutatása

Ezután mindkét ismérvet definiálni kell, a benne szereplő ismérvváltozatok számának segítségével. A sorváltozót (kérdés kódok) 1-től 3-ig, míg az oszlopváltozót (klasztercsoportok száma) szintén 1-től 3-ig. A többi beállításon nem változtatva futtassuk le az elemzést. A keletkező eredmények a következő ábrákon követhetjük nyomon.

Correspondence Table

Értékelje, hogy milyen fontos a következő környezeti szempont: Csökkenti az éghajlat változást	3 klaszter			Active Margin
	1	2	3	
Nem fontos	141	54	153	348
Közepesen fontos	114	93	132	339
Nagyon fontos	102	51	66	219
Active Margin	357	198	351	906

42. ábra: A korrespondancia táblázat

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	,128	,016			,617	,617	,032	-,013
2	,101	,010			,383	1,000	,033	
Total		,026	23,907	,000 ^a	1,000	1,000		

a. 4 degrees of freedom

43. ábra: A statisztikai különbséget bemutató táblázat

Overview Row Points^a

Értékelje, hogy milyen fontos a következő környezeti szempont: Csökkenti az éghajlat változást	Mass	Score in Dimension			Inertia	Contribution				
		1	2			Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
						1	2	1	2	Total
Nem fontos	,384	,444	-,077	,010	,593	,022	,977	,023	1,000	
Közepesen fontos	,374	-,347	-,271	,009	,353	,273	,675	,325	1,000	
Nagyon fontos	,242	-,169	,541	,008	,054	,704	,110	,890	1,000	
Active Total	1,000			,026	1,000	1,000				

a. Symmetrical normalization

44. ábra: A sorokat kialakító változó értékeinek eredményei

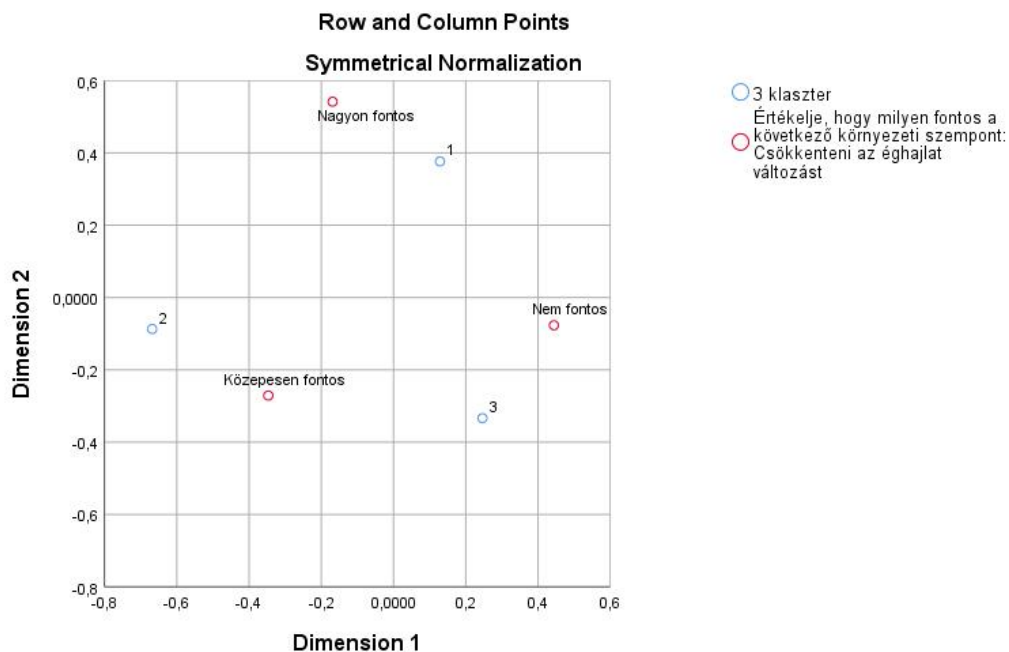
Overview Column Points^a

3 klaszter	Mass	Score in Dimension			Inertia	Contribution				
		1	2	1		2	Of Point to Inertia of Dimension		Of Dimension to Inertia of Point	
							1	2	1	2
1	,394	,129	,376	,006	,051	,555	,129	,871	1,000	
2	,219	-,668	-,087	,013	,765	,017	,987	,013	1,000	
3	,387	,246	-,334	,007	,184	,429	,409	,591	1,000	
Active Total	1,000			,026	1,000	1,000				

a. Symmetrical normalization

45. ábra: Az oszlopokat kialakító változó értékeinek eredményei

A keletkező eredmények közül a grafikus ábrázolást vizsgálva, láthatóvá válnak az összetartozó értékek is.



46. ábra: A korrespondancia-analízis grafikus megjelenítése

Az ábra alapján jól elkülönülnek az egyes klaszterek és az éghajlatváltozás fontosságának szintjei. Megfigyelhetjük, hogy az első klaszter szorosan összefügg azokkal, akik azt választották, hogy számukra nagyon fontos az éghajlatváltozás, mint környezeti szempont. A második klaszterbe azok különültek el, akiknek közepesen fontos volt ez a szempont. Az utolsó klaszterbeliek számára elhanyagolható az éghajlatváltozás fontosságának figyelembevétele.

9. A diszkriminancia analízis bemutatása

A diszkriminanciaanalízis az az eljárás, ami a következő kérdésre ad választ: milyen csoportba tartoznak célcsoportunk alanyai? A csoportot itt igen tágan értelmezhetjük. Tulajdonképpen bármilyen csoportosítás lehetséges, amelyet értelmesen vizsgálni lehet. A diszkriminanciaanalízis olyan adatelemzési módszer, amelyet kategóriába tartozás előrejelzésére lehet használni. Alkalmazásával alacsony mérési szintű függő változót magas mérési szintű független változók segítségével magyarázunk.

Azt vizsgáljuk, hogy a csoporthoz tartozás mekkora százalékban becsülhető a független változókkal (pl. azt, hogy valaki drog függő vagy nem, mekkora mértékben magyarázza az életkor, iskolázottság, stb.). Az eljárás során, akár csak a lineáris regresszió esetében olyan egyenest keresünk, amely a legjobban szétválasztja az elemzendő csoportokat.

A diszkriminanciaanalízis során tehát azt a problémát járjuk körül, hogyan lehet az emberek egyes csoportjait valamilyen vizsgált jellemzők alapján szétválasztani, az egyes csoportokat azonosítani, valamint a csoporttagságokat az előbb említett vizsgált jellemzők alapján előrejelezni.

Példa: megfelelő-e egy jelölt az állásra vagy sem, hajlamos-e depresszióra vagy nem, visszaeső bűnöző lesz-e a személy vagy a büntetés után feladja a bűnözői karrierjét.

Hogy a csoporttagságokat előre tudjuk jelezni, valamilyen jellemző vagy képesség vizsgálata szükséges, amely vagy amelyek alapján a becslésünket meg tudjuk tenni. Például a jelölt alkalmasságának vizsgálatához tesztelhetjük a munkavégzés szempontjából kulcsfontosságú képességeit. A depresszióra való hajlam vizsgálatánál megnézhetjük a családbeli előfordulását a depresszióknak, nézhetjük a személyt érő stressz mennyiségét, stb. Mindezek, és az ehhez hasonló vizsgálatok végrehajtására alkalmas eljárás a diszkriminancia-analízis.

Lineáris diszkriminancia-analízis a statisztikában, minta-felismerésben és gépi tanulásban használt módszer, amely a független változók olyan lineáris kombinációját képes megtalálni, amely a függő változó alapján kialakított csoportokat a lehető legjobban megkülönbözteti (diszkriminálja). A diszkriminancia-analízis szorosan kapcsolódik a varianciaanalízishez és a regresszióanalízishez, amelyek úgyszintén egy függő változót igyekeznek kifejezni más változók lineáris kombinációjaként. Azonban míg e két utóbbi eljárásnál a függő változó folytonos változó, addig a diszkriminancia-analízisnél ez kategorikus változó. Ellentétben a varianciaanalízissel, ahol kategorikus független változókkal magyarázzuk a folytonos függő változókat, a diszkriminancia-analízis esetében folytonos független változók mellett kategorikus függő változókat használunk. A regresszióanalízis pedig abban különbözik a diszkriminancia elemzéstől, hogy esetében magas mérési szintű folytonos függő és független változók szerepelhetnek az elemzésben. Kategorikus független változók esetén az ekvivalens eljárás a megfelelési diszkriminancia-analízis.

A diszkriminancia-analízis központi lépése a diszkrimináló függvény(ek) kiszámítása. A szükséges diszkrimináló függvények száma úgy számítható ki, hogy a függő változó lehetséges kimeneteleinek száma -1 és a független változók száma közül a kisebbet kell venni. Ha tehát

két csoportunk van és két folytonos független változó (ú.n. prediktor változó), akkor egy diszkrimináló függvényünk lesz. Ellenben négy csoport és két folytonos prediktor változó esetén kettő. A diszkrimináló függvény általános képlete az alábbi:

$$D_j = d_0j + d_1jx_1 + d_2jx_2 + \dots + D_kjx_k ,$$

ahol j az adott diszkrimináló függvény sorszám, az x_i -k a mért független változók, d_0 konstans, a d_{ij} a x_i mért változó j -edik diszkrimináló függvényhez tartozó együtthatója. A függvény akkor optimális, ha a függő változó által meghatározott csoportok közötti külső négyzetösszeg és a csoportokon belüli négyzetösszeg hányadosa maximális. A négyzetösszeg a varianciaanalízisben használt heterogenitást kifejező átlagos négyzetes eltérést jelenti. A Wilks-féle lambda, amely a csoportokon belüli átlagos négyzetes eltérés és teljes átlagos eltérés aránya, megadja a diszkrimináló függvény jószágát. Értéke egy 0 és 1 közötti szám. 0-hoz közelítő értékek esetén a csoportokon belüli variabilitás kicsi, ami azt is jelzi, hogy függvényünk jól diszkriminál a csoportok között. Ezzel szemben az 1 közeli érték azt jelzi, hogy a csoporton belüli négyzetösszeg közel áll a teljes négyzetösszeghez, és így a csoportok közötti négyzetösszeg kicsi, ami azt mutatja, hogy a függvény kevésbé tudja a csoportokat jól megkülönböztetni.

A marketing területén a diszkriminancia-analízis felhasználható arra, hogy egy empirikusan összegyűjtött adatsor alapján meghatározzuk, mely faktorok különítik el a vásárlókat vagy termékeket két vagy több csoportra. Manapság erre a célra a logisztikus regresszió szélesebb körben használt eljárás. A diszkriminancia-analízis használata a marketingben az alábbiak szerint foglalható össze:

A kutatási kérdés megfogalmazása és az adatok összegyűjtése. Először meg kell határozni azokat a kitüntetett jellemzőket, amelyek alapján a vásárlók értékelik a terméket. Ezután kvantitatív marketing technikákkal (pl. kérdőíves felmérésekkel) a potenciális vásárlók egy csoportján fel kell mérni az adott termék minden lényeges sajátosságát. Ez az adatgyűjtési szakasz általában marketing szakemberekre hárul. A kérdőívben 1-től 5-ig (vagy 7-ig ill. 10-ig) kell értékelni a terméket több (átlagosan 5 és 20 közötti) sajátosság tekintetében. Ezek az alábbiak lehetnek: a használat egyszerűsége, súly, pontosság, tartósság, szín, ár vagy méret, stb... Az értékelési szempontok nagyban függenek a termék mibenlététől. Ugyanezen szempontok alapján értékelik a potenciális vásárlók a többi terméket. Az adatokat lekódozzák, beviszik egy statisztikai programba, mint pl. R, SPSS vagy a SAS. (Eddig a lépésig a teendők megegyeznek a faktor-analízissel.)

A diszkrimináló függvények kiszámítása és statisztikai szignifikancia és validitás meghatározása. A diszkrimináló függvények olyan függvények, amelyek a legnagyobb különbséget produkálják a kategorikus függő változó által definiált csoportok között. Az első lépés a megfelelő diszkriminancia-analízis eljárás kiválasztása. A közvetlen módszer esetében a független változók egyszerre kerülnek be az eljárásba, és így számítjuk ki a diszkrimináló függvényeket. A lépcső-módszer esetében egymást követően kerülnek be a független változók a modellbe. A kétmodelles eljárás alkalmazandó abban az esetben, ha függő változónak két szintje van. A többszörös diszkriminancia-analízis pedig akkor szükséges, ha három vagy több függő változónk van. Az SPSS-ben a Wilks-féle lambda, míg a SAS programban az F-statisztika tájékoztat bennünket a szignifikanciáról. A diszkrimináló függvények alapján a

statisztikai programok lehetőséget biztosítanak arra, hogy a függő változó csoportjaiba tartozó személyeket vagy termékeket újra klasszifikáljuk, így tesztelve, hogy a diszkriminancia-analízisben kialakított magyarázóter mennyire hatékony.

Kétszintű függő változó esetén felrajzolhatóak az eredmények egy kétdimenziós diagramon. A termékek vagy csoportok távolsága jelzi, hogy azok mennyire különbözőek. A diagram tengelyeit a kutatóknak kell elnevezniük. Az ábra értelmezése sokszor erősen szubjektív.

```

DISCRIMINANT
/GROUPS=CLU3_1(1 3)
/VARIABLES=K1_1 K1_2 K1_3 K1_4 K1_5 K2_1 K2_2 K2_3
/ANALYSIS ALL
/SAVE=CLASS
/PRIORS EQUAL
/STATISTICS=MEAN STDDEV UNIVF BOXM CORR TABLE CROSSVALID
/PLOT=COMBINED
/CLASSIFY=NONMISSING POOLED.

```

47. ábra: A Diszkriminancia analízis lefuttatásához szükséges parancssor bemutatása

Az első táblázat „Group Statistics” az elemzésbe bevont összes változó csoportok szerinti és összesített átlagát, szórását, súlyát mutatja.

Group Statistics

3 klaszter (Hierarchikus)		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
1	Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre	4,47	,608	201	201,000
	Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	4,05	,904	201	201,000
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyek	3,93	,636	201	201,000
	Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	3,76	,430	201	201,000
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	3,95	1,021	201	201,000
	Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van	3,90	,837	201	201,000

	Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni	4,13	,723	201	201,000
	Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben	3,40	,996	201	201,000
2	Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre	3,43	,899	417	417,000
	Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	3,00	1,116	417	417,000
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyek	4,15	,593	417	417,000
	Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	4,28	,460	417	417,000
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	3,72	,932	417	417,000
	Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van	2,50	1,181	417	417,000
	Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni	3,32	1,194	417	417,000
	Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben	2,24	,969	417	417,000
3	Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre	3,42	1,145	288	288,000
	Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	2,69	,875	288	288,000
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyek	3,21	,969	288	288,000
	Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	2,58	,703	288	288,000
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	2,58	,915	288	288,000

	Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van	2,98	1,022	288	288,000
	Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni	3,55	1,137	288	288,000
	Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben	2,51	,875	288	288,000
Total	Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre	3,66	1,028	906	906,000
	Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	3,13	1,119	906	906,000
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyek	3,80	,849	906	906,000
	Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	3,62	,918	906	906,000
	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	3,41	1,107	906	906,000
	Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van	2,96	1,192	906	906,000
	Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni	3,57	1,131	906	906,000
	Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben	2,59	1,048	906	906,000

48. ábra: A változók leíró statisztikáját bemutató táblázat

Az ezt követő ábrán feltüntettük azt a táblázatot, amelyben meg tudjuk vizsgálni, hogy a független változók milyen mértékben járulnak hozzá a létrejövő függvényhez. A változók szignifikáns voltának tesztelésére az F-érték mellett, a Wilks'-Lambda statisztika is szerepel.

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre	,823	97,280	2	903	,000
Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	,794	116,818	2	903	,000
Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyek	,761	142,100	2	903	,000
Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	,351	836,640	2	903	,000
Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	,731	166,162	2	903	,000
Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van	,793	117,580	2	903	,000
Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni	,923	37,718	2	903	,000
Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben	,814	103,238	2	903	,000

49. ábra: A független változók hozzájárulásainak eredményei

Látható, hogy minden változónak szignifikáns hatása van. A Wilks'-Lambda értéke 0 és 1 közé eső értékek, melyek közül a mindig a nullához közeli értékekhez tartozó változóknak (*Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni*) van a legjelentősebb hatása a diszkriminancia függvényre.

A következő ábrán egy alapfelvetés tesztelése történik meg. A Pooled Within-Groups Matrices táblázat eredményei segítségével a multikollinearitást teszteljük.

Pooled Within-Groups Matrices

	Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre	Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyenek	Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van	Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni	Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben
Correlation	1,000	,510	,331	,332	,341	-,023	,031	-,153
Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre								
Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	,510	1,000	,191	,183	,300	-,074	-,133	-,160
Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyenek	,331	,191	1,000	,510	,278	-,023	,022	-,119
Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	,332	,183	,510	1,000	,438	,103	,113	,078

50. ábra: A multikollinearitás tesztelésének adatai

A következő fontos táblázat (Eigenvalues), mely során először kapunk információt a keletkező függvényről. Megfigyelhető, hogy két függvény keletkezett. A függvények számát megállapíthatjuk, ha a csoportok száma, illetve a független változók száma közül a kevesebbikből egyet kivonunk. A két függvény fontosságának megállapításában, a sajátérték segíti a kutatót. A táblázat sajátértékei és magyarázott variancia értékei alapján az első függvény lesz fontosabb számunkra. A kanonikus korreláció (0,834) azt jelenti, hogy az adott függvény igen számottevő részt magyaráz a teljes varianciából. A kapott értékek négyzete megmutatja, hogy a függő változó varianciájának, hány százalékát magyarázzák a független változók csoportja (70%).

Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2,276 ^a	76,6	76,6	,834
2	,694 ^a	23,4	100,0	,640

a. First 2 canonical discriminant functions were used in the analysis.

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	,180	1541,596	16	,000
2	,590	474,277	7	,000

51. ábra: A sajátérték és a Wilks'-Lambda eredmények bemutatása

A megjelenő Wilks'-Lambda táblázat a függvények szignifikanciájának tesztelését végzi. Láthatóan mindkét függvény szignifikáns, de az első hatása jelentősebb.

A korrelációs együttható mátrixa (Structure Matrix) hasonlóan értelmezendő, mint a faktoranalízisnél a Component Matrix, hiszen a független változók és a diszkriminancia függvények közti, csoportonkénti átlagolt (Pooled within groups) Pearson féle lineáris korrelációk.

Structure Matrix

	Function	
	1	2
Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni	,874*	,405
Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyek	,353*	,211
Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre	,036	,607*
Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre	-,042	,552*
Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben	-,129	,524*
Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy ruhája van	-,176	,523*
Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a lakóhelyem környezetét	,313	,458*
Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni	-,089	,307*

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

*. Largest absolute correlation between each variable and any discriminant function

52. ábra: A struktúra mátrix eredményeinek bemutatása

Ez alapján az első függvény a „környezetbarát termékekért hajlandó vagyok többet fizetni” és a „hajlandó vagyok gyengébb anyagminőséget elfogadni azért, hogy környezetbarát legyek”

szempontokat, míg a második az összes többit foglalja magában, mely alapján a kutató el tudja nevezni a dimenziókat (hasonlóan a faktorelemzéshez).

A következő ábrán a csoportok középpontértékeit mutatjuk be. Megállapíthatjuk, hogy második és a harmadik csoport (klaszter) magas értékekkel rendelkezik az első dimenzióban, míg az első klaszter magas értékei a második dimenzió mentén jelentkeznek. Ezeket a koordinátákat fogja a program felhasználni a grafikus megjelenítéskor is.

Functions at Group Centroids

3 klaszter (Hierarchikus)	Function	
	1	2
1	-,436	1,539
2	1,526	-,318
3	-1,905	-,614

Unstandardized canonical discriminant functions evaluated at group means

53. ábra: A csoportok középpont értékei

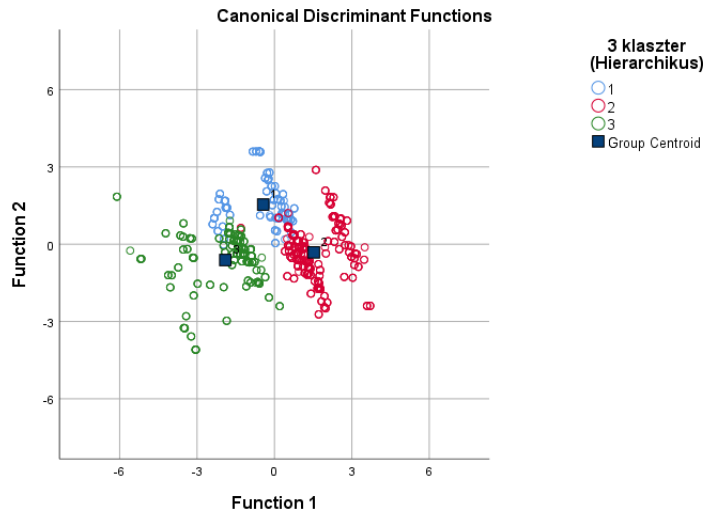
Az elemzés folytatásaként a program kiszámolja az un. klasszifikációs statisztikai mutatókat. Az egyik ilyen eredmény a következő ábrán látható. Látszik, hogy a csoportokba kerülés esélye 33,3 százalék volt.

Prior Probabilities for Groups

3 klaszter (Hierarchikus)	Prior	Cases Used in Analysis	
		Unweighted	Weighted
1	,333	201	201,000
2	,333	417	417,000
3	,333	288	288,000
Total	1,000	906	906,000

54. ábra: A klasszifikációs statisztika eredményei

A következőkben a grafikus ábrázolás történik, ahol a tengelyek maguk a függvények (dimenziók). Az ábra az analízisbe bevont egyedek értékeit és a centrumközéppontokat ábrázolja.



55. ábra: A független változók hozzájárulásainak eredményei

A helyesen kategorizált csoporttagságok arányát a klasszifikációs eredmények elnevezésű táblázatban (Classification Results) láthatjuk.

Classification Results^{a,c}

		Predicted Group Membership				
		3 klaszter (Hierarchikus)	1	2	3	Total
Original	Count	1	173	7	21	201
		2	18	399	0	417
		3	4	3	281	288
	%	1	86,1	3,5	10,4	100,0
		2	4,3	95,7	,0	100,0
		3	1,4	1,0	97,6	100,0
Cross-validated ^b	Count	1	173	7	21	201
		2	18	399	0	417
		3	4	3	281	288
	%	1	86,1	3,5	10,4	100,0
		2	4,3	95,7	,0	100,0
		3	1,4	1,0	97,6	100,0

a. 94,2% of original grouped cases correctly classified.

b. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

c. 94,2% of cross-validated grouped cases correctly classified.

56. ábra: A klasszifikációs eredmények

A táblázat alján láthatjuk, hogy a modell 94,2%-ban tudta helyesen kategorizálni a megadott független változó mentén.

Felhasznált irodalom

- Tamus Antaln  (2009): A marketing kutat s gyakorlata. K roly R bert Kutat  – Oktat  K zhaszn  Nonprofit Kft., Gy ngy s ISBN: 9789639941083
- Sz kelyi M. – Barna I. (2002): T l l k szlet az SPSS-hez. T bbv ltoz s elemz si technik kr l t rsadalomkutat k sz m ra. 4. kiad s TYPOTEX Elektronikus Kiad  Kft., ISBN:9789632790121
- Huzsvai L. (2012): STATISZTIKA gazdas gelemz k r sz re Excel  s R alkalmaz sok. SENECA BOOKS, 173. o. ISBN 978-963-08-5016-2
- Huzsvai L. – Balogh, P. (2015): Line ris modellek az R-ben. SENECA BOOKS, 150. o. ISBN 978-615-801172-0-6
- Hair J.F., Anderson R.E., Tatham R.L., Grablowsky B.J. (2014): MULTIVARIATE DATA ANALYSIS: Pearson New International Edition. 7th edition, Pearson Publisher, UK, ISBN: 9781292021904
-  cs P. (2009): Sporttudom nyai kutat sok m dszertana. P cs 291. o. ISBN: 9789636422752
- Simon J. (2006): A klaszterelemz s alkalmaz si lehet ségei a marketingkutat sban. Statisztikai Szemle 85(7) 627-650. o.
- F st s et al. (2004): Alakfelismer s. UMK. Budapest.
- Hajdu O. (2003): T bbv ltoz s statisztikai sz m t sok. KSH. Budapest.
- Backhaus K., Erichson B., Plinke W., Weiber R. (2016): Multivariate Analysemethoden. 10. Auflage, Springer Gabler, Berlin, Heidelberg 647. o. ISBN 978-3-662-46075-7
- Fishbein, M. (1967): Attitude and the Prediction of Behaviour. In: Fishbein, M. (Ed.): Readings in Attitude Theory and Measurement. Wiley, New York
- Green, M. C. – Keegan, W. J. (2020): Global Marketing. Global Edition, 10th Edition, Pearson, Boston
- Kotler, P. – Armstrong, G (2020): Principles of Marketing. Global Edition, 18/E, Pearson, Harlow, etc.
- Kotler, P. – Keller, K. L. (2017): Marketingmenedzsment. Akad miai Kiad , Budapest
- Malhotra, N. K. (2010): Marketing Research – An Applied Orientation. 6th Edition, Prentice Hall, Boston etc.
- Malhotra, N. K. – Simon, J. (2016): Marketingkutat s. Akad miai Kiad , Budapest
- Petrovics, P. – G czi-Papp, R. (2021): Keresztt bla elemz s az SPSS-ben. Oktat si seg danyag, Miskolci Egyetem, https://gtk.uni-miskolc.hu/files/12362/10_SPSS+kereszt%3%A1bla.pdf (let lt s ideje: 2021.06.31.)
- Sajtos, L – Mitev. A. (2007): SPSS kutat si  s adatelemz si k zik nyv. Alinea Kiad , Budapest
- SAGE (2019): Learn to Use the Eta Coefficient Test in R With Data From the NIOSH Quality of Worklife Survey (2014). SAGE Publications, Ltd., <https://methods.sagepub.com/base/download/DatasetStudentGuide/eta-coefficient-niosh-qwl-2014-r> (let lt s ideje: 2021.06.10.)
- SPSS (2021): IBM SPSS Statistics honlapja: <https://www.ibm.com/products/spss-statistics> (let lt s ideje: 2021.06.18.)

- Baráth CS.-né - Ittész A. - Ugrósdly GY.:1996. Biometria: módszertan és a MINITAB programcsomag alkalmazása. Mezőgazda Kiadó, Budapest
- Cochran, W. G., and G. M. Cox 1957. Experimental Designs. 2d. ed. New York: Wiley.
- Dunn, O. J., and V. A. Clark. 1987. Applied Statistics: Analysis of Variance and Regression. 2d. ed. New York: Wiley.
- Freund, J. and Perles, B. "A New Look at Quartiles of Ungrouped Data." American Stat. 41, 200-203, 1987.
- Hoaglin, D.; Mosteller, F.; and Tukey, J. (Ed.). Understanding Robust and Exploratory Data Analysis. New York: Wiley, pp. 39, 54, 62, 223, 1983.
- Hunyadi L., Vita L.: Statisztika I. Aula Kiadó, Budapest, 2008. 1-348. o.
- Hunyadi L., Vita L.: Statisztikai képletek és táblázatok (oktatási segédlet), Aula Kiadó, Budapest, 2008. 1-51. o.
- John, P.W.M. 1971. Statistical Design and Analysis of Experiments. New York: MacMillan.
- J.P. Marques de Sá (2007): Applied Statistics, Using SPSS, STATISTICA, MATLAB and R. Springer-Verlag Berlin Heidelberg, ISBN 978-3-540-71971-7.
- Katona Tamás - Lengyel Imre (szerk.): Statisztikai ismerettár - fogalmak, képletek, módszerek Excel és SPSS alkalmazásokkal. JATEPress, Szeged, 1999. 121 oldal, (közgazdász, jogász, kísérletes és társadalomtudomány)
- Kenney, J. F. and Keeping, E. S. "Quartiles." §3.3 in Mathematics of Statistics, Pt. 1, 3rd ed. Princeton, NJ: Van Nostrand, pp. 35-37, 1962.
- KIRK, R. E. 1982 Experimental Design. 2d ed. Monterey, CA: Brooks/Cole Publishing Co.
- Lothar Sachs: 1985. Statisztikai módszerek. Mezőgazdasági Kiadó, Budapest
- Mendenhall, W. and Sincich, T. L. Statistics for Engineering and the Sciences, 4th ed. Prentice-Hall, 1995.
- Mérő L. (1992): A pszichológiai skálázás matematikai alapjai. Tankönyvkiadó, Budapest, 1992. 15. o.
- Moksony Ferenc: Gondolatok és adatok: Társadalomtudományi elméletek empirikus ellenőrzése. Budapest, Osiris Kiadó, 1999.
- Moore, D. S. and McCabe, G. P. Introduction to the Practice of Statistics, 4th ed. New York: W. H. Freeman, 2002.
- Neter, J., W. Wassermann, and M. H. Kutner. 1985. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. 2d ed. Homewood, Illinois.: Richard D.Irwin, Inc.
- PENG, K. C. 1967. The Design and Analysis of Scientific Experiments. Reading, MA: Addison-Wesley.
- Sváb, J. 1981. Biometriai módszerek a kutatásban. Mezőgazdasági Kiadó. Budapest.
- Szűcs István Szerk.: Alkalmazott statisztika. Agroiinform Kiadó, 2002.
- Whittaker, E. T. and Robinson, G. The Calculus of Observations: A Treatise on Numerical Mathematics, 4th ed. New York: Dover, pp. 184-186, 1967.
- Winer, B. J. 1971. Statistical Principles in Experimental Design, 2d. ed. New York:

Mellékletek

1. sz. melléklet. Az ásványvíz fogyasztást és vásárlást elemző kérdőív kérdéseinek bemutatása

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Sorszam	Numeric	8	0		None	None	8	Right	Nominal	Input
2	K1_1	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? A palackozott ásványvíz vásárlása negatívan hat a környezetre.	{1, Egyáltalán...}	None	18	Right	Scale	Input
3	K1_2	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? A zöld csomagolóanyagok használata pozitív hatással van a környezetre.	{1, Egyáltalán...}	None	20	Right	Scale	Input
4	K1_3	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok gyengébb anyagminőséget elfogadni a környezetvédelem érdekében.	{1, Egyáltalán...}	None	20	Right	Scale	Input
5	K1_4	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? A környezetbarát termékekért hajlandó vagyok többet fizetni.	{1, Egyáltalán...}	None	20	Right	Scale	Input
6	K1_5	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Hajlandó vagyok több adót fizetni azért, hogy védjem a környezetet.	{1, Egyáltalán...}	None	13	Right	Scale	Input
7	K2_1	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Csodálom azokat, akiknek drága kocsija, lakása vagy más luxuscikke van.	{1, Egyáltalán...}	None	16	Right	Scale	Input
8	K2_2	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Boldogabb lennék, ha több minden dolgot meg tudnék venni.	{1, Egyáltalán...}	None	14	Right	Scale	Input
9	K2_3	Numeric	8	0	Mennyire ért egyet a következő kijelentéssel? Szeretem a nagy luxust az életemben.	{1, Egyáltalán...}	None	15	Right	Scale	Input
10	K3_1	Numeric	8	0	Értékeje, hogy milyen fontos a következő környezeti szempont: Csökkenteni az éghajlat változást okozó széndioxid-kibocsátást.	{1, Nem fontos...}	None	20	Right	Ordinal	Input
11	K3_2	Numeric	8	0	Értékeje, hogy milyen fontos a következő környezeti szempont: Több tevékenység a természetvédelem érdekében.	{1, Nem fontos...}	None	14	Right	Ordinal	Input
12	K3_3	Numeric	8	0	Értékeje, hogy milyen fontos a következő környezeti szempont: A szemét és a háztartási hulladék csökkentése.	{1, Nem fontos...}	None	15	Right	Ordinal	Input
13	K4	Numeric	8	0	A használt vízes palackokat visszazinném-e a vásárlás helyszínére?	{1, Nem...}	None	5	Right	Ordinal	Input
14	K5_1	Numeric	8	0	Figyelembe veszi-e a MÁRKÁT, amikor palackozott vizet vásárol?	{0, Nem...}	None	5	Right	Nominal	Input
15	K5_2	Numeric	8	0	Figyelembe veszi-e a MÉRETET, amikor palackozott vizet vásárol?	{0, Nem...}	None	5	Right	Nominal	Input
16	K5_3	Numeric	8	0	Figyelembe veszi-e a FORMÁT, amikor palackozott vizet vásárol?	{0, Nem...}	None	5	Right	Nominal	Input
17	K5_4	Numeric	8	0	Figyelembe veszi-e a SÚLYT, amikor palackozott vizet vásárol?	{0, Nem...}	None	5	Right	Nominal	Input
18	K5_5	Numeric	8	0	Figyelembe veszi-e a CSOMAGOLÁS DESIGNJÁT, amikor palackozott vizet vásárol?	{0, Nem...}	None	5	Right	Nominal	Input
19	K5_6	Numeric	8	0	Figyelembe veszi-e a TERMÉK VÉDELMEJÉT, amikor palackozott vizet vásárol?	{0, Nem...}	None	5	Right	Nominal	Input
20	K5_7	Numeric	8	0	Figyelembe veszi-e az ANYAG MINŐSÉGÉT, amikor palackozott vizet vásárol?	{0, Nem...}	None	5	Right	Nominal	Input
21	K5_8	Numeric	8	0	Figyelembe veszi-e a ZÖLD CSOMAGOLÁST, amikor palackozott vizet vásárol?	{0, Nem...}	None	5	Right	Nominal	Input
22	K5_9	Numeric	8	0	Figyelembe veszi-e az ÁRAT, amikor palackozott vizet vásárol?	{0, Nem...}	None	5	Right	Nominal	Input
23	K6_1	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben. Kötetem a környezetbarát témákat.	{1, Soha...}	None	5	Right	Scale	Input
24	K6_2	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben. Kényelmetlenséget is vállaltam azért, hogy környezetbarát legyen a választásom.	{1, Soha...}	None	11	Right	Scale	Input
25	K6_3	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben. Vásárlás közben a saját bevásárló táskám használatát részesítettem előnyben.	{1, Soha...}	None	5	Right	Scale	Input
26	K6_4	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: A háztartásomban újrahasznosítottam a dobozokat.	{1, Soha...}	None	5	Right	Scale	Input
27	K6_5	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: A vásárlásaim során olyan termékeket választottam, amelyek környezetbarátak.	{1, Soha...}	None	5	Right	Scale	Input
28	K6_6	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben: A nem környezetbarát viselkedéseimért büntetnem magam.	{1, Soha...}	None	5	Right	Scale	Input
29	K6_7	Numeric	8	0	Környezetbarát viselkedés gyakorisága az elmúlt 5 évben. Adományozok a környezetvédelemmel kapcsolatos szervezeteknek.	{1, Soha...}	None	5	Right	Scale	Input
30	K7	Numeric	8	0	Nemek	{0, Nő...}	None	11	Right	Nominal	Input
31	K8	Numeric	8	0	Életkor kategóriák	{1, 25-34...}	None	12	Right	Ordinal	Input
32	K9	Numeric	8	0	Legmagasabb iskolai végzettség	{1, Maximum...}	None	18	Right	Ordinal	Input
33	K10	Numeric	8	0	Havi jövedelem (ezer Ft)	{0, Nem vált...}	0	20	Right	Ordinal	Input
34	K4_két_kategória	Numeric	8	0	A használt vízes palackokat visszazinném-e a vásárlás helyszínére?	{0, Nem / T...}	None	20	Right	Nominal	Input
35	K1_összesen	Numeric	8	0	Környezetbarát attitűd összesen	None	None	14	Right	Scale	Input
36	K6_összesen	Numeric	8	0	Környezetbarát viselkedés összesen	None	None	14	Right	Scale	Input

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
37	FAC1_1	Numeric	11	5	környezetvédel...	None	None	13	Right	Scale
38	FAC2_1	Numeric	11	5	kényelem és p...	None	None	13	Right	Scale
39	FAC3_1	Numeric	11	5	környezet iránti...	None	None	13	Right	Scale
40	TSC_5674	Numeric	10	0	3 klaszter	{-1, Outlier ...}	None	8	Right	Nominal
41	QCL_1	Numeric	8	0	3 klaszter (K...	None	None	10	Right	Nominal
42	CLU3_1	Numeric	8	0	3 klaszter (Hier...	None	None	10	Right	Nominal
43	Discriminancia_1	Numeric	10	0	Predicted Grou...	{-1, Outlier ...}	None	8	Right	Nominal
44	Discriminancia_2	Numeric	8	0	Predicted Grou...	None	None	10	Right	Nominal
45	Discriminancia_3	Numeric	8	0	Predicted Grou...	None	None	10	Right	Nominal
46										
47										
48										
49										
50										
51										
52										
53										

Forrás: Balogh (2022)