Kemény Ildikó, Simon Judit, Berezvai Zombor, Kun Zsuzsanna

MARKETINGKUTATÁS KVANTITATÍV MÓDSZEREI – SEGÉDANYAG SPSS PROGRAM HASZNÁLATÁHOZ

Budapest
2021

Kemény Ildikó, Simon Judit, Berezvai Zombor, Kun Zsuzsanna

MARKETINGKUTATÁS KVANTITATÍV MÓDSZEREI – SEGÉDANYAG SPSS PROGRAM HASZNÁLATÁHOZ

Marketing Intézet

ISBN: 978-963-503-864-0

Kiadó: Budapesti Corvinus Egyetem

2021. Budapest
1. kiadás

Tartalom

ELŐSZÓ 1
1. ADATBEVITEL, KÓDOLÁS 2
GYAKORLÓ FELADATOK 4
2. LEÍRÓ STATISZTIKA 5
GYAKORLÓ FELADATOK 11
3. ADATBÁZIS MŰVELETEK 12
GYAKORLÓ FELADATOK 20
4. KERESZTTÁBLA ELEMZÉS 21
GYAKORLÓ FELADATOK 26
5. VARIANCIAELEMZÉS 27
5.1. EGYSZEMPONTOS VARIANCIAELEMZÉS (egyszempontos ANOVA) 27
GYAKORLÓ FELADATOK 33
5.2. TÖBBSZEMPONTOS VARIANCIAELEZMÉS (többszempontos ANOVA) 34
GYAKORLÓ FELADATOK 38
6. GYAKROLÓ ESETEK: leíró statisztika, kereszttábla, varianciaelemzés 39
7. LINEÁRIS REGRESSZIÓS ELEMZÉS 41
GYAKORLÓ FELADATOK 50
8. FAKTORELEMZÉS 51
GYAKORLÓ FELADAT 65
9. KLASZTERELEMZÉS 66
GYAKORLÓ FELADAT 77

ELŐSZÓ

A sikeres marketingkutatási projektek egy jó kutatási brieffel kezdődnek, mely alapján lehetséges a megfelelő kutatási terv elkészítése, a kutatás megvalósítása, elemzése és értékelése. A Marketingkutatás kvantitatív módszerei – Segédanyag SPSS program használatához című könyv a tudományos és a gyakorlati kutatásban egyik legtöbbet használt szoftvercsomag, az IMB SPSS Statistics 25 program használatát mutatja be. Az elméleti megalapozáshoz a Malhotra – Simon (2009): Marketingkutatás könyv nyújt kiváló magyarázatot (https://mersz.hu/malhotra-simon-marketingkutatas). Ebben a könyvben egyszerű, gyakorlati példákon keresztül mutatjuk be a különböző statisztikai módszerek marketingkutatási használatát, és a gyakorló feladatok teszik lehetővé a program használatának alaposabb elsajátítását. A feladatok során használt Italfogyasztási szokások adatbázis a COBEREN együttműködés kutatási projektjén alapul (Consumer Behaviour Erasmus Network, http://www.coberen.eu, Projektszám: 156089-LLP-1-2009-1-ES-ERASMUS-ENW.) Az adatbázis online elérhető, és letölthető.

A könyvben az egyszerűbb adatbázisműveletek után a leíró statisztikai elemzések, majd a kereszttábla, egy- és többszempontos varianciaelemzés, lineáris regresszió elemzés, faktorelemzés és a klaszterelemzés módszere kerül bemutatásra. A módszereket ugyanazon adatbázis felhasználásával mutatjuk be, ez a megközelítés azzal az előnnyel is jár, hogy megismerhető, hogy egy marketingkutatási projektben, adott adatbázis esetében hogyan választhatjuk ki a kutatási kérdésekhez és hipotézisekhez a megfelelő statisztikai és elemzési módszert, adott esetben akár több módszert is kipróbálva.

A könyv több éves tanítási gyakorlaton alapul, folytatva a Marketing Intézetben több évtizede bevezetett és használt segédanyagok sorát. Ezúton is szeretnénk megköszönni a Budapesti Corvinus Egyetem Marketing mesterszakos hallgatóinak a segítségét, akik a Marketingkutatás és piacelemzés tárgy keretében az elmúlt évek során visszajelzéseikkel segítettek a segédanyag fejlesztését.

Minden Olvasónak sikeres és élményekben gazdag marketingkutatási elemzéseket kívánunk!

A könyv Szerzői

1. ADATBEVITEL, KÓDOLÁS

A fejezet célja, hogy megismertesse a kérdőív válaszainak rögzítését, kódolását papíralapú, és online megkérdezés esetén, valamint az SPSS adatfájlok felcímkézését.
A témához kapcsolódó legfontosabb fogalmak:

- mérési szintek:
  - nem metrikus (nominális, ordinális)
  - metrikus (intervallum, arány)
- SPSS-hez kapcsolódó kifejezések:
  - label, value, missing, measure, data/variable view

FELADAT

Papír alapú megkérdezés adatainak bevitele, SPSS-ben exportálása és felcímkézése
***Felhasznált fájlok**: 01_Könyvtárhasználati szokások*

FELADAT

A kitöltött kérdőívek (01_Könyvtárhasználati szokások_kitöltött kérdőívek.pdf) adatainak Excel táblában történő kódolása, és rögzítése.

MEGOLDÁS

1. Kérdések kódolása (sszám, v1, v2, v3_1, v3_2, ….) → Excel tábla első sora szolgál majd a változók elnevezéseként az SPSS-ben
2. Válaszlehetőségek kódolása, minden nem metrikus változót numerikus adatokra cserélünk (CTRL+F funkciók az Excelben), mivel az SPSS numerikus adatokkal dolgozik. A metrikus változók esetében erre nincs szükség, mivel azok már eleve numerikus adatok.
3. Excel tábla feltöltése

FELADAT

A kitöltött Excel tábla SPSS-be történő importálása

1. SPSS program elindítása
2. File / Open / Data

MEGOLDÁS

SPSS adatfájl felcímkézése

Variable view

1. **Name**: kérdés kódja (nem tartalmazhat speciális karaktereket, szóközt, valamint nem kezdődhet számmal). Érdemes a kérdőív struktúráját követni a változó nevek esetében, így egyfajta navigációként is szolgálnak a későbbiekben.
2. **Type**: szám (Numeric) vagy szöveg (String)
3. **Width**: Karakterek hossza
4. **Decimals**: Tizedes jegyek száma

5. **Label**: Pontos kérdés, címke, amely meg fog jelenni a későbbiekben az output táblákban is, ezért érdemes nem túl hosszú, de kifejező elnevezéseket adni.
6. **Value**: Az adott kérdés esetén a számokhoz tartozó értékek megjelölése, mely megjelenik majd az outputokban is a számok helyett, pl. 1 – Férfi, 2 – Nő. Ennek nem metrikus adatok esetén van jelentősége.
7. **Missing**: Hiányzó értékek felvétele → számolások során nem veszi figyelembe az SPSS az itt rögzített értékeket az adott változóból. Jellemzően 9, 99, 999 jelöléseket alkalmazunk, illetve a system missing-et az üresen maradt cellák esetében.
8. **Measure**: Mérési szint → Nominal / Ordinal / Scale, ahol az utóbbi mindkét metrikus mérési szintet (intervallum és arány) magába foglalja.

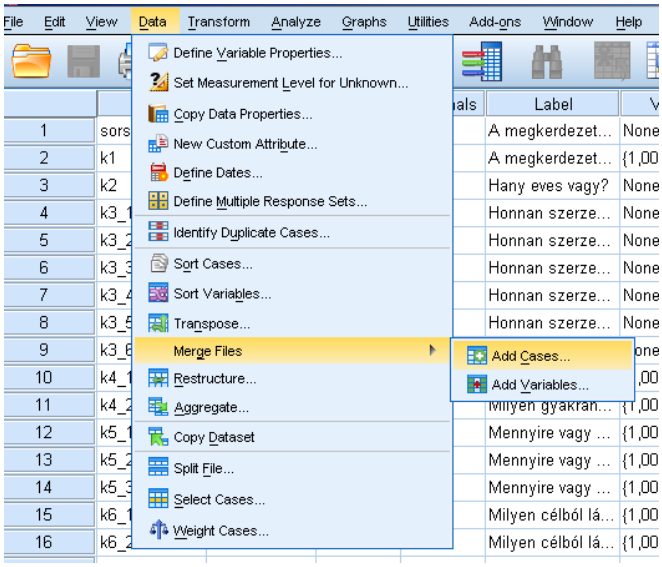
FELADAT

SPSS adatbázisok egyesítése
***Felhasznált fájlok**: 03_Könyvtár_adatbázis*

A két SPSS adatbázis egyesítése (03_konyvtar adatbázis_1fele.sav, 03_konyvtar adatbázis_2fele.sav)

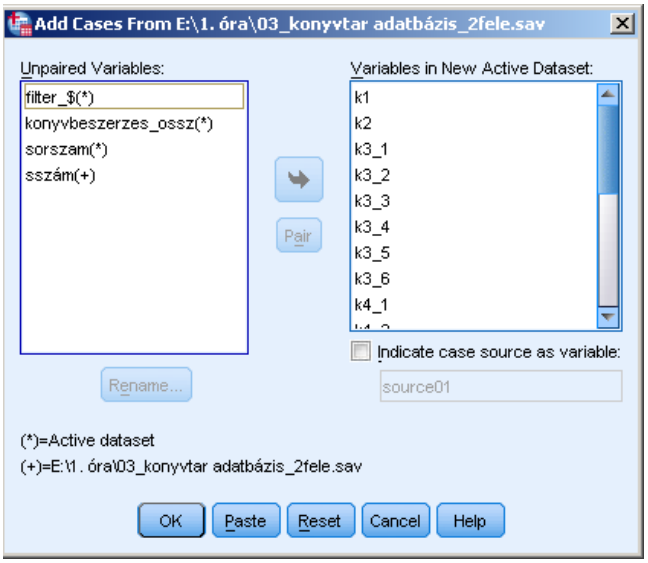
MEGOLDÁS

Elérés útvonala: Data / Merge files / Add cases



Az egyesítés előtt el kell dönteni, hogy ezekkel a változókkal mi történjen: átkerüljenek az egyesített adatbázisba, vagy pedig maradjanak ki belőle.

Az egyesíteni próbált adattáblák változóinak egyezése és különbsége (*unpaired variables*).



SYNTAX

```
*Első adattábla megnyitása.
GET
  FILE='C: \ Eléréshelye\03_konyvtar '+ 'adatbázis_1fele.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.

*Második adattábla hozzáadása nem egyező változókkal együtt.
ADD FILES /FILE=*
  /RENAME (filter_$ konyvbeszerzes_ossz sorszam=d0 d1 d2)
  /FILE='C:\ Eléréshelye \03_konyvtar '+ 'adatbázis_2fele.sav'
  /RENAME (sszám=d3)
  /DROP=d0 d1 d2 d3.
EXECUTE.

*Második adattábla hozzáadása csak a nem egyező változók kihagyásával.
ADD FILES /FILE=*
  /FILE='C:\ Eléréshelye \03_konyvtar '+ 'adatbázis_2fele.sav'
EXECUTE.
```

GYAKORLÓ FELADATOK

1. Importálja az online lekérdezéshez tartozó Excelt táblát SPSS-be! Az így nyert SPSS adattáblát címkézze fel!

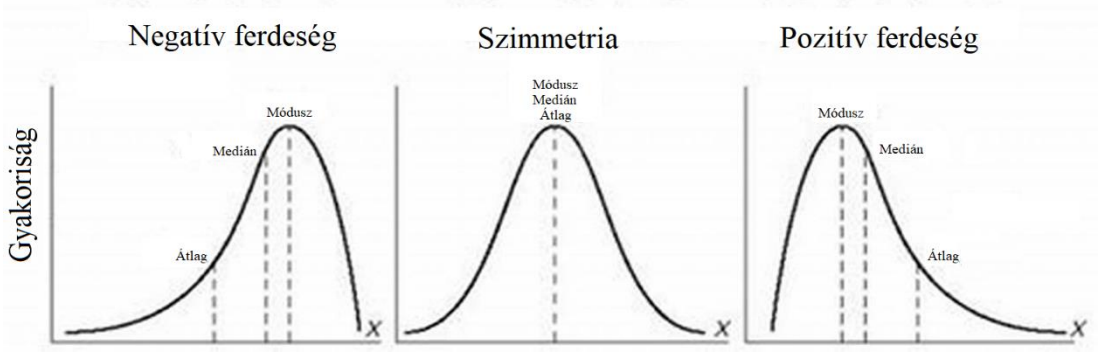
2. LEÍRÓ STATISZTIKA

Az adatelemzés első lépése az adatok áttekintése és leíró statisztikai vizsgálata. Ennek több oka is van:

- Bármilyen elemzés elvégzése előtt fontos megismerkednünk az adatokkal, tisztában lenni azzal, hogy milyen változókat vizsgálunk.
- Az esetleges adatfelvételi és/vagy értelmezési hibák a leíró statisztikai vizsgálat során könnyebben észrevehetők, mint az elemzés során, így segítenek megakadályozni hibás vagy félreérthető eredmények közlését.
- Az esetleges kiugró értékek (outlierek) a leíró statisztikai vizsgálatok során azonosíthatók be legkönnyebben.

A leíró statisztikai vizsgálat során grafikusan lehet ábrázolni a változók eloszlását, gyakorisági táblát lehet készíteni, illetve többféle mutatószám is elérhető:

- **Módusz**: a leggyakrabban előforduló érték a mintában.
- **Medián**: a minta (növekvő sorrendbe állítva) középső eleme, a megfigyelések 50%-a ennél kisebb (vagy egyenlő), míg másik 50%-a ennél nagyobb (vagy egyenlő) értéket vesz fel.
- **Átlag**: az adott változó mintabeli összege osztva a mintaelemszámmal.
- **Szórás**: az átlagtól vett átlagos eltérés.
- **Variancia**: a szórás négyzete.
- **Minimum**: a legkisebb érték a mintában.
- **Maximum**: a legnagyobb érték a mintában.
- **Tartomány**: a maximum és a minimum különbsége.
- **Kvartilisek**: negyedelőpontok; az első kvartilis esetén a megfigyelések 25%-a ennél kisebb (vagy egyenlő), míg 75%-a ennél nagyobb (vagy egyenlő) értéket vesz fel; a második kvartilis a medián; a harmadik kvartilis esetén a megfigyelések 75%-a ennél kisebb (vagy egyenlő), míg 25%-a ennél nagyobb (vagy egyenlő) értéket vesz fel.
- **Ferdeség** (skewness): az eloszlás szimmetriáját vizsgáló mutatószám; 0 esetén az eloszlás szimmetrikus, pozitív szám esetén jobbra elnyúló (balra ferde), míg negatív esetben balra elnyúló (jobbra ferde) az eloszlás alakja.



- **Csúcsosság** (kurtosis): az eloszlás csúcsosságát, a szélső értékek elhelyezkedését vizsgáló mutatószám. Az SPSS Statistics által kiszámított értéke normális eloszlás esetén 0, negatív értékek esetén a normális eloszlásnál lapultabb (kevesebb szélső értéket tartalmazó) az eloszlás, míg pozitív értékek esetén a normális eloszlásnál csúcsosabb (több szélső értéket tartalmazó) az eloszlás.

Az alkalmazható leíró statisztikai eszköztár alapvetően a vizsgálni kívánt változó mérési szintjének függvénye.

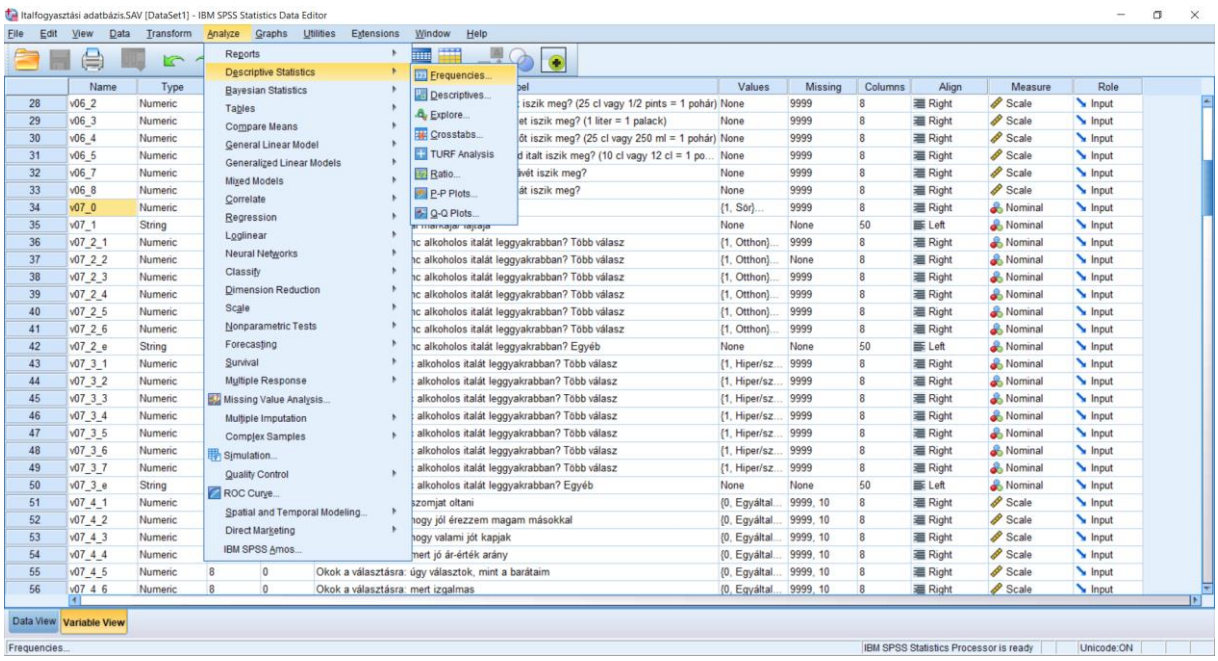
- Nominális skála: gyakorisága tábla készíthető, ahol látszódik, hogy az egyes válaszopciókat hányan (és a minta hány százaléka) választotta. Az egyetlen alkalmazható mutatószám a módusz.
- Ordinális skála: gyakorisága tábla készíthető, ahol látszódik, hogy az egyes válaszopciókat hányan (és a minta hány százaléka) választotta. A módusz mellett a medián is értelmes.
- Intervallumskála: mivel az intervallumskálán mért változók általában nagyon sokféle értéket vehetnek fel, gyakorisági tábla helyett az eloszlás (a sűrűségfüggvény, hisztogram) ábrázolása terjedt el. Emellett az összes bemutatott mutatószám kiszámítható és értelmes.
- Arányskála: mivel az arányskálán mért változók általában nagyon sokféle értéket vehetnek fel, gyakorisági tábla helyett az eloszlás (a sűrűségfüggvény, hisztogram) ábrázolása terjedt el. Emellett az összes bemutatott mutatószám kiszámítható és értelmes.

FELADAT

Vizsgáljuk meg, hogy mely alkoholos italt mennyien kedvelnek.
***Felhasznált fájlok**: Italfogyasztási szokások.sav*[1]

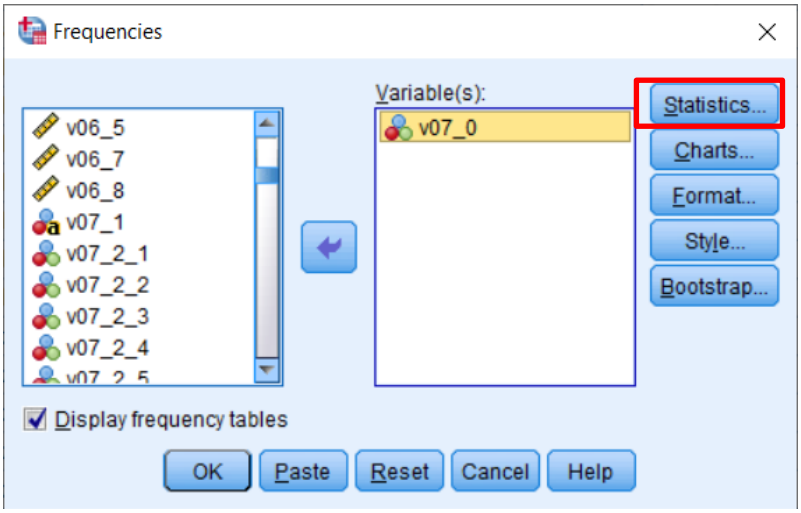
MEGOLDÁS

Elérés útvonala: Analyze/Descriptive Statistics/Frequencies



[1] A felhasznált *Italfogyasztási szokások* adatbázis a COBEREN együttműködés kutatási projektjén alapul (Consumer Behaviour Erasmus Network, http://www.coberen.eu). Projektszám: 156089-LLP-1-2009-1-ES-ERASMUS-ENW. Az adatbázis online elérhető, és letölthető.

Változó kiválasztása: v07_0

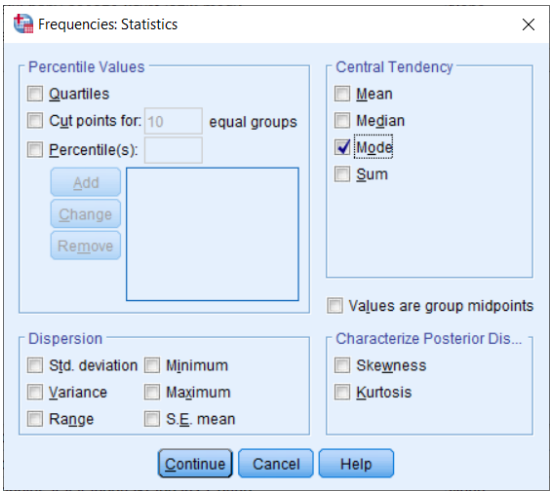


Mivel ez nominális skálán mért változó, így gyakorisági táblát és móduszt vizsgálunk csak. A gyakorisági tábla alapértelmezetten be van állítva, mert a 'Display frequency tables' opció be van pipálva.

Lekért adatok:

Statistics

- ✓ Mode



ÉRTELMEZÉS

Statistics

v07_0 Kedvenc alkoholos ital

N	Valid	894
	Missing	0
Mode		3

v07_0 Kedvenc alkoholos ital

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Sör	275	30.8	30.8	30.8
	2 Cider	27	3.0	3.0	33.8
	3 Bor	345	38.6	38.6	72.4
	4 Tömény	56	6.3	6.3	78.6
	5 Kevert ital	69	7.7	7.7	86.4
	6Nem iszom alkoholt	122	13.6	13.6	100.0
	Total	894	100.0	100.0	

A sört 275 fő (30,8%) jelölte kedvenc italának, a cidert 27 fő (3,0%), a bort 345 fő (38,6%), a tömény italokat 56 fő (6,3%), a kevert italokat 69 fő (7,7%), míg 122 fő (13,6%) nem iszik alkoholt. A Percent és a Valid Percent csak akkor tér el, ha vannak olyanok, akik nem válaszoltak a kérdésre. A Percent a teljes mintaelemszámra vetítve mutatja a válaszok megoszlását, míg a Valid Percent csak a valid válaszokat adókra.

A legtöbben (módusz) a 3-as opciót, vagyis a bort választották kedvenc alkoholos italuknak.

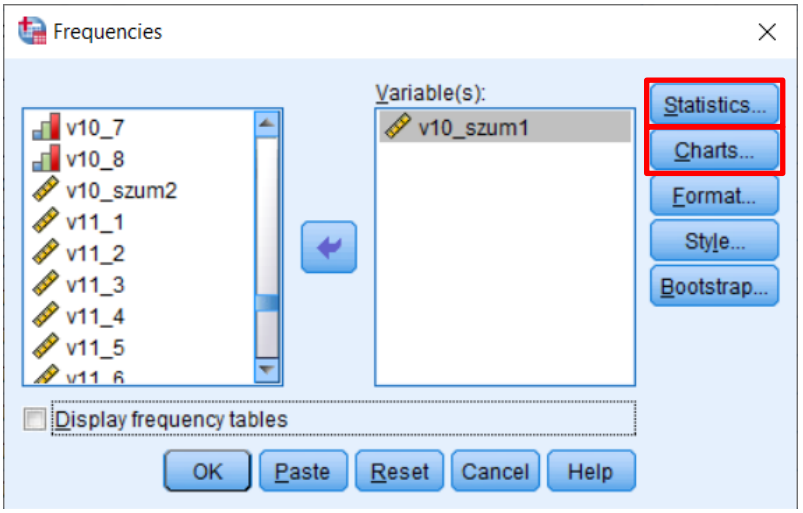
FELADAT

Vizsgáljuk meg a heti alkoholos italköltés alakulását.
***Felhasznált fájlok**: Italfogyasztási szokások.sav*

MEGOLDÁS

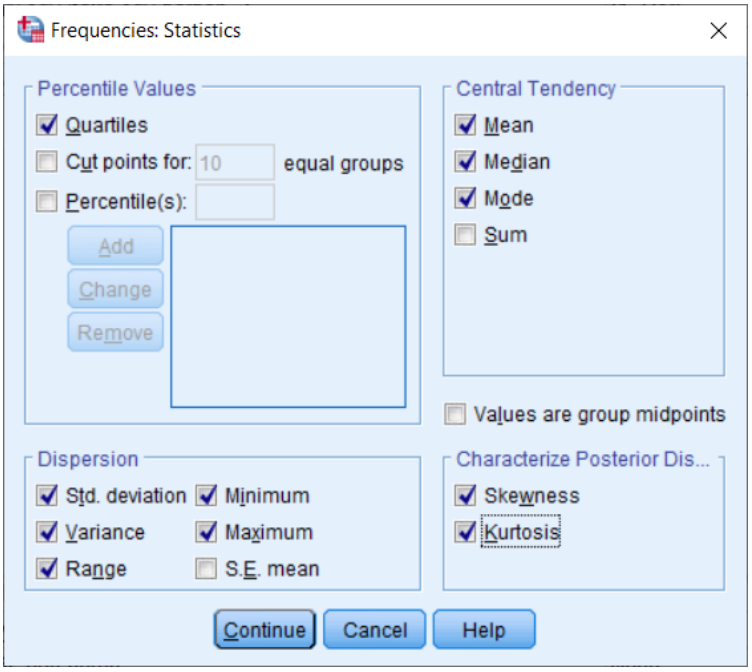
Elérés útvonala: Analyze/Descriptive Statistics/Frequencies

Változó kiválasztása: v10_szum1

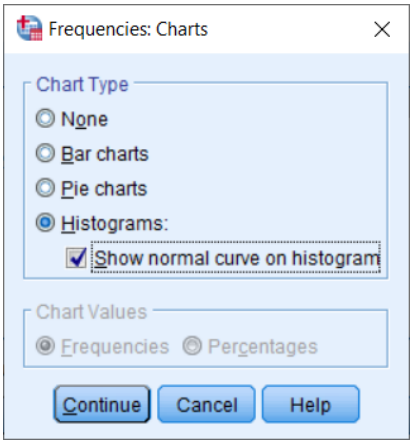


Mivel arányskálán mért változóról van szó, ami rengeteg különböző értéket vehet fel, így a gyakorisága táblának nincs értelme, a ’Display frequency tables’ elől kivettük a pipát.

Lekért adatok:
Statistics: Quartiles, Mean, Median, Mode, Std. deviation, Variance, Range, Minimum, Maximum, Skewness, Kurtosis



Charts: Histrogram

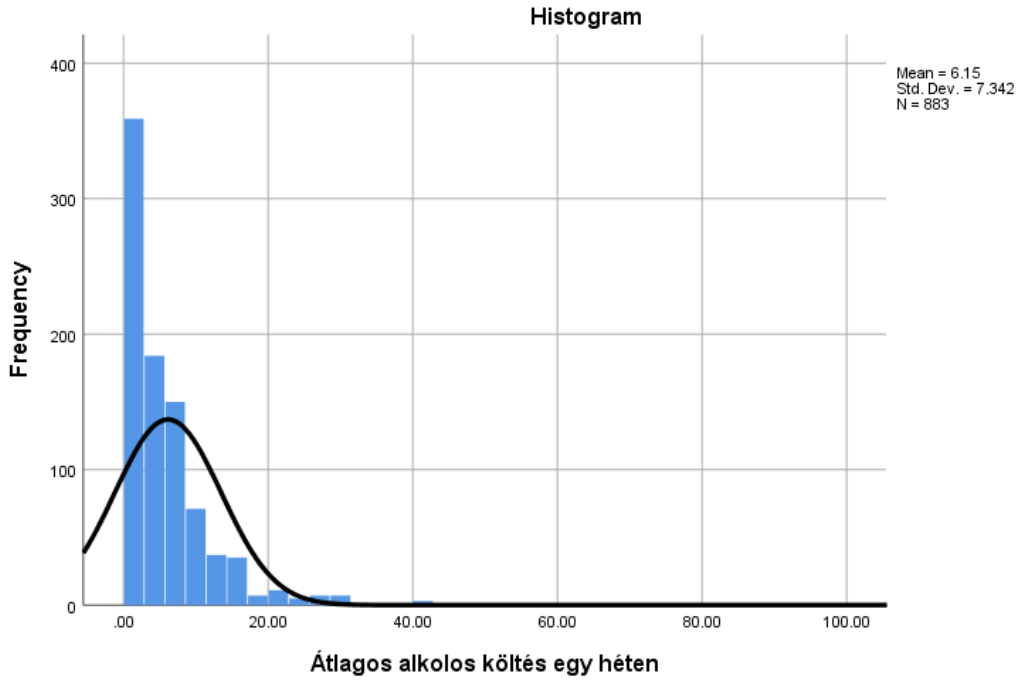


ÉRTELMEZÉS

Statistics

v10_szum1 Átlagos alkolos költés egy héten

N	Valid	883
	Missing	11
Mean		6.1523
Median		4.6667
Mode		2.33
Std. Deviation		7.34228
Variance		53.909
Skewness		3.784
Std. Error of Skewness		.082
Kurtosis		24.333
Std. Error of Kurtosis		.164
Range		81.67
Minimum		.00
Maximum		81.67
Percentiles	25	2.3333
	50	4.6667
	75	7.0000



A heti átlagos alkoholos italköltés a kitöltők körében 6,15 euró, a medián 4,67 euró (tehát a kitöltők fele ennél kevesebbet, fele pedig ennél többet költ hetente alkoholos italokra), a módusz 2,33 euró, a szórás (tehát az átlagtól való átlagos eltérés) 7,34 euró. A legkevesebbet költő semmit sem költ alkoholos italokra, míg a legtöbbet költő heti 81,67 eurót költ alkoholos italokra. Az eloszlás jobbra elnyúló (balra ferde), ami látható a hisztogramról is. Az eloszlás csúcsos, ahogy ezt a pozitív kurtosis és a hisztogram is mutatja.

SYNTAX

*Nem metrikus változó esetében.

```
FREQUENCIES VARIABLES=v07_0
/STATISTICS=MODE
/ORDER=ANALYSIS.
```

*Metrikus változó esetében.

```
FREQUENCIES VARIABLES=v10_szum1
/FORMAT=NOTABLE
/NTILES=4
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SKEWNESS SESKEW KURTOSIS SEKURT
/HISTOGRAM NORMAL
/ORDER=ANALYSIS.
```

GYAKORLÓ FELADATOK

1. Jellemezze a mintát a demográfiai változók alapján! (v17-től v24-ig)
2. A Pickwick szeretne egy promóciós kampányt megvalósítani és ehhez egy kutatást tervez. A kampány megtervezéséhez az alábbi információkra van szüksége. (1) Hogyan jellemezhető a minta a kedvenc alkoholmentes ital alapján (v08_0)? (2) Hogyan alakul a minta teafogyasztása (v06_8)?
3. Egy alkoholosital gyártó cég szeretné megérteni, hogy az egyének miért választja kedvenc alkoholos italukat (v07_4_1 - v07_4_17). Melyek a leginkább és legkevésbé jellemző választási indokok?
4. Melyek a legkedveltebb italok, melyek a kitöltők kedvenc zenéjük hallgatása közben választanak (v09_12_1) És ünneplések közben (v09_14_1)?

3. ADATBÁZIS MŰVELETEK

Az adatelemzés során gyakran van szükségünk arra, hogy a meglévő változóinkat átalakítsuk, összesítsük, szűrjük. A szemináriumon során a leggyakrabban használt adatbázis műveletek elvégzését ismerjük meg az SPSS programban:

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

Data / Select cases

Jóformán elképzelhetetlen, hogy egy adatbázis elemzése során ne lenne szükségünk az adatok szűrésére. Például csak a férfiak válaszai érdekelnek bennünket, csak a három legmagasabb értékesítést magáénak tudható áruház adatait akarjuk elemezni, esetleg rá akarunk fókuszálni a termékskála elmúlt évben bevezetett variánsaira, ésatöbbi. A **select cases** parancsot használjuk szűrés céljából (ld. az Excel filter / szűrés alkalmazását).

Legtipikusabban a *select* mező *if condition satisfied* lehetőségét választjuk, ahol a szelektálásra kijelölt változó segítségével definiálhatjuk a kívánt tartományt. Ezt a lehetőséget alkalmazzuk, hogy a fókuszba kívánt opciókat leválogassuk, a nem érdekes elemeket pedig kizárjuk az aktuális elemzésből. Az éppen nem elemzendő eseteket törölhetjük, vagy kizárhatjuk az analízisből (*delete unselected cases* vagy *filter out unselected cases*), esetleg egy új adattáblát hozhatunk létre a leválogatott esetekből (*copy selected cases to a new dataset*). A három lehetőség közül főleg a *filter* megoldást használjuk.

Transform / Compute Variable

Az elemzések során előfordul, hogy a változók valamilyen matematikai transzformációjára van szükségünk. Összesíteni, átlagolni, szorozni stb. szeretnénk az adatainkat, létrehozva ezzel egy új változót. A létrejövő új változó nevét a *target variable* mezőben kell nevesítenünk, míg a *numeric expression* mezőben a matematikai műveletet kell definiálnunk. A *function group* mező segítségével előre beépített függvények közül választhatunk (pl: statistical: max, mean, mode, median, sum, variance, min). A számított új változó pedig megjelenik az adatbázisban a változó lista végén a *target variable* mezőben megadott névnek megfelelően. Ezt követően az új változó felcímkézése még szükséges a további elemzések megkönnyítése céljából.

Transform / Recode

Előfordulhat, hogy a meglévő változóink átalakítására van szükség. Például kiinduló kategorizált változó esetében a meglévő 10 kategóriát szeretnénk 5 kategóriában megjeleníteni, vagy a kiinduló metrikus változónkat szeretnénk kategorizált változóvá átalakítani. Ebben az esetben újra kell kódolnunk a *recode* segítségével. A menüsorból választhatunk a *recode into same variable* és a *recode into different variable* lehetőségek közül, melyek közül jellemzően az utóbbit választjuk, mivel ez érintetlenül meghagyja az eredeti változót is, míg a második lehetőség átírja azt az új kategóriáknak megfelelően. Amennyiben új változót hozzunk létre az átkategorizálással, akkor az átalakítandó változót kiválasztását követően meg kell adnunk a létrehozandó változó nevét (*name)* és *label*-ét is, melyet a *change* lenyomásával rögzíthetünk. Ezt követően az *old and new values* -ra kattintva a definiálási ablak jelenik meg, ahol az *old value* területen az alakítandó régi értékeket, míg a *new value* területen az új értékeket kell kijelölnünk. Az *old -> new* mező a már rögzített változtatásokat listázza. Az átkódolást követően létrejön egy új változó a változó lista végén, melynek felcímkézése ebben az esetben is tanácsos.

Data / Split file
Ezt az lehetőséget akkor használjuk, ha az elemzési eredményeinket csoportosítva szeretnénk látni. A csoportosítás alapjául nem metrikus változót kell választani a *Groups based on:* mező használatával. A csoportosítás létrejöhet egy output táblán belül, ilyenkor a *Compare groups* lehetőséget választjuk, vagy külön-külön output táblaként minden egyes kategóriára az *Organize output by groups* lehetőséggel élve. A parancsot végrehajtva az adattábla jobb alsó sarkában megjelenik a *Split by ....* felirat, és a csoportosítás egészen addig fennmarad, míg azt ki nem kapcsoljuk az *Analyze all cases, do not create groups* opcióval.

FELADAT

Egy szeszesital gyártással foglalkozó vállalat szeretné jobban megismerni az angolszász piacokat. Készítsük elő az adatbázist a szűréssel.
***Felhasznált fájlok**: Italfogyasztási szokások.sav*

MEGOLDÁS

Elérés útvonala: Data/Select cases

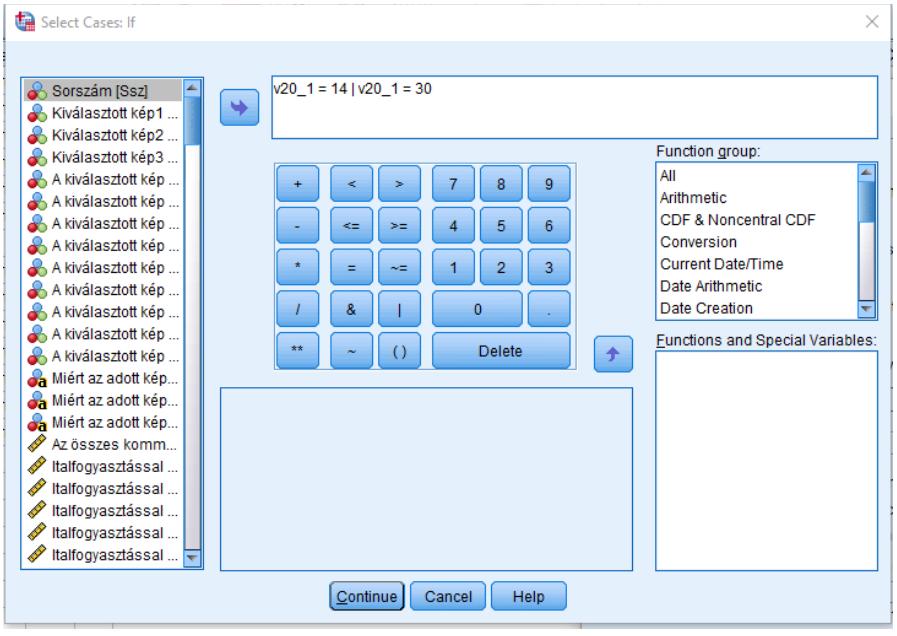
A válaszadók származási helye a v20-as változóban került rögzítésre, 14 Írország, míg 30 Anglia kóddal. Az ő szokásaikat kívánjuk figyelni, ezért ők a szűrés fókusza.

Data / Select cases / Select / If condition satisfied / v20_1 = 14 | v20_1 = 30

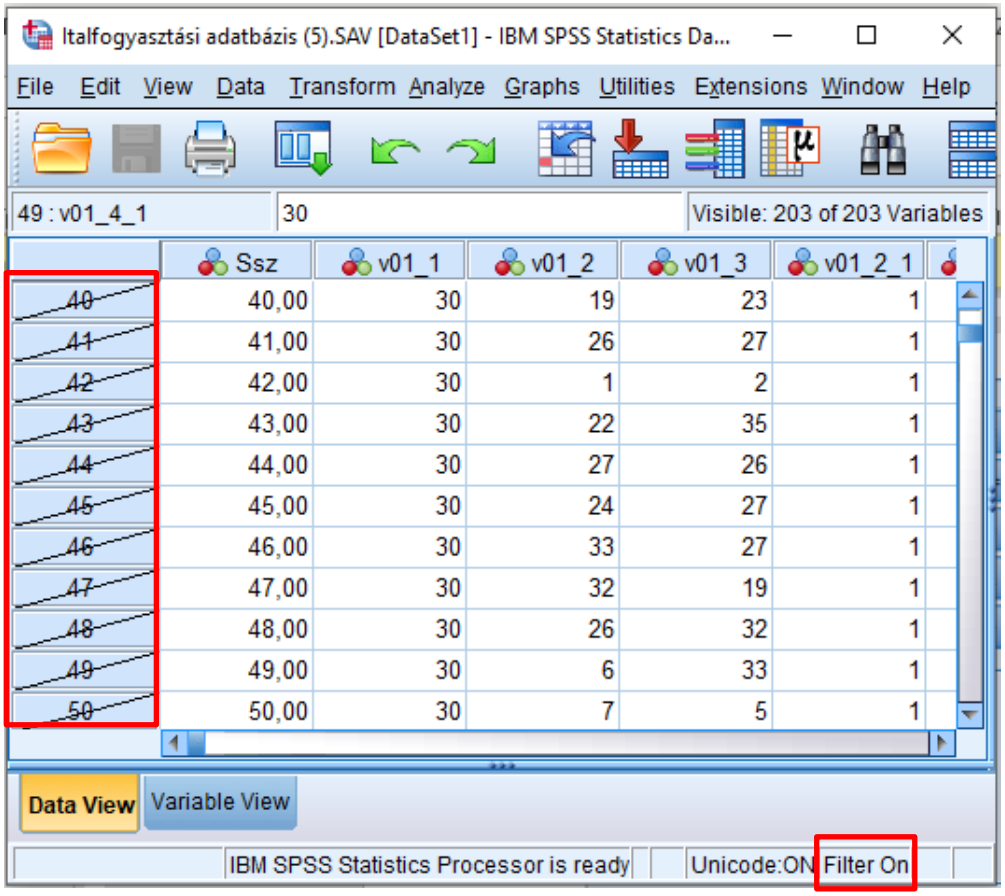
Több feltétel egy változón belüli teljesüléséhez VAGY kapcsolat, azaz „I” logikai jel szükséges.

Több feltétel egyidejű teljesüléséhez több változó által definiálva ÉS kapcsolat, azaz & jel szükséges.

Output / filter out unselected cases



A szűrés során kizárt esetek sorszámát *Data View*-ban áthúzza az SPSS, illetve a jobb alsó sarokban megjelenik a *Filter On* felirat.



A szűrést követően érdemes ellenőrizni, hogy helyesen szűrtünk-e. Ezt legegyszerűbben egy gyakorisági táblával tehetjük meg.

Elérés útvonala: *Analyze / Descriptive statistics / Frequences*

Variable(s): v20
Display frequency tables

Származás helye_kódolt

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Írország	178	50,0	50,0	50,0
	Anglia	178	50,0	50,0	100,0
	Total	356	100,0	100,0	

Látható, hogy már csak az angolszász országok maradtak bent a vizsgálandó esetek között. FONTOS! A szűrési feltétel egészen addig érvényben marad, amíg azt ki nem kapcsoljuk a következőképpen:

Data / Select cases / Select / All cases

FELADAT

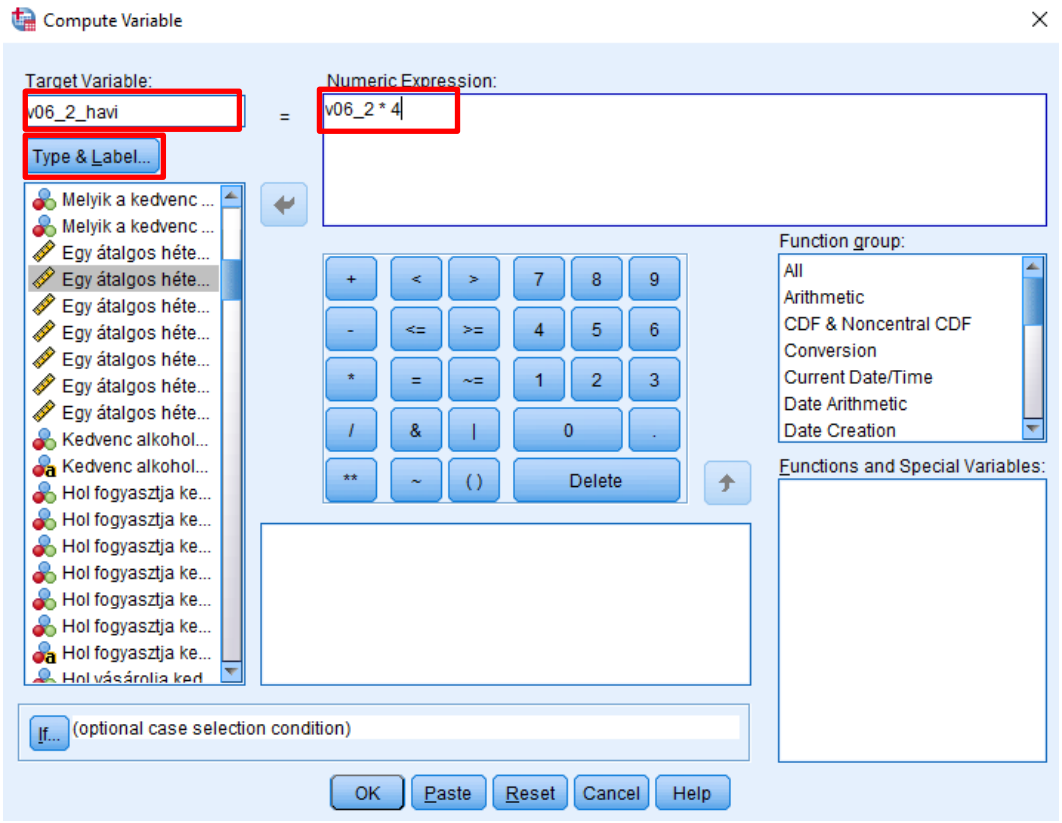
A szeszesital gyártó cég egy új sört szeretne piacra dobni, ezért szeretné megérteni, hogy az angolszász válaszadók (tehát a szűrési feltételünk marad!) egy hónap alatt mennyi sört fogyasztanak.
***Felhasznált fájlok**: Italfogyasztási szokások.sav*

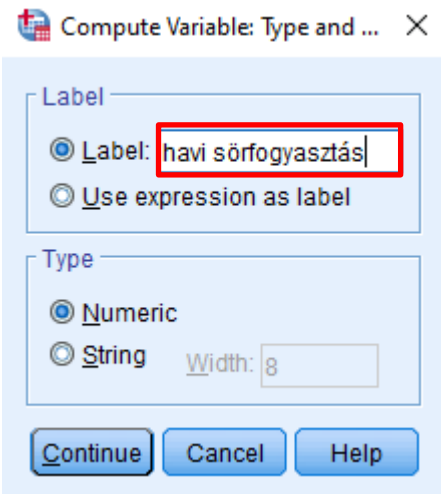
MEGOLDÁS

Az adatfelvétel során heti fogyasztást kérdeztünk (v06_2) a válaszadóktól, azonban egy egyszerű szorzással havi adattá alakíthatjuk.

Elérés útvonala: Transform / Compute variable

Meg kell adnunk egy *Target Variable*-t az új változónak: v06_2_havi, aminek labelt-t (*type & label*) is kell adnunk: havi sörfogyasztás.
A *numeric expression* mezőben pedig definiáljuk a megfelelő műveletet: v06_2 * 4





A változó lista végén megjelenik egy új változó, a fentieknek megfelelően definiált névvel és labellel, mely a havi (azaz a 4 heti) sörfogyasztási adatokat tartalmazza. Amennyiben szükséges további beállításokat tehetünk a változóra vonatkozóan a *Variable view*-ban.

FELADAT

A szeszesital gyártó cég tovább elemzi az angolszász válaszadókat (tehát a szűrési feltételünk marad!). Az Eurostat adatai alapján arra a következtetésre jutnak, hogy túl sok jövedelemkategóriát (v24) tartanak nyilván, ezért úgy döntenek, hogy a 10%-nál kisebb kategóriák összevonásra kerülnek.
Felhasznált fájlok: *Italfogyasztási szokások.sav*

MEGOLDÁS

Először vizsgáljuk meg az eredeti jövedelemkategóriákat.
Elérés útvonala*: Analyse / Descriptive Statistics / Frequencies / d24 / Display frequency tables*

Kiinduló gyakorisági tábla a jövedelem kategóriák esetében mutatja, hogy a 3 legalacsonyabb jövedelmi kategóriában 10% alatti válaszadó tartozik, ezért ezt a 3 kategóriát összevonjuk.

A háztartás nettó havi jövedelme

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Több mint 1.200.000 Ft (5.000 euró)	38	10,7	11,1	11,1
	700.-1.200.000 Ft (3-5.000 euró)	108	30,3	31,5	42,6
	500-700.000 Ft (2-3.000 euró)	90	25,3	26,2	68,8
	250 - 500.000 Ft (1-2.000 euró)	67	18,8	19,5	88,3
	120 - 250.000 Ft (500-1.000 euró)	25	7,0	7,3	95,6
	50 - 120.000 Ft (2-500 euró)	11	3,1	3,2	98,8
	Kevesebb mint 50.000 Ft (200 euró)	4	1,1	1,2	100,0
	Total	343	96,3	100,0	
Missing	9999	13	3,7		
Total		356	100,0		

A kiinduló kategorizálás így néz ki (*data view / v24 / values*):

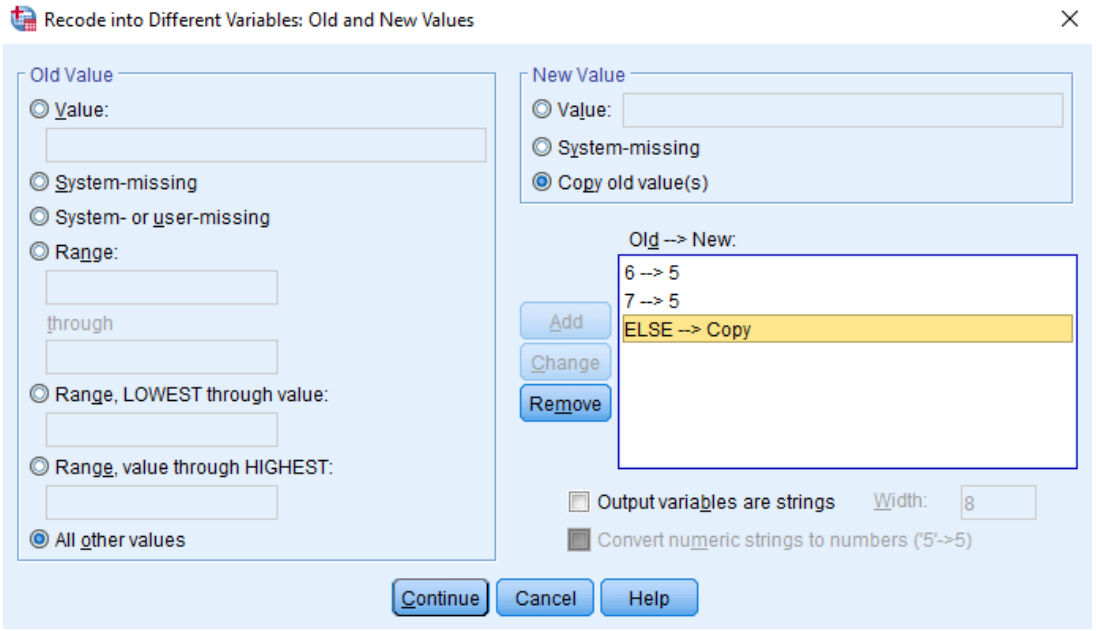
1 = "Több mint 1.200.000 Ft (5.000 euró)"
2 = "700.-1.200.000 Ft (3-5.000 euró)"
3 = "500-700.000 Ft (2-3.000 euró)"
4 = "250 - 500.000 Ft (1-2.000 euró)"
5 = "120 - 250.000 Ft (500-1.000 euró)"
6 = "50 - 120.000 Ft (2-500 euró)"
7 = "Kevesebb mint 50.000 Ft (200 euró)"

Tehát az 5, 6, 7 kategóriákat kívánjuk egyesíteni.

Elérés útvonala: *Transform / recode into different variable*

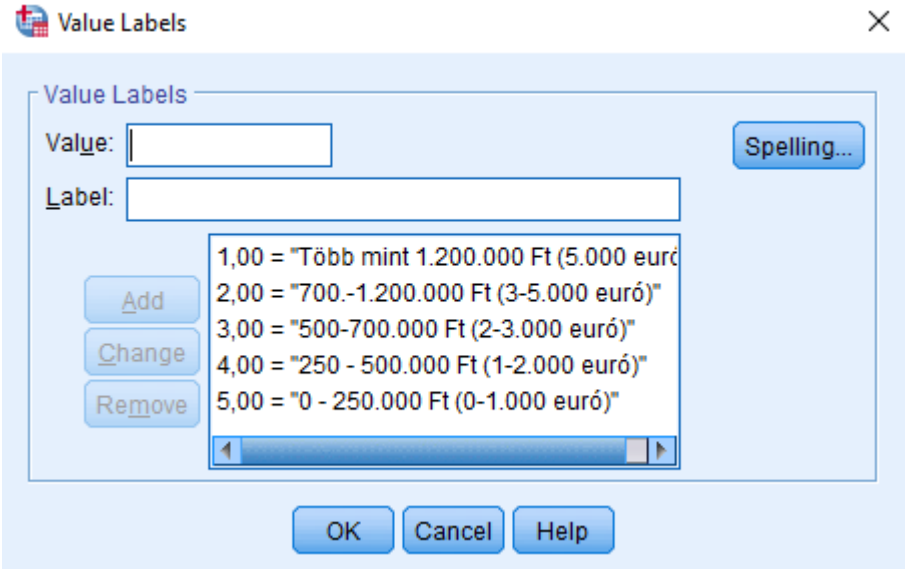
input variable: v24
output variable: name: v24_recode label: háztartási jövedelem – kódolt
>> change

old and new values:
old value: 6 >> new value: 5
old value: 7 >> new value: 5
all other values >> copy old values



A változó lista végén megjelenik egy új változó, a fentieknek megfelelően definiált névvel és labell-el, mely az átkódolt adatokat tartalmazza. A variable view-ban szükséges rögzítenünk az új kategóriákat a *values*-ban.

Elérés útvonala: *Analyse / Descriptive Statistics / Frequencies / d24_recode / Display frequency tables*



Ellenőrizzük le a 10% feltételt és az átkódolást egy gyakorisági táblával.

háztartási jövedelem - kódolt

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Több mint 1.200.000 Ft (5.000 euró)	38	10,7	10,7	10,7
	700.-1.200.000 Ft (3-5.000 euró)	108	30,3	30,3	41,0
	500-700.000 Ft (2-3.000 euró)	90	25,3	25,3	66,3
	250 - 500.000 Ft (1-2.000 euró)	67	18,8	18,8	85,1
	0 - 250.000 Ft (0-1.000 euró)	40	11,2	11,2	96,3
	9999,00	13	3,7	3,7	100,0
	Total	356	100,0	100,0	

FELADAT

A szeszesital gyártó cég a férfi és a női (v17) angolszász válaszadókat (tehát a szűrési feltételünk marad!) külön szeretné megvizsgálni. A kedvenc alkoholos italokat szeretné összehasonlítani (v07_0).
***Felhasznált fájlok**: Italfogyasztási szokások.sav*

MEGOLDÁS

Data / Split file / Compare groups / Groups based on V17
majd
Analyze / Descriptive statistics / Frequencies / Display frequancy tables / v07_0

Kedvenc alkoholos ital

Válaszadó neme			Frequency	Percent	Valid Percent	Cumulative Percent
Férfi	Valid	Sör	80	44,9	44,9	44,9
		Cider	11	6,2	6,2	51,1
		Bor	45	25,3	25,3	76,4
		Tömény	16	9,0	9,0	85,4
		Kevert ital	2	1,1	1,1	86,5
		Nem iszom alkoholt	24	13,5	13,5	100,0
		Total	178	100,0	100,0	
Nő	Valid	Sör	14	7,9	7,9	7,9
		Cider	8	4,5	4,5	12,4
		Bor	104	58,4	58,4	70,8
		Tömény	26	14,6	14,6	85,4
		Kevert ital	7	3,9	3,9	89,3
		Nem iszom alkoholt	19	10,7	10,7	100,0
		Total	178	100,0	100,0	

SYNTAX

```
*Szűrés.
USE ALL.
COMPUTE filter_$=(v20 = 14 | v20 = 30).
VARIABLE LABELS filter_$ 'v20 = 14 | v20 = 30 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMATS filter_$ (f1.0).
FILTER BY filter_$.

*Compute parancs.
COMPUTE v06_2_havi=v06_2 * 4.
VARIABLE LABELS  v06_2_havi 'havi sörfogyasztás'.
EXECUTE.

*kiinduló gyakoriság
FREQUENCIES VARIABLES=v24
  /ORDER=ANALYSIS.
*átkódolás
  RECODE v24 (6=5) (7=5) (ELSE=Copy) INTO v24_recode.
VARIABLE LABELS  v24_recode 'háztartási jövedelem - kódolt'.
EXECUTE.
*ellenőrző gyakoriság
FREQUENCIES VARIABLES=v24_recode
  /ORDER=ANALYSIS.

*sorba rendezés
SORT CASES  BY v17.
SPLIT FILE LAYERED BY v17.

FREQUENCIES VARIABLES=v07_0
  /ORDER=ANALYSIS.
```

GYAKORLÓ FELADATOK

1. Egy kapszulás kávégépeket gyártó cég egy design központú kávégép megalkotásában érdekelt Ezért szeretné jobban megismerni azokat a női (v17=2) fogyasztókat, akiknek a kedvenc nem alkoholos itala a kávé (v08_0=6). Készítsük elő az adatbázis a szűréssel.
2. A legújabb kutatások szerint a nők számára kiemelkedően fontos a vásárlás során, hogy a fenntarthatóság szempontjait érvényesíthessék. Ezért a kávékedvelő nők esetében (szűrés marad!) a kávégépeket gyártó cég szeretne egy átlagos környezettudatossági indexet létrehozni vásárlási attitűdök tekintetében. Készítsük el az indexet a további elemzésekhez.
3. A kávékedvelő nők életkorát (v18) kategorizálni szeretnénk 15 éves intervallumok segítéségével. Készítsük el az új, kategorizált változót.

4. KERESZTTÁBLA ELEMZÉS

A kereszttábla elemzés célja, hogy megvizsgálja, hogy van-e kapcsolat két nem metrikus változó között. A kereszttábla elemzés tesztje a Chi-négyzet próba, melynek nullhipozétise a következő:
H0: Nincs kapcsolat a két változó között.

Amennyiben a Chi-négyzet próba eredménye alapján H0-t elvetjük, tehát van kapcsolat a két változó között, akkor annak mértékét is vizsgálni tudjuk a Cramer V, illetve a Phi mutatókkal. Az elemzés során a kapcsolat irányáról nem kapunk információt.
A hipotézisvizsgálat mellett választ kapunk a két ismérv szerinti gyakoriságokra, és százalékos megoszlásokat (sor, oszlop, teljes).
Fontos! Figyelni arra, hogy ne legyen az összes cellaszám 20%-ánál több olyan cella, melyben 5-nél kevesebb elem található, mivel ebben az esetben semmilyen kapott eredmény nem értelmezhető! Megoldás: kategorizálás illetve nagyobb intervallumú kategóriák létrehozása

A szemináriumi munka során a következő fogalmakkal foglalkozunk:

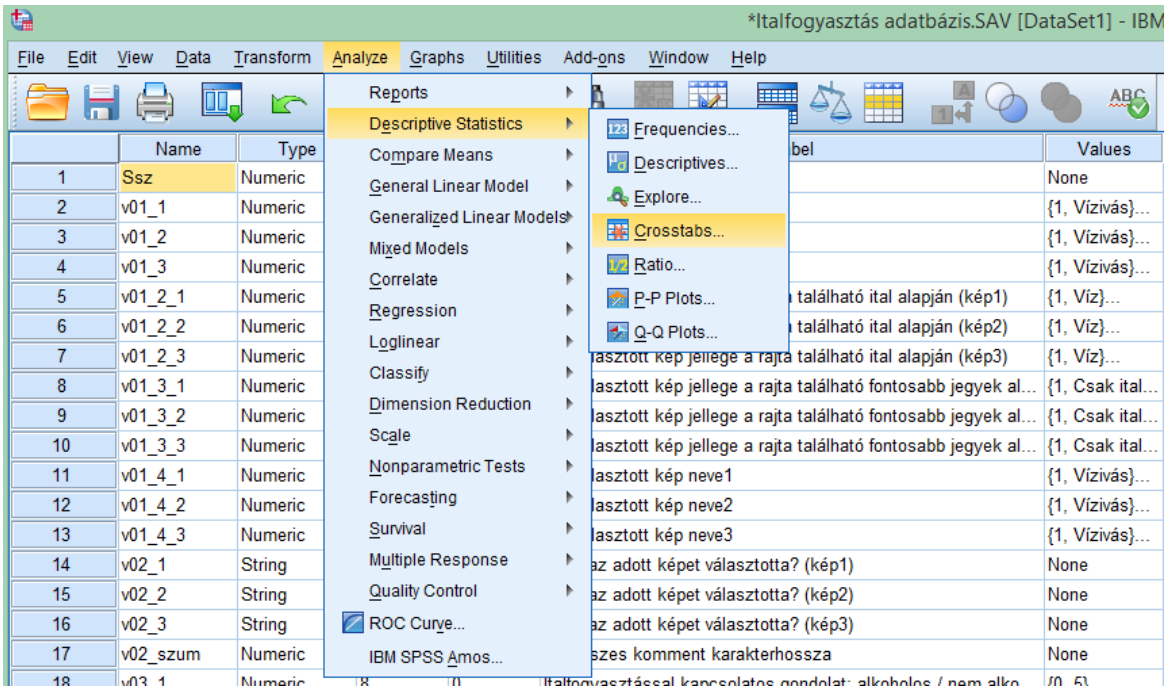
- Chi-négyzet próba
- Cramer V és Phi mutató

FELADAT

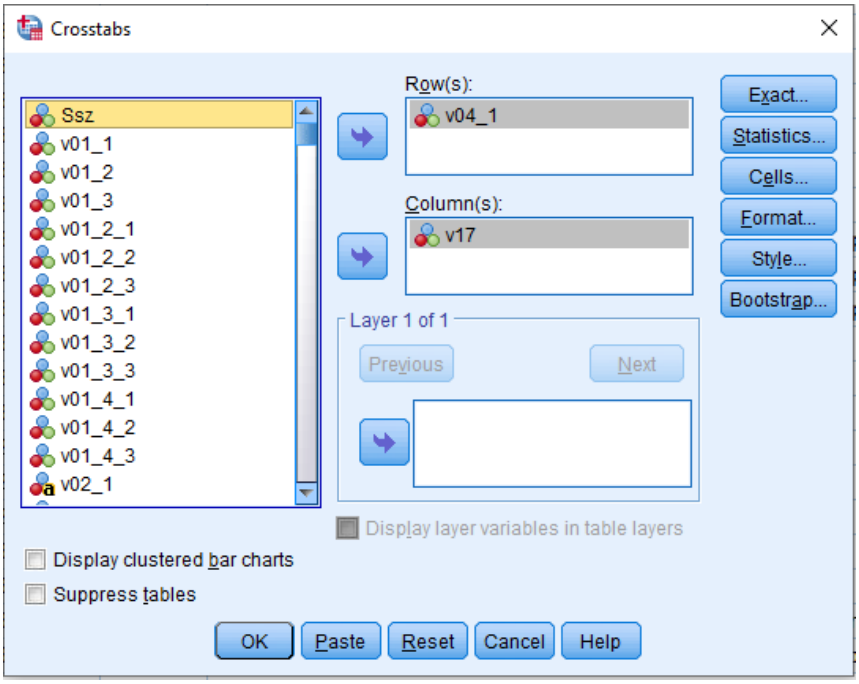
Vizsgáljuk meg, hogy van-e összefüggés a kedvenc alkoholos ital típusa (v04_1) és a válaszadó neme (v17) között?
***Felhasznált fájlok**: Italfogyasztási szokások.sav*

MEGOLDÁS

Elérés útvonala: Analyze/Descriptive Statistics/Crosstabs



Változók bevitele: Row(s): v04_1, Column(s): v17

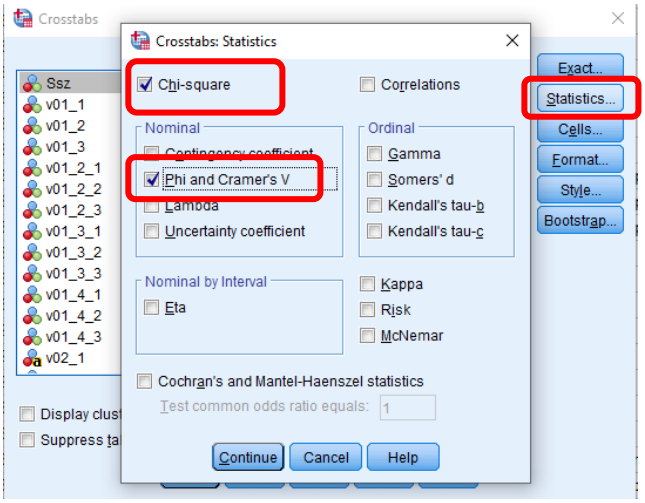


Mivel a kereszttábla elemzés során a kapcsolat irányáról nem mondunk semmit, ezért a változók sorokban (függő) és oszlopokban (független) való elhelyezése a kutató ízlésére van bízva. Érdemes azonban végig gondolni, hogy vajon melyik változó lehet a független és függő változó.

Lekért adatok:

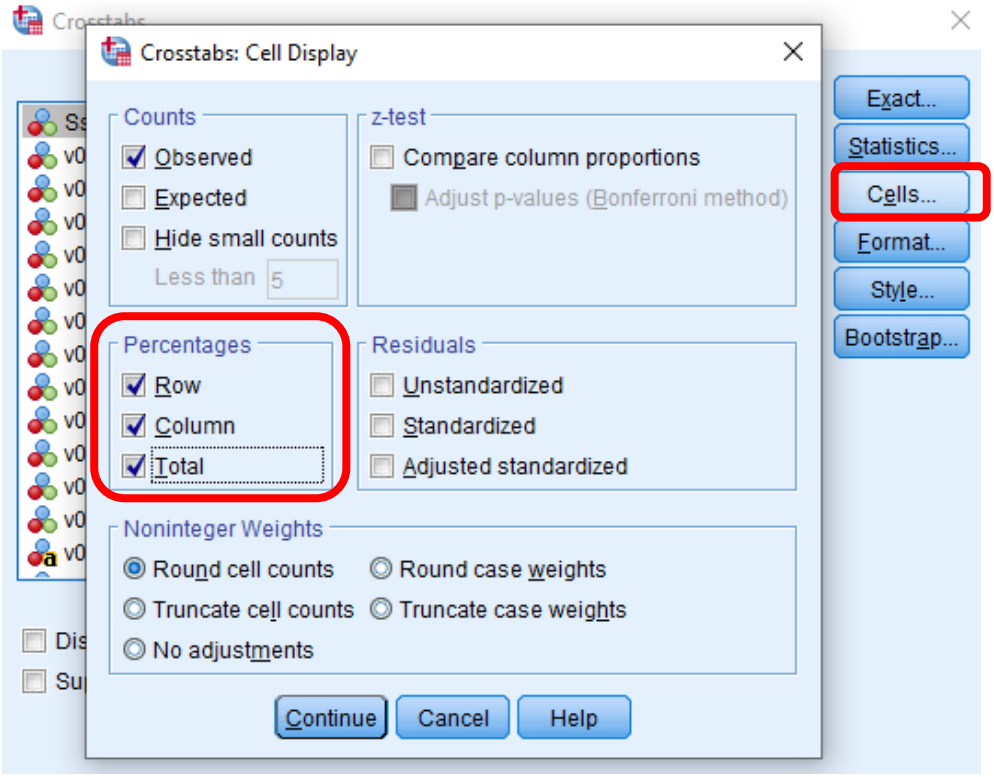
Statistics:

- ✓ Chi-Square → ipotézisvizsgálat próbája
- ✓ Phi and Cramer'V → kapcsolat szorosságának mutatója

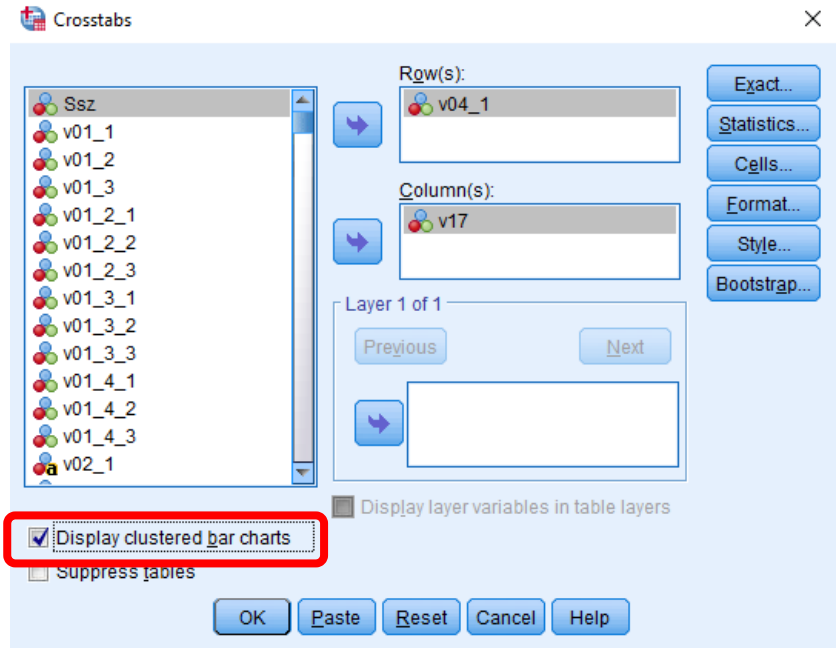


Cells:

- ✓ Percetanges (Row, Column, Total) → százalékos gyakoriságok.
  Jellemzően – amennyiben azonosítható – a független változó szerinti százalékot szokás lekérni, de bármelyik százalék értelmezhető.



Display clustered bar charts → grafikus megjelenítés



ÉRTÉKELÉS

Hipotézisvizsgálat eredménye a Chi négyzet teszt alapján

Mivel Pearson féle Chi négyzet teszthez tartozó szignifikanciaszint 0,00 (1. táblázat), ezért H0 hipotézist elvetjük, tehát szignifikáns kapcsolat van a válaszadó neme és a kedvenc alkoholos ital típusa között.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	**130,911[a]**	**5**	**,000**
Likelihood Ratio	136,896	5	,000
Linear-by-Linear	69,907	1	,000
N of Valid Cases	894		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 13,50.

Mivel a cellák kevesebb, mint 20%-a (0,0%) tartalmaz 5 válaszadónál kevesebbet, ezért az eredményeink megbízhatóak.

A kapcsolat erősségének vizsgálata

A válaszadó neme és kedvenc alkoholos ital típusa alapján a Phi és Cramer V mutatók alapján is közepes kapcsolat figyelhető meg (0,383).

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,383	,000
	Cramer's V	,383	,000
N of Valid Cases		894	

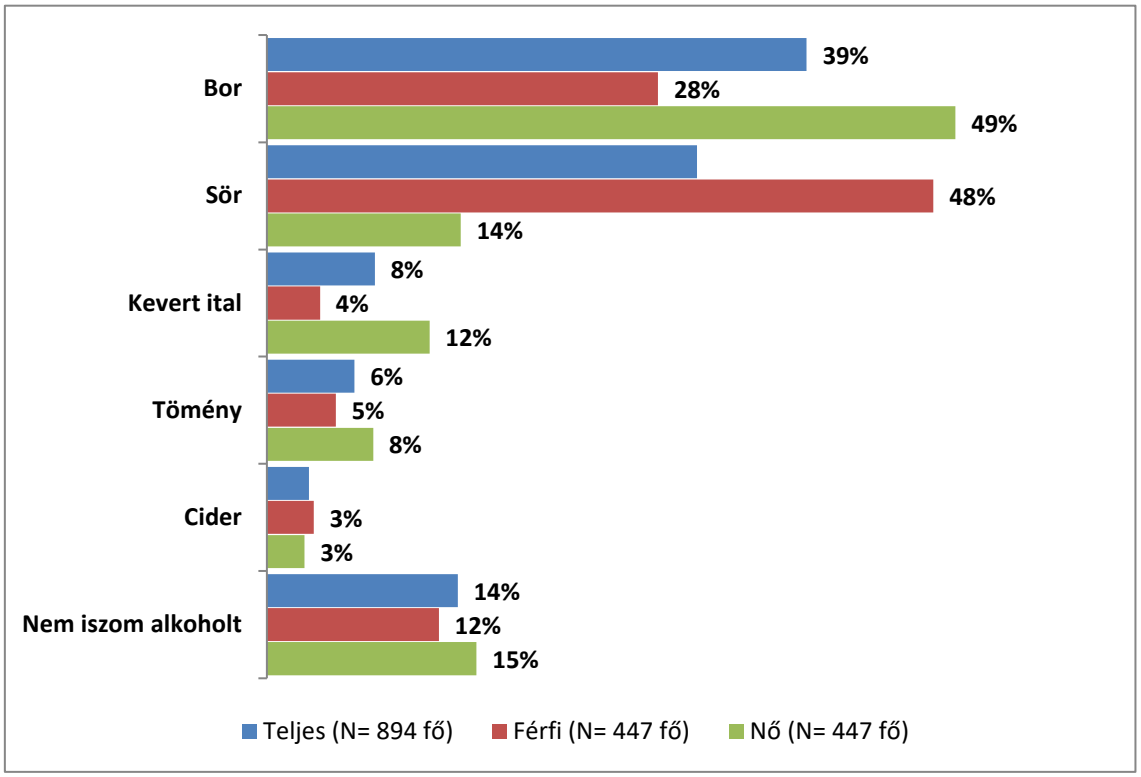
Az eredmények alapján megállapítható, hogy a férfi körében szignifikánsan magasabb a sört kedvenként megjelölők aránya (47,7%), míg a nők esetében a bort választók aránya jelentősebb (49,2%). A sört kedvelők 77,5%-a férfi, 22,5%-a nő, míg a bor esetében a nők aránya 63,8%, a férfiaké pedig 36,2%. A teljes mintában a sört kedvelő férfiak aránya 23,8%, míg a bort kedvelő nők aránya 24,6%.

A tömény és kevert italt kedvencként jelölők között is magasabb a nők aránya (60,7%, 75,4%). Az eredményeket táblázatos vagy grafikus formában is lehet ábrázolni (3. táblázat, 1. ábra). A jelen esetben bemutatott táblázatott kutatási jelentésbe nem célszerű betenni, elég vagy csak az oszlop, vagy csak a sor szerinti megoszlásokat bemutatni. Jelen esetben az értelmezés bemutatása miatt minden egyes százalékos megoszlás bemutatásra került.

A kedvenc alkoholos ital és válaszadó neme közötti kereszttábla. N=894 fő.

Melyik a kedvenc alkoholos italod (max. 2, rangsorold)_1 * Válaszadó neme Crosstabulation					
			Válaszadó neme		Teljes
			Férfi	Nő	
Melyik a kedvenc alkoholos italod (max. 2, rangsorold)_1	Sör	Count	213	62	275
		% within Melyik a kedvenc alkoholos italod (max. 2, rangsorold)_1	77,5%	22,5%	100,0%
		% within Válaszadó neme	47,7%	13,9%	30,8%
		% of Total	23,8%	6,9%	30,8%
	Cider	Count	15	12	27
		% within Melyik a kedvenc alkoholos italod (max. 2, rangsorold)_1	55,6%	44,4%	100,0%
		% within Válaszadó neme	3,4%	2,7%	3,0%
		% of Total	1,7%	1,3%	3,0%
	Bor	Count	125	220	345
		% within Melyik a kedvenc alkoholos italod (max. 2, rangsorold)_1	36,2%	63,8%	100,0%
		% within Válaszadó neme	28,0%	49,2%	38,6%
		% of Total	14,0%	24,6%	38,6%
	Tömény	Count	22	34	56
		% within Melyik a kedvenc alkoholos italod (max. 2, rangsorold)_1	39,3%	60,7%	100,0%
		% within Válaszadó neme	4,9%	7,6%	6,3%
		% of Total	2,5%	3,8%	6,3%
	Kevert ital	Count	17	52	69
		% within Melyik a kedvenc alkoholos italod (max. 2, rangsorold)_1	24,6%	75,4%	100,0%
		% within Válaszadó neme	3,8%	11,6%	7,7%
		% of Total	1,9%	5,8%	7,7%
	Nem iszom alkoholt	Count	55	67	122
		% within Melyik a kedvenc alkoholos italod (max. 2, rangsorold)_1	45,1%	54,9%	100,0%
		% within Válaszadó neme	12,3%	15,0%	13,6%
		% of Total	6,2%	7,5%	13,6%
Total		Count	447	447	894
		% within Melyik a kedvenc alkoholos italod (max. 2, rangsorold)_1	50,0%	50,0%	100,0%
		% within Válaszadó neme	100,0%	100,0%	100,0%
		% of Total	50,0%	50,0%	100,0%

A kedvenc alkoholos ital típusának megoszlása nemek szerint. N=894 fő, $N_{férfi}$=447 fő, $N_{nő}$=447 fő



SYNTAX

```
*kereszttábla.
CROSSTABS
/TABLES=V04_1 BY V17
/FORMAT=AVALUE TABLES
/STATISTICS=CHISQ PHI ETA
/CELLS=COUNT
/COUNT ROUND CELL
/BARCHART.
```

GYAKORLÓ FELADATOK

1. Vizsgáljuk meg, hogy van-e összefüggés a kedvenc alkoholmentes ital típusa (v05_1) és a válaszadó neme (v17) között?
2. Vizsgáljuk meg, hogy van-e összefüggés a háztartás havi nettó jövedelme (v24) és a származási ország között (v20)?
3. Vizsgáljuk meg, hogy van-e összefüggés aközött, hogy valaki inkább otthon vagy nem otthon fogyasztja a kedvenc alkoholos italát (v07_2_1-ból új változó egy „*Otthon fogyasztom*” és egy „*Nem otthon fogyasztom*” opcióval, ahol a Missing is a „*Nem otthon fogyasztom*” kategóriába tartozzon), és hogy milyen korcsoportba tartozik (v18-ból új változó, ahol a kategóriák: max 30, 31-50, 50+ év)?

5. VARIANCIAELEMZÉS

A varianciaelemzés célja, hogy megvizsgálja, hogy van-e különbség két vagy több csoport átlaga között. A varianciaelemzés tesztje az F-próba, melynek nullhipozétise a következő:

H0: A csoportok átlaga nem tér el szignifikánsan egymástól.

Amennyiben az F-próba eredménye alapján H0-t elvetjük, tehát van szignifikáns különbség a csoportok átlaga között, akkor annak mértékét is vizsgálni tudjuk az Eta^2 mutatóval (0 – 1).

Fontos! A varianciaelemzés feltétele a szóráshomogenitás teljesülése, mely a Levene-teszt lefuttatásával ellenőrizhető. A Levene-teszt nullhipotézise a következő:

H0: A szórásnégyzetek egyenlők.

Amennyiben a Levene-teszt H0 hipotézisét nem vetjük el, vagyis a hozzá tartozó érték magasabb, mint 0,1 (10%-os szignifikanciaszint esetén), akkor az F-próba nyugodt szívvel elvégezhető. A H0 hipotézis elvetése esetén az F-próba nem végezhető el kellő megbízhatósággal, így az eredményeink sem értelmezhetők kellő megbízhatóság mellett. Ugyanakkor fontos megemlíteni, hogy – főként nagy mintaelemszám esetén – ez nem jelent érdemi problémát, az F-teszt eredménye kellően robusztus az előfeltétel megsértésére.
A varianciaelemzés során megkülönböztetünk egyszempontos és többszempontos varianciaelemzést. Az egyszempontos varianciaelemzés során egy függő metrikus és egy független nem metrikus változó szerepel az elemzésünkben, míg a többszempontos varianciaelemzés során kettő vagy több független nem metrikus változónk van és egy metrikus függő változónk.

A varianciaelemzés során használt legfontosabb fogalmak:

- F- próba
- Levene-teszt
- Eta2

5.1. EGYSZEMPONTOS VARIANCIAELEMZÉS (egyszempontos ANOVA)

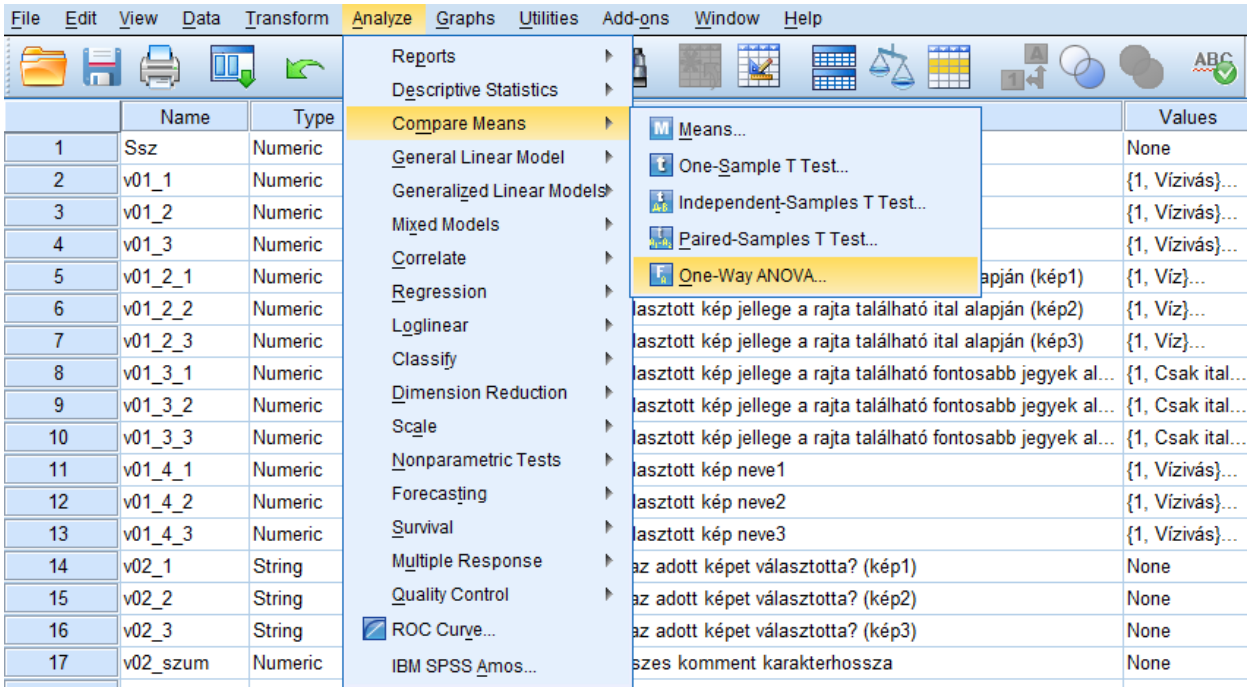
Az SPSS programon belül a varianciaelemzés során két parancsot kell lefuttatnunk. A One-Way Anova parancs futtatása során választ kapunk a Levene-teszt valamint a varianciaelemzéshez tartozó F-próba eredményéről. Szignifikáns kapcsolat esetén ennek erősségét a Means menüpont lefuttatásával tudjunk bemutatni.

FELADAT

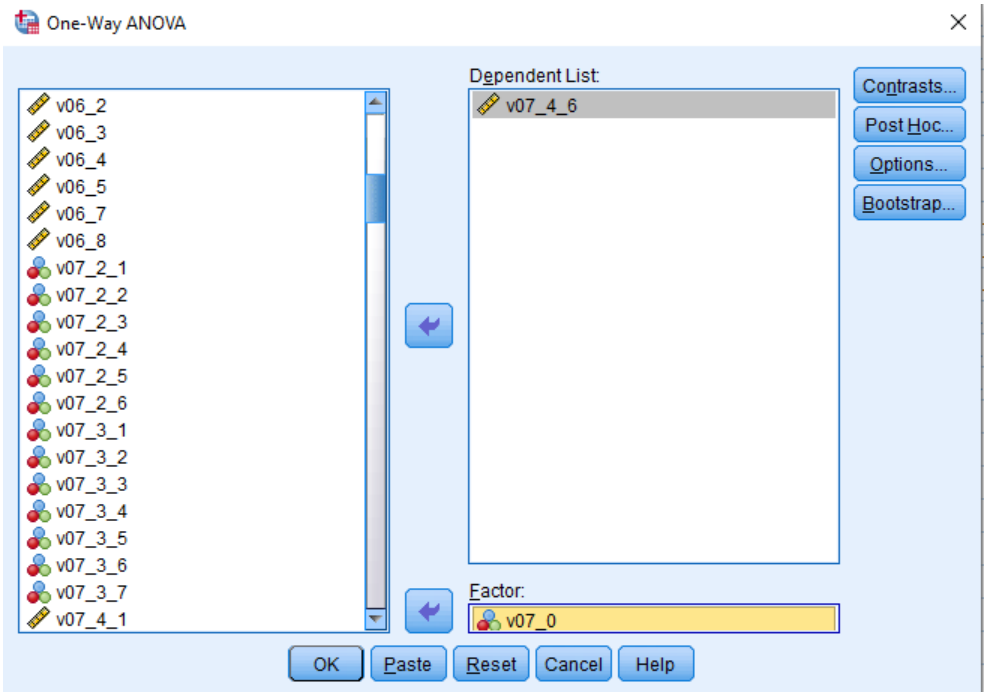
A kedvenc alkoholos italt típusa (v07_0) hatással van-e arra, hogy a megkérdezett az izgalmasság miatt választja-e az adott italt (v07_4_6)?
***Felhasznált fájlok**: Italfogyasztási szokások.sav*

MEGOLDÁS

Elérés útvonala: Analyze/Compare Means/One-Way ANOVA



Változók bevitele: Dependent list: v07_4_6, Factor: v07_0

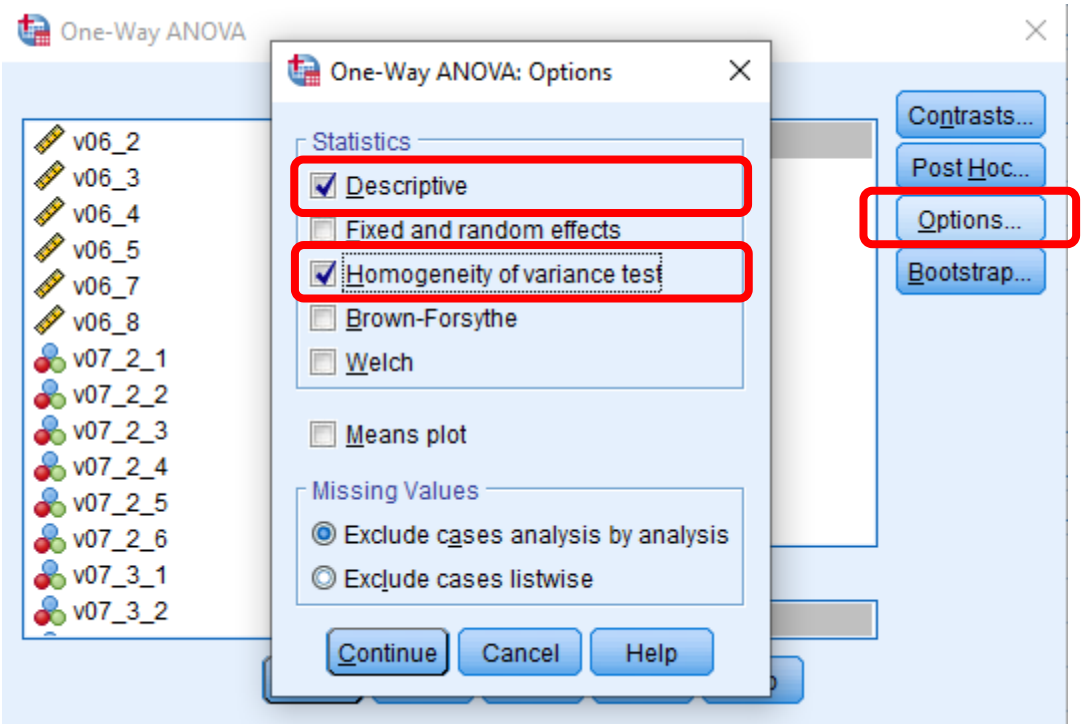


Mindig a metrikus a függő, a nem metrikus a független (faktor) változó.

Lekért adatok:

Options:

- ✓ Descriptives → független változó által képzett csoportokhoz valamint a teljes mintához tartozó leíró statisztikákra ad válasz (válaszadói számot, átlagot, szórást, minimumot, maximumot).
- ✓ Homogenity of variance test → szóráshomogenitás tesztje, Levene-teszt



ÉRTÉKELÉS

A feltételek teljesülése

A Levene-teszthez tartozó szignifikanciaszint 0,125, tehát a H0 hipotézist nem vetjük el, vagyis a szóráshomogenitás fennáll, így a varianciaelemzéshez tartozó F-próba eredménye kellő megbízhatósággal elvégezhető.

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
v07_4_6 Okok a választásra: mert izgalmas	Based on Mean	1.809	4	726	.125
	Based on Median	1.181	4	726	.318
	Based on Median and with adjusted df	1.181	4	717.450	.318
	Based on trimmed mean	1.532	4	726	.191

Az F-próba eredménye: az ANOVA táblázatban az F-próbához tartozó szignifikanciaszint 0,017, tehát a kedvenc alkoholos ital típusaihoz tartozó átlagok az italválasztás izgalmassági szintje alapján szignifikánsan különböznek.

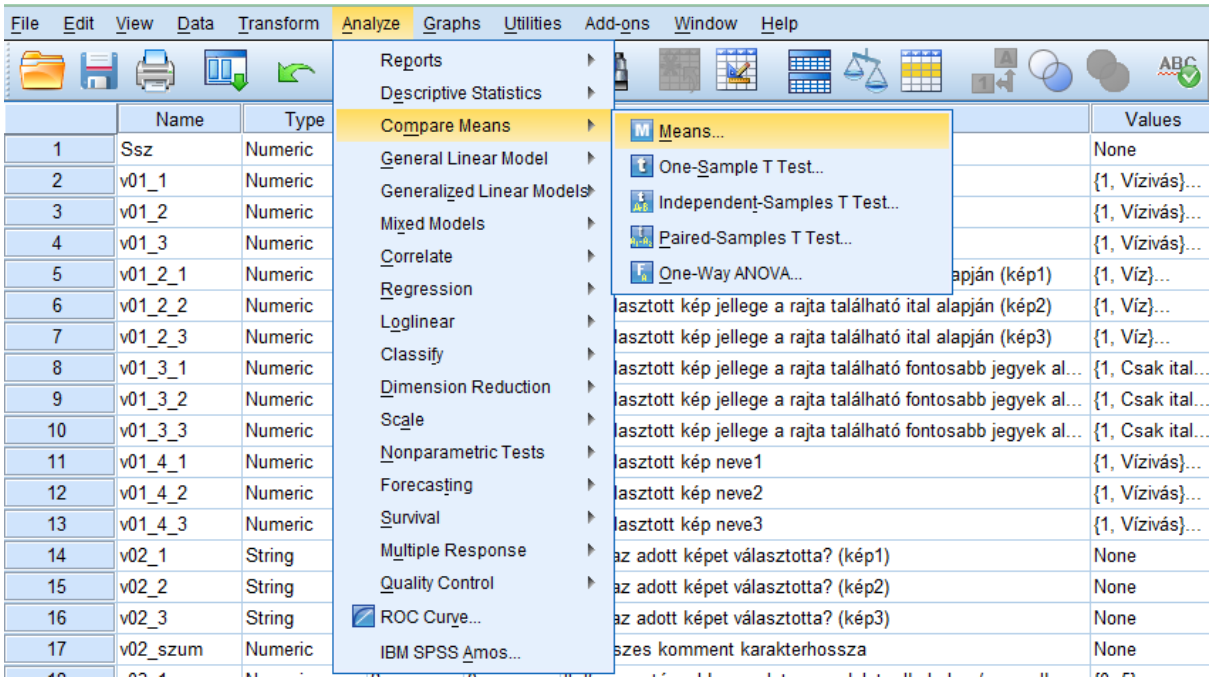
ANOVA

v07_4_6 Okok a választásra: mert izgalmas

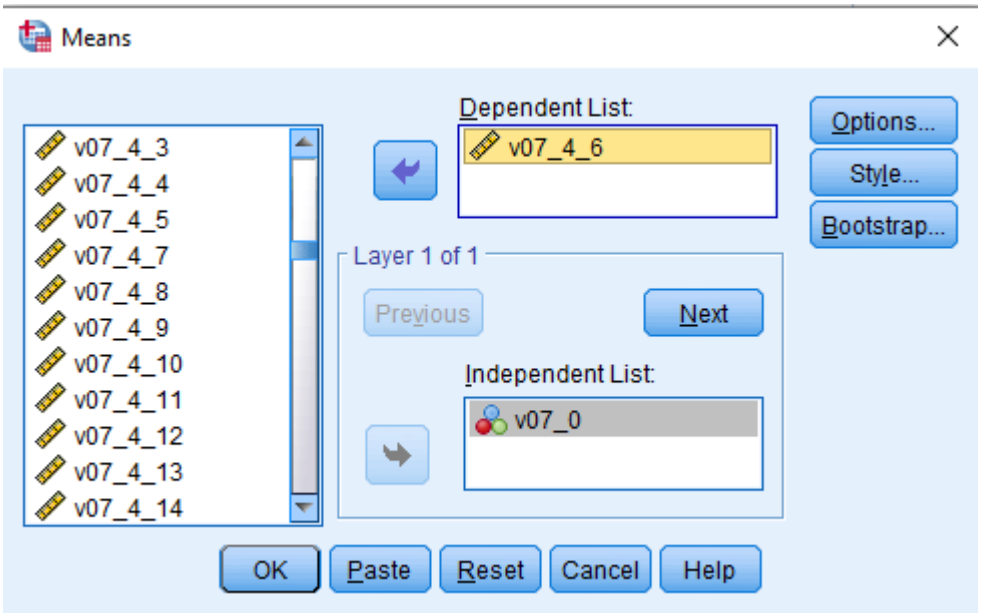
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	63.468	4	15.867	3.017	.017
Within Groups	3818.440	726	5.260		
Total	3881.908	730			

Mivel a feltételek teljesülnek és szignifikáns kapcsolat is van a vizsgálatba bevont változók között, ezért a kapcsolat erősségét is érdemes megvizsgálni.

Elérés útvonala: Analyze/Compare Means/Means



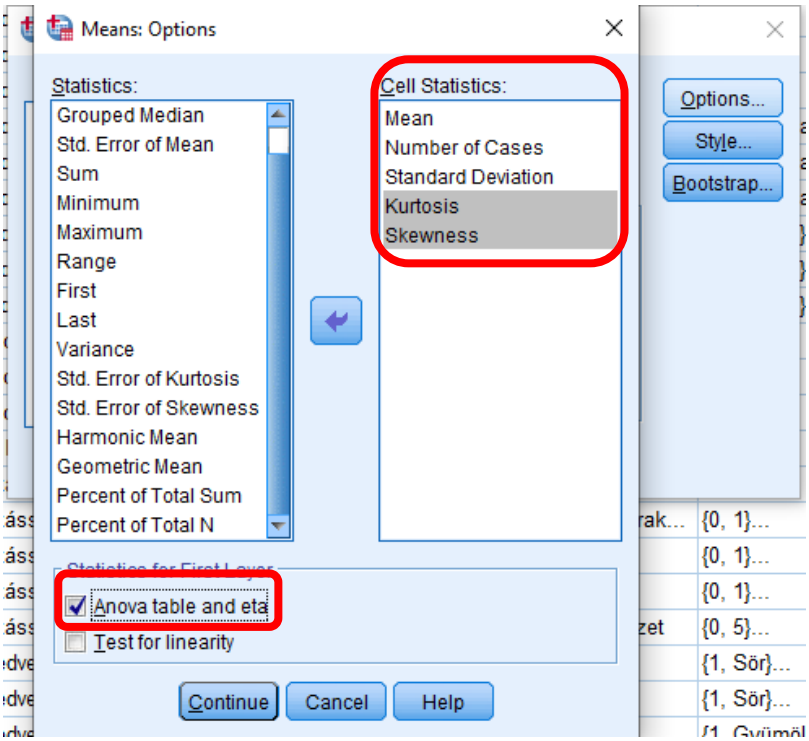
Változók bevitele: Dependent list: v07_4_6, Independent list: v07_0



Lekért adatok:

Options:

- ✓ Cell Statistics: Mean, Number of Cases, St. Devation, Skewness, Kurtosis →független változó által képzett csoportokhoz valamint a teljes mintához tartozó kijelölt leíró statisztikákra ad válasz (válaszadói számot, általagot, szórást, ferdeség, csúcsosság)
- ✓ ANOVA and Eta



ÉRTÉKELÉS

Az kedvenc alkoholos ital típusa és a választás izgalmassága közötti kapcsolat azonban gyengének tekinthető, mivel az Eta^2 mutató értéke 0,016.

Measures of Association

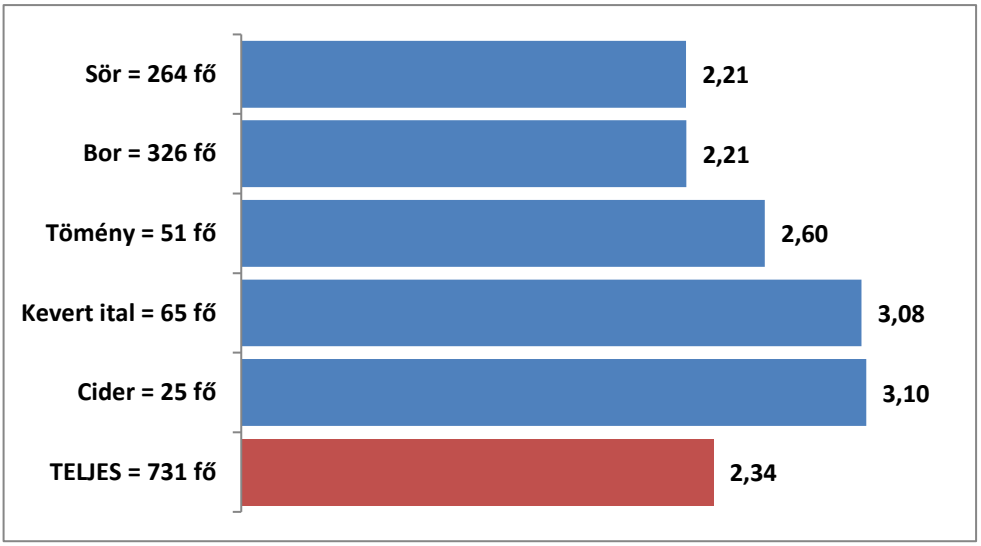
	Eta	Eta Squared
v07_4_6 Okok a választásra: mert izgalmas * v07_0 Kedvenc alkoholos ital	.128	.016

Az eredmények alapján megállapítható, hogy a cidert és kevert italokat kedvelők körében jellemző leginkább az izgalmasság miatti kedveltség: 0-7,5-ig terjedő skálán, ahol a 0 az egyáltalán értek vele egyet, a 7,5 pedig a teljes mértékben egyet értek vele (a diszkrét skála értékei 0,2,5,5,7,5) a cidert kedvencként megjelölők átlaga 3,1, míg a kevert italokat kedvencként jelölők átlaga 3,08. Legalacsonyabb egyetértés a kijelentéssel a sört és bort kedvelők körébe volt. Mindkét esetben az átlag 2,21. A teljes mintában a kijelentéssel való egyetértés átlagosan 2,34. Az eredményeket táblázatban vagy grafikusan is lehet ábrázolni.

Az izgalmasságot fogyasztási okként jelölések átlaga a különböző kedvenc alkoholos ital csoportokon belül (0-7,5 skála, 0 –egyáltalán nem értek egyet, 7,5 – teljes mértékben egyetértek)**.** N = 731 fő.

Kedvenc alkoholos ital	**Átlag**	N (fő)	Szórás	Ferdeség	Csúcsosság
Sör	**2,21**	264	2,231	,589	-,709
Cider	**3,10**	25	2,727	,317	-1,173
Bor	**2,21**	326	2,243	,643	-,577
Tömény	**2,60**	51	2,396	,487	-,773
Kevert ital	**3,08**	65	2,530	,168	-1,151
Total	**2,34**	731	2,306	,566	-,733

Az izgalmasságot fogyasztási okként jelölések átlaga a különböző kedvenc alkoholos ital csoportokon belül (0-7,5 skála, 0 –egyáltalán nem értek egyet, 7,5 – teljes mértékben egyetértek). N = 731 fő



SYNTAX

```
*Varianciaelemzés + Levene-teszt.
ONEWAY V07_4_6 BY V07_0
/STATISTICS DESCRIPTIVES HOMOGENEITY
/MISSING ANALYSIS.

*Varianciaelemzés + eta mutató.
MEANS TABLES=V07_4_6 BY V07_0
/CELLS MEAN COUNT STDDEV SKEW KURT
/STATISTICS ANOVA.
```

GYAKORLÓ FELADATOK

1. A heti szinten átlagosan elfogyasztott sör mennyiségére (v06_2) hatással van-e a válaszadó neme (v17)?
2. A különböző foglalkozásúak (v22) heti szinten elfogyasztott kávémennyisége különbözik-e (v06_7)?
3. Megfigyelhető-e szignifikáns különbség az idős és fiatal (v18 – új változó – kategóriák: max 45, 45+ év) sört kedvelők (v07_0) között a heti alkoholos italra költött összeg mennyiségében (v10_szum1)?

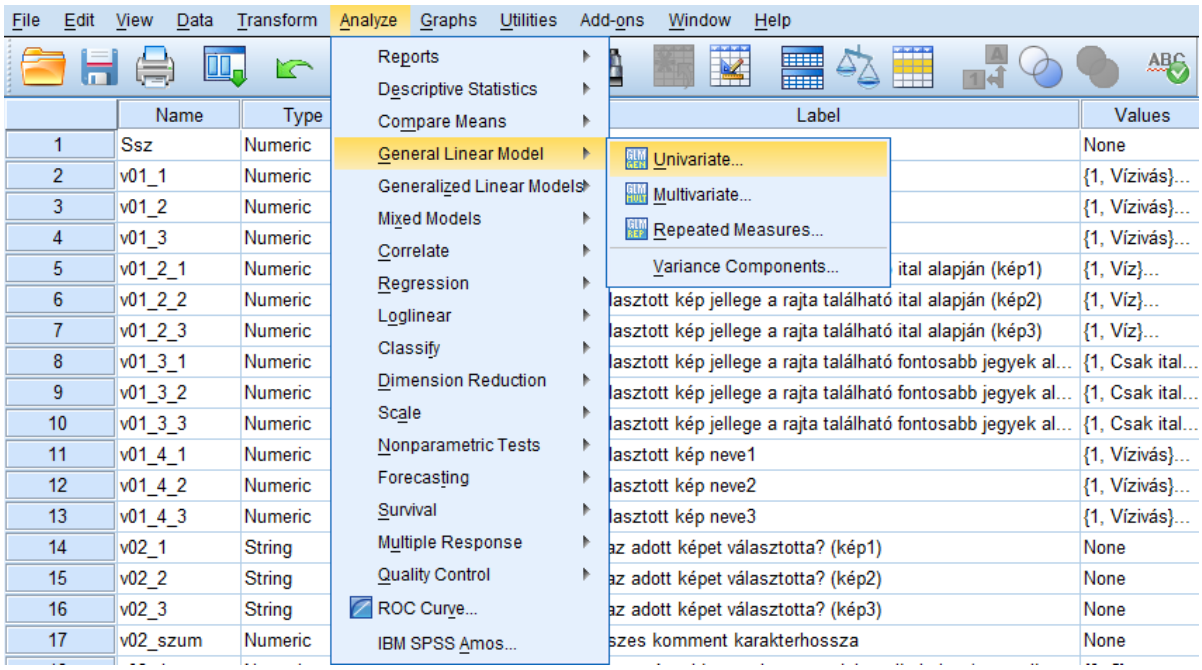
5.2. TÖBBSZEMPONTOS VARIANCIAELEZMÉS (többszempontos ANOVA)

FELADAT

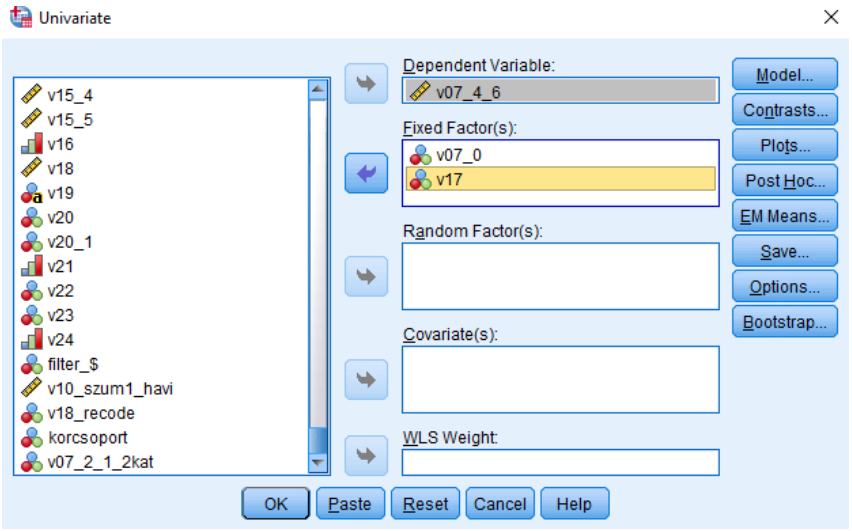
A kedvenc alkoholos italt típusa (v07_0) és a válaszadó neme (v17) hatással van-e arra, hogy a megkérdezett az izgalmasság miatt választja-e az adott italt (v07_4_6)?
***Felhasznált fájlok**: Italfogyasztási szokások.sav*

MEGOLDÁS

Elérés útvonala: Analyze/General Linear Model /Univariate



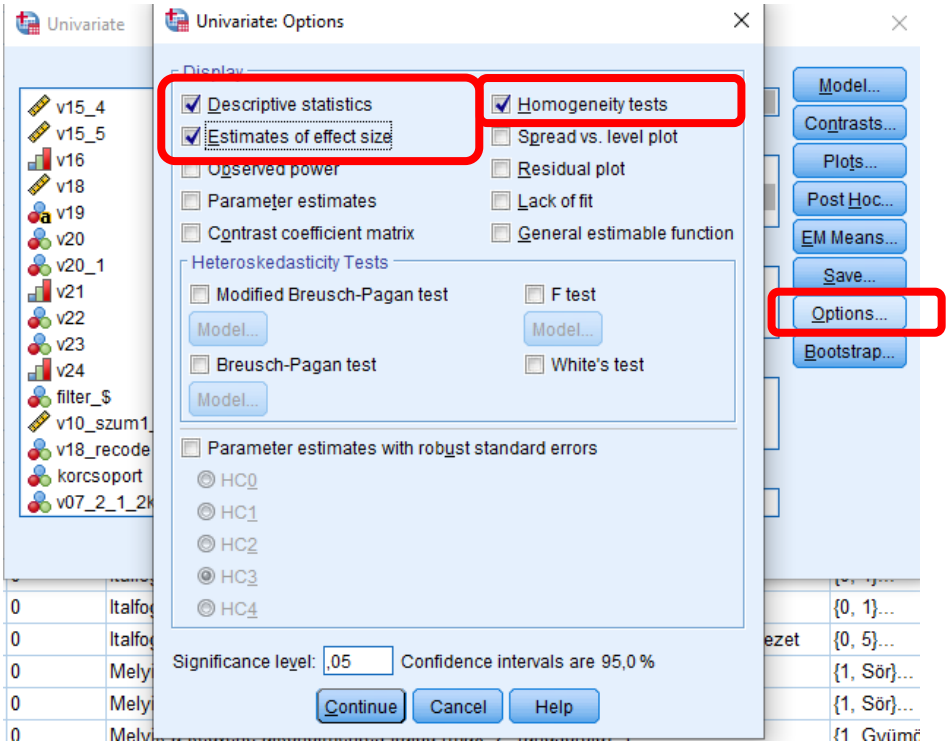
Változók bevitele: Dependent list: v07_4_6, Fix factor(s): v07_0, v17



Lekért adatok:

Options

- ✓ Descriptive Statistics → a független változók által képzett csoportokhoz valamint a teljes mintához tartozó leíró statisztikákra ad válasz (válaszadói számot, átlagot, szórást, minimumot, maximumot).
- ✓ Estimate of effect size → kapcsolat szorosságának mutatóját adja meg (Eta^2)
- ✓ Homogenity test → szóráshomogenitás tesztje, Levene-teszt



ÉRTÉKELÉS

A feltételek teljesülése: A Levene-teszthez tartozó szignifikanciaszint 0,377, tehát a H0 hipotézist nem vetjük el, vagyis a szóráshomogenitás fennáll, így a varianciaelemzéshez tartozó F-próba eredménye kellő megbízhatósággal elvégezhető.

Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
v07_4_6 Okok a választásra: mert izgalmas	Based on Mean	1.077	9	721	.377
	Based on Median	.814	9	721	.603
	Based on Median and with adjusted df	.814	9	707.462	.603
	Based on trimmed mean	.909	9	721	.517

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Dependent variable: v07_4_6 Okok a választásra: mert izgalmas
b. Design: Intercept + v07_0 + v17 + v07_0 * v17

Az F-próba eredménye: A többszempontos varianciaelemzéshez tartozó F-próba alapján elmondhatjuk, hogy amennyiben az izgalmasság való egyetértést, mint választási okkal való egyetértést vizsgáljuk a válaszadók nemét és kedvenc alkoholos italát figyelembe véve, akkor sem a nem, sem a kedvenc alkoholos ital típusa, sem pedig ezek együttesen nem befolyásolják a kijelentéssel való egyetértés szintjét, mivel minden esetben az F-próbához tartozó szignifikanciaszint magasabb, mint 0,05. (v07_0 sig=0,120, v17 sig=0,929, v07_0*v17 sig=0,673).

Tests of Between-Subjects Effects
Dependent Variable: v07_4_6 Okok a választásra: mert izgalmas

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	78.919[a]	9	8.769	1.662	.094	.020
Intercept	1852.778	1	1852.778	351.264	.000	.328
v07_0	38.715	4	9.679	1.835	.120	.010
v17	.042	1	.042	.008	.929	.000
v07_0 * v17	12.362	4	3.091	.586	.673	.003
Error	3802.989	721	5.275			
Total	7893.750	731				
Corrected Total	3881.908	730				

a. R Squared = .020 (Adjusted R Squared = .008)

Amennyiben bármely változó szerinti szignifikáns kapcsolat állna fenn, annak erősségét a Partial Eta Squared oszlop mutatná meg.

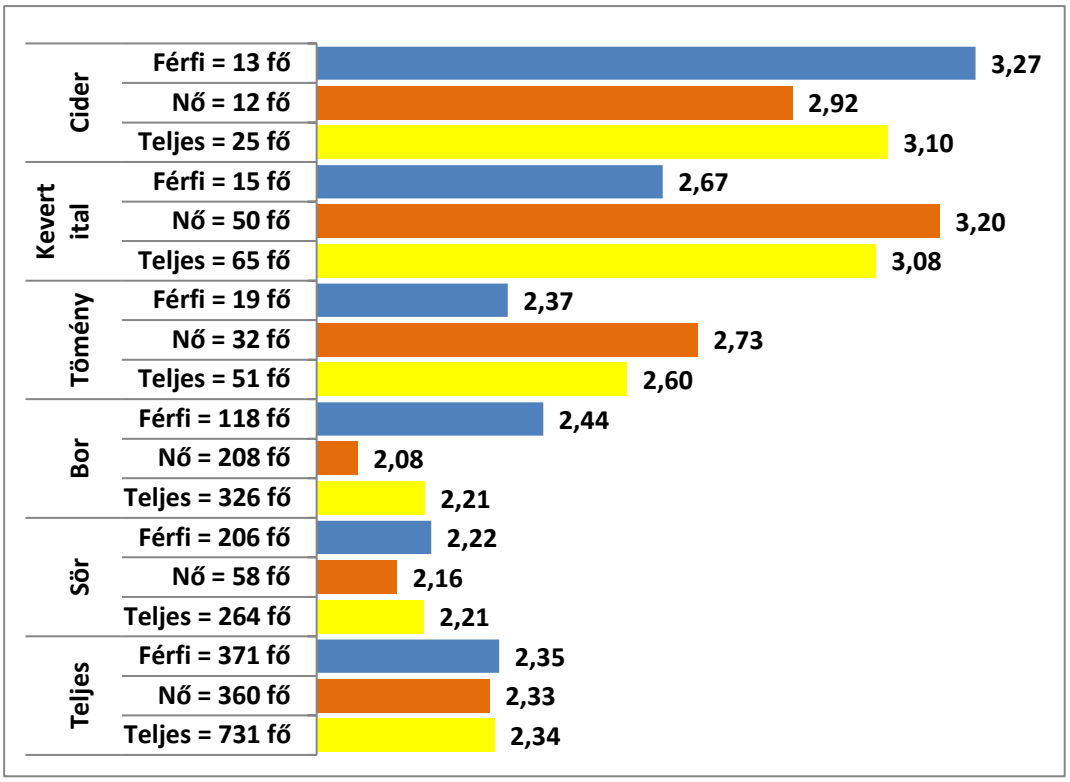
Habár nincs szignifikáns különbség a csoportok között, még az eredményeket tudjuk értelmezni. Az izgalmasság, mint választási ok leginkább a cidert kedvelő férfiakat (átlag=3,27) valamint a kevert italt kedvelő nőket (átlag=3,2) jellemzi. Az egyetértés

mértékének mérésére egy 0-7,5-ig terjedő skála került használtra, ahol a 0 az egyáltalán nem értek vele egyet, a 7,5 pedig a teljes mértékben egyet értek vele szintet jelöli. A diszkrét skála értékei 0,2,5,5,7,5. Az eredményeket ismét táblázatban, vagy grafikusan is tudjuk ábrázolni.

Az izgalmasságot fogyasztási okként jelölések átlaga a különböző kedvenc alkoholos ital és nemek szerinti csoportokon belül (0-7,5 skála, 0 –egyáltalán nem értek egyet, 7,5 – teljes mértékben egyetértek). N = 731 fő.

Kedvenc alkoholos ital		**Átlag**	**Szórás**	**N (fő)**
Sör	**Férfi**	2,22	2,21	206
	Nő	2,16	2,32	58
	Teljes	2,21	2,23	264
Cider	**Férfi**	3,27	2,77	13
	Nő	2,92	2,79	12
	Teljes	3,10	2,73	25
Bor	**Férfi**	2,44	2,28	118
	Nő	2,08	2,22	208
	Teljes	2,21	2,24	326
Tömény	**Férfi**	2,37	2,28	19
	Nő	2,73	2,49	32
	Teljes	2,60	2,40	51
Kevert ital	**Férfi**	2,67	2,40	15
	Nő	3,20	2,58	50
	Teljes	3,08	2,53	65
Teljes	**Férfi**	2,35	2,26	371
	Nő	2,33	2,35	360
	Teljes	2,34	2,31	731

Az izgalmasságot fogyasztási okként jelölések átlaga a különböző kedvenc alkoholos ital és nemek szerinti csoportokon belül (0-7,5 skála, 0 –egyáltalán nem értek egyet, 7,5 – teljes mértékben egyetértek)**.** N = 731 fő.



SYNTAX

```
*Többszempontos variancia-elemzés.
UNIANOVA V07_4_6 BY V07_0 V17
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/EMMEANS=TABLES(OVERALL)
/PRINT=ETASQ HOMOGENEITY DESCRIPTIVE
/CRITERIA=ALPHA(.05)
/DESIGN=V07_0 V17 V07_0*V17.
```

GYAKORLÓ FELADATOK

1. A kedvenc alkoholos italt típusa (v07_0) és válaszadó neme (v17) hatással van-e arra, hogy a megkérdezett a családi hagyomány miatt választja az adott italt (v07_4_12)?

6. GYAKROLÓ ESETEK: leíró statisztika, kereszttábla, varianciaelemzés

1. Az AGYŐ pálinkafőzde tavaly alakul Magyarországon, és tervezi, hogy a nemzetközi piacra is szeretne betörni. Az alapító két jóbarátnak több kapcsolata is van Angliában, ezért először ott szeretne belépni a piacra. A döntése meghozatala előtt azonban szeretné megismeri a kinti alkohol fogyasztási szokásokat, különös figyelmet szánva a töményital fogyasztásnak. A kutatás során életkor szerint az alábbi korcsoportbontás érdekes a márka számára (1) 30 év vagy annál fiatalabb, (2) 31-50 év, (3) 51 év vagy annál idősebb. A kutatási reportban a megbízó a következő kérdésekre szeretne választ kapni:
1) Melyek az angolok kedvenc alkoholos itala? Az érintett korcsoportok szerint megfigyelhető-e különbség? Van-e különbség a férfiak és nők között?
2) Mennyi tömény italt fogyasztanak az angolok egy átlagos héten? Az érintett korcsoportok szerint megfigyelhető-e különbség? Van-e különbség a férfiak és nők között? A korcsoport és a válaszadók neme együttesen vizsgálva befolyásolják az elfogyasztott mennyiséget?
3) Akiknek a kedvenc italuk a tömény, mennyi pénzt költenek rövidre átlagosan egy héten?

2. A Kacagás borászat 10 éve alakult Magyarországon, és 5 piacon is jelen vannak (Magyarország, Ausztria, Németország, Írország, Anglia). Az alapítók szeretnének egy új márkát bevezetni az olcsóbb borok szegmensében mind az 5 országban, és ehhez szeretnék megvizsgálni azon emberek szokásait, akik átlagosan hetente 0-15 eurót költenek borra (v10_1). A kutatás alapján a következő kérdésekre szeretnének választ kapni.
1) Átlagosan mennyit költenek alkoholra a célcsoport tagjai? (v10_szum1)
2) Mennyi az átlagos teljes italköltés (alkoholos és alkoholmentes összesen) egy átlagos héten (v10_szum1, v10_szum2)?
3) Hogyan jellemezné a 0-15 eurót borra költőket a kedvenc alkoholos italuk alapján (v07_0)?
4) A célcsoporton belül van-e szignifikáns kapcsolat a válaszadók neme és kedvenc italuk között? (v17, v07_0)
5) Van-e különbség az alkoholra költött összegben a 40 év alattiak, és a legalább 40 éves esetében (v10_szum1, v18)?
6) Van-e különbség a célcsoport férfi és nő tagjai között az alapján, hogy a bort kedvenc italként azért választják, mert ez családi hagyomány? (v17, v08_4_12)

3. A Zwack szeretné megismerni a tömény italt nem fogyasztók italfogyasztási szokásait (v06_5). A kutatás megkezdése előtt az alábbi kérdéseik merültek fel
1) Jellemezze a célcsoportot a kedvenc alkoholos itala alapján és az alapján, hogy az átlagos héten mennyit költenek alkoholra (v07_0, v10_szum1).
2) A megbízó a célcsoportot 2 alcsoportra bontja: az alkoholos italra nem költők, és az alkoholos italra költőkre. Megfigyelhető-e különbség a célcsoport esetében a különböző alcsoportba tartozó egyének között
3) Megfigyelhető-e összefüggés a célcsoport esetében a válaszadó kedvenc alkoholos itala és aközött, hogy melyik alcsoportba tartozik? (v07_0, v10_szum1 - új).

4. A Szentkirályi szeretné megismerni a palackozott vizet fogyasztó férfiak italfogyasztási szokásait (v06_3. v17). A kutatás megkezdése előtt az alábbi kérdéseik merültek fel.

1) Jellemezze a célcsoportot a kedvenc alkoholmentes itala alapján és az alapján, hogy egy átlagos héten mennyit költenek alkoholmentes italra. (v08_0, v10_szum2).

2) A megbízó a célcsoportot 3 alcsoportra bontja: az ásványvizet kedvencnek tartók, a csapvizet kedvencnek tartók és az egyéb alkoholmentes italokat kedvencnek tartókra. Megfigyelhető-e összefüggés a célcsoporton belül a származási ország és aközött, hogy melyik alcsoportba tartozik valaki? (v20, v08_0 - új).

3) Megfigyelhető-e különbség a célcsoport esetében a különböző alcsoportokba tartozó egyének között az átlagos életkor tekintetében? (v08_0 - új, v18).

5. A Segafredo a 60 év feletti, férfi korosztályban szeretne egy promóciós kampányt megvalósítani (v17, v18), és ehhez egy kutatást tervez. A kampány megtervezéséhez az alábbi információkra van szüksége a cégnek:

1) Hogyan jellemezhető a célcsoport a kedvenc alkoholmentes ital alapján (v08_0)?

2) Van-e szignifikáns különbség a heti átlagos alkoholköltési mennyiségben a különböző kedvenc alkoholmente italt választók esetében? (v10_szum2, v08_8)

3) Hogyan jellemezhető a célcsoport NAPI kávéfogyasztása? (új változó v06_7-ből)

4) Az előbb létrehozott új változót, vagyis a napi szinten elfogyasztott kávémennyiséget soroljuk két csoportban az alábbi feltételek szerint: átlagosan legfeljebb 1 csésze kávé/nap, átlagosan legalább 1,01 csésze kávé/nap. A két csoport megoszlása szignifikánsan különbözik-e a származási ország alapján (v20)?

7. LINEÁRIS REGRESSZIÓS ELEMZÉS

A regresszióelemzés célja, hogy egy vagy több független változó függő változó értékére való hatását meg tudjuk becsülni.

- A függő változó mindig metrikus változó.
- A független változók lehetnek metrikus és nem metrikus (ún. dummy) változók is, azonban jelen segédanyag keretén belül mindössze a metrikus változókkal foglalkozunk.

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

$$\mathbf{Y = \beta_0+\beta_1X_1+\beta_2X_2+...+\beta_kX_k+u,}$$

ahol

Y – a függő változó (vagy eredményváltozó),
$X_1, X_2, \ldots, X_k$ – a független változók (vagy magyarázóváltozók),
β_0 egy konstans,
$\beta_1, \beta_2, \ldots, \beta_k$ – a független változókhoz tartozó regressziós együtthatók.
u – az ún. reziduum, vagy hibatag, a megfigyelések azon része, amelyet a regressziós modell nem magyaráz meg. Azaz a valós értékek regressziós modelltől való eltérései.

A **regressziós együtthatók** megmutatják a független változókban bekövetkező változások függő változóra gyakorolt hatásait, minden egyéb tényező változatlansága mellett (*ceteris paribus*).

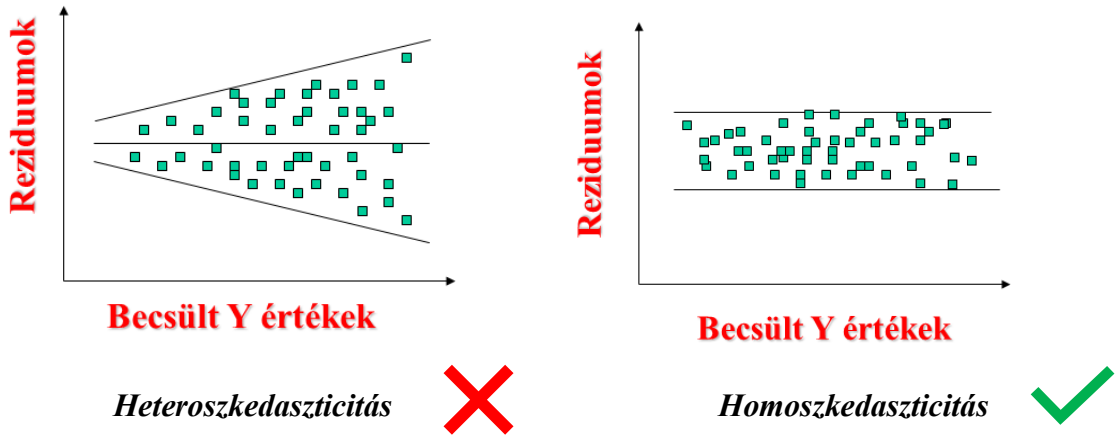
β_0 a felírni kívánt egyenes **Y tengellyel alkotott metszéspontját** mutatja,
β_1 az egyenes meredekségét adja meg.

Egyváltozós regresszió esetén egy független változóval, míg a **többváltozós regresszió** során több független változóval magyarázzuk a függő változót.

A lineáris regressziónak hat **feltétele** van**:**

1. ***Linearitás***, vagyis a változók közötti lineáris kapcsolat (amelyet a modell felírásakor feltételezünk) a valóságban is megállja a helyét.
2. ***Véletlen minta***, vagyis az adatfelvétel során az alapsokaságból véletlenszerűen kerültek kiválasztásra a kitöltők, minden egyednek ugyanakkor volt a bekerülési valószínűsége.
3. ***Tökéletes multikollinearitás hiánya***, vagyis a független (magyarázó) változók között nincs lineáris kapcsolat (+1 vagy -1 korreláció). Ugyanakkor a független változók közötti erős, de még nem tökéletes multikollinearitás is problémás lehet.
4. ***Exogenitás,*** vagyis a független változók és az elméleti hibatag (u) között ne legyen kapcsolat (korreláció). Ez lényegében azt jelenti, hogy a modell teljesnek tekinthető, minden fontos változót figyelembe vettünk. Ezt a feltételt nem tudjuk tesztelni, logikai alapon lehet következtetni a teljesülésére, illetve nem teljesülésére.
5. ***Homoszkedaszticitás***, vagyis a reziduumok szórása legyen állandó, a magyarázó változók értékétől független. Ha ez nem teljesül, akkor heteroszkedaszticitásról beszélhetünk. A vizsgálat a standardizált reziduum értékek és a standardizált függő változó értékek ábrázolásával történik (bár formális tesztek is létetznek). Heteroszkedaszticiátás esetén az ábrázolt standardizált reziduumok nem foghatók közre két párhuzamos egyenessel.

A standardizált reziduum értékek és a standardizált függő változó értékek ábrája



6. ***A reziduumok normális eloszlást követnek.*** Nagy mintáknál (n > 100) a centrális határeloszlási tétel alapján ezt a feltételt adottnak vesszük.

<u>A regresszióelemzés során jellemzően két próbát, két hipotézisvizsgálatot kell alkalmaznunk.</u>

Az *F*-próba célja, hogy megvizsgáljuk, hogy a feltételezett modell létezik-e, fennáll-e az összefüggés (azaz van-e kapcsolat) a függő és független változók között.
H0: $\beta_1 = \beta_2 = ... = \beta_k = 0$, vagyis nincs összefüggés a függő és független változók között.
(A modell nem szignifikáns, minden becsült paramétere 0.)
(Grafikusan ez azt jelenti, hogy a pontokra illesztett egyenes vízszintes, az X tengellyel párhuzamos)
H1: fennáll az összefüggés a függő és független változók között. (A modell szignifikáns, legalább egy becsült paraméter szignifikánsan eltér 0-tól.)

A *t*-próba célja megvizsgálni, hogy egy adott regressziós együttható különbözik-e nullától (azaz szignifikáns-e a modellben), vagyis szükség van-e az adott magyarázó változóra a regresszióban.
Ho: $\beta_j = 0$, vagyis a regressziós együttható nem különbözik szignifikánsan 0-tól.
H1: $\beta_j \neq 0$, vagyis a regressziós együttható szignifikánsan különbözik 0-tól.

FONTOS! Amennyiben egy együttható a modellben nem lesz szignifikáns, azaz a *t*-próba *p*-értéke (vagyis szignifikancia szintje) a választott szignifikanciaszinthez $(1 - \alpha)$ tartozó α-nál (általában 5%-nál) nagyobb értéket vesz fel, abban az esetben az adott együtthatóhoz tartozó magyarázó változónak nincs szignifikáns hatása a függő változóra, így azt el kell távolítani a modellből, és újra kell futtatni a modellt enélkül a magyarázó változó nélkül.

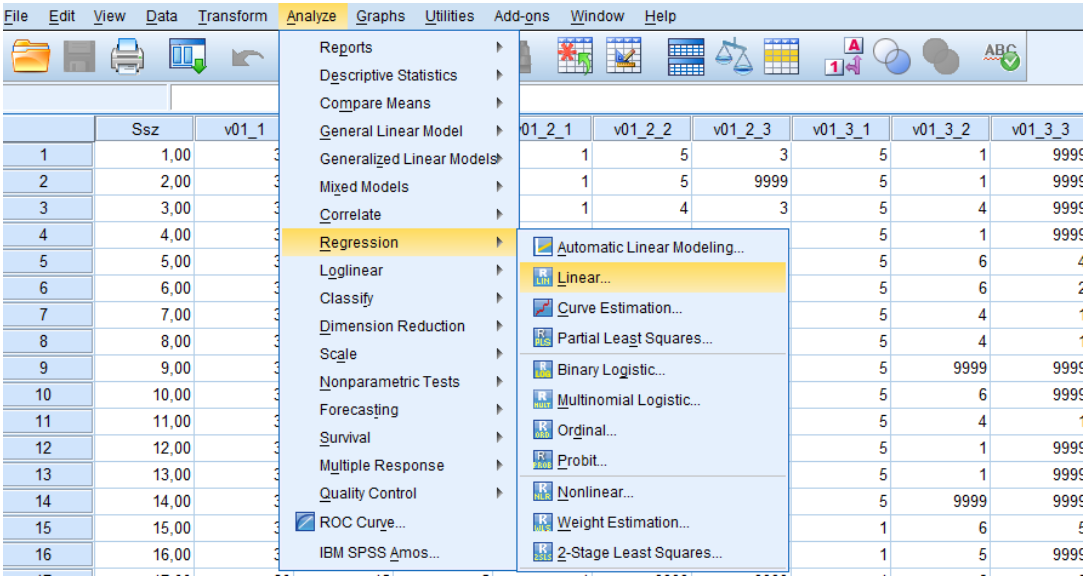
FELADAT

Hány pohár bort iszik meg egy átlagos héten egy 30 éves fiatal, aki átlagosan 20 eurót költ egy héten alkoholos italra? (v06_1, v18, v10_szum1)
***Felhasznált fájlok**: Italfogyasztási szokások.sav*

MEGOLDÁS

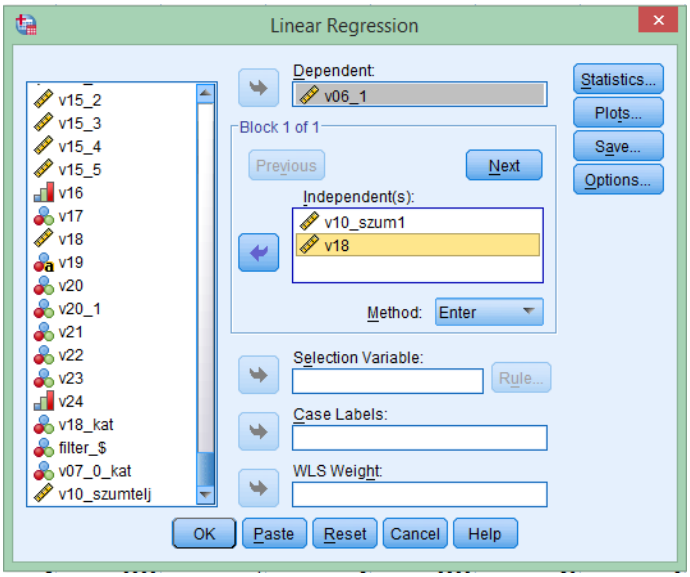
Mivel több független (magyarázó) változónk van, azért ez egy többváltozós regresszió lesz. Egyváltozós regresszió esetén a lépések ugyanezek lennének, de nem kell multikollinearitást vizsgálni.

Elérés útvonala: Analyze/ Regression/ Linear *(az egyváltozós regresszió is ugyanígy működik, és ugyanazokat a paramétereket kell beállítani)*



Változók bevitele: Dependent: v06_1, Independent(s): v10_szum1, v18

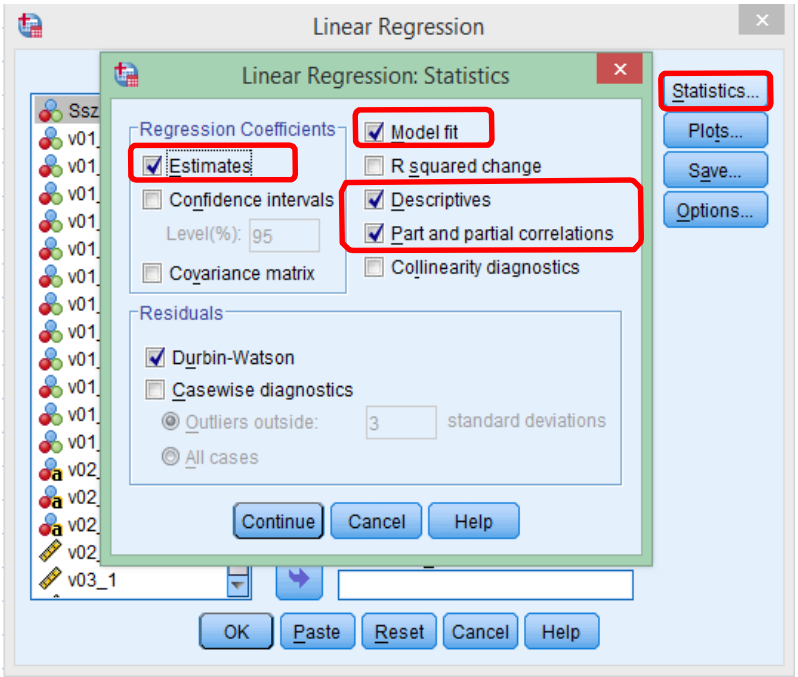
Azaz a pohár borok száma a függő változó (eredményváltozó), és az életkor és a heti átlagos italköltés pedig a függő (magyarázó) változók.



Lekért adatok:

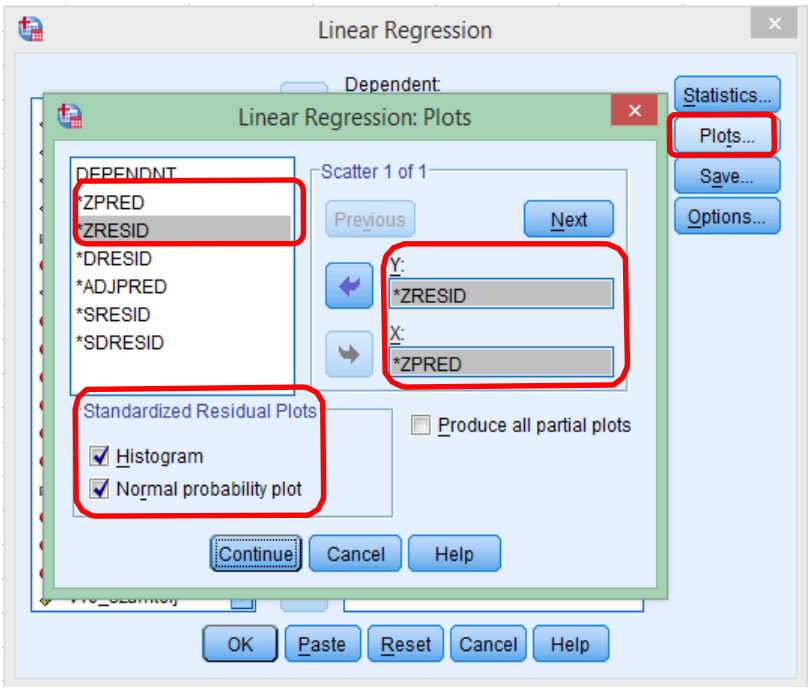
Statistics:

- ✓ Estimates → regressziós együtthatók, *t*-próba
- ✓ Model fit → F-próba
- ✓ Descriptives
- ✓ Part and partial correllation → Multikollinearitás miatt



Plots:

- ✓ X: ZPRED, Y:ZRESID → homoszkedaszticitás vizsgálatára
- ✓ Standardized residual plots: → reziduumok normál eloszlásának vizsgálatára
- ✓ Histogram
- ✓ Normal probability plot



ÉRTÉKELÉS:

FELTÉTELEK TELJESÜLÉSE

a. Tökéletes multikollineritás

A korrelációs táblában meg kell nézni, hogy a független változók között milyen erős korreláció áll fenn (zöld terület), illetve az adott korrelációk szignifikánsak-e. **Amennyiben az átlón kívüli értékek közül a magyarázó változók között 0,7-nél nagyobb korrelációs értéket találunk, akkor a független változók között már erősnek tekinthető összefüggés van.** Ilyenkor érdemes átgondolni, hogy mindegyik magyarázó változót meg akarjuk-e tartani a modellben (lehetséges, hogy a magas, bár nem tökéletes multikollinearitás ellenére nem távolítunk el egyetlen változót sem a modellből).

Az eredmények alapján megállapítható, hogy az átlagos alkoholos költés és a válaszadó kora között gyenge, pozitív (0,034), nem szignifikáns (Sig = 0,168) kapcsolat áll fenn. Ez alapján nincs jelentős mértékű multikollinearitás a modellben. Mivel nincs tökéletes multikollinearitás a modellben, ezért a feltétel teljesül.

Az átlagos borfogyasztás és az alkoholos italköltés közötti kapcsolat közepes, pozitív (0,339), szignifikánsnak mondható (Sig = 0,000), míg az átlagos borfogyasztás és a válaszadó kora között gyenge, pozitív (0,194), szignifikáns kapcsolat van.

Az eredményváltozó és az első magyarázó változó közötti korreláció mértéke, azaz r = 0,339 (*p* = 0,000)

Az eredményváltozó és a második magyarázó változó közötti korreláció mértéke, azaz r = 0,194 (*sig* = 0,000)

Correlations

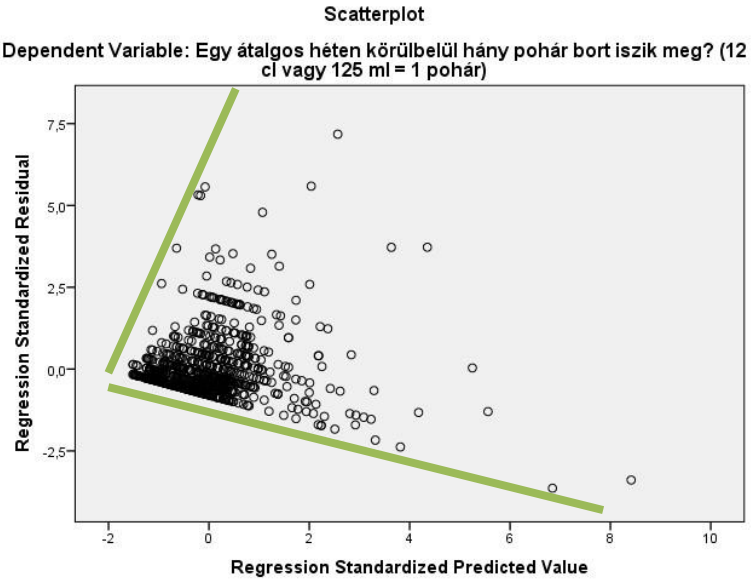
		Egy átlagos héten körülbelül hány por… ort iszik meg? (12 cl vagy 125 ml = 1 pohár)	Átlagos alkoholos költés egy héten	…sza …ora
Pearson Correlation	Egy átlagos héten körülbelül hány pohár bort iszik meg?	***1,000***	***,339***	***,194***
	Átlagos alkoholos költés egy héten	***,339***	***1,000***	***,034***
	Válaszadó kora	***,194***	***,034***	***1,000***
Sig. (1-tailed)	Egy átlagos héten körülbelül hány pohár bort iszik meg?		***,000***	***,000***
	Átlagos alkoholos költés egy	***,000***		***,168***
	Válaszadó kora	***,000***	***,168***	
N	Egy átlagos héten körülbelül hány pohár bort iszik meg?	814	814	814
	Átlagos alkoholos költés egy	814	814	814
	Válaszadó kora		814	814

Ez a két magyarázó változó közötti korrelációhoz tartozó *p*-érték, azaz *sig* = 0,168

Ez a két magyarázó változó közötti korreláció értéke, azaz r = 0,034

b. Homoszkedaszticitás

Az ábra alapján megállapítható, hogy a *reziduumok nem foghatók közre két párhuzamos egyenessel*, tehát a modell hibatagjai *heteroszkedasztikusak*. Ennek ellenére az elemzést elvégezzük, mivel a heteroszkedaszticitás következménye „mindössze" a magyarázóerő csökkenése, tehát ennek a feltételnek a nem teljesülését megtűrjük.



c. A reziduumok normális eloszlása

Alapvetően n > 100 esetén a centrális határeloszlás tétele alapján adottnak vesszük. Hisztogram és normális eloszlás teszt (Analyze / Descriptive Statistics / Explore) alapján vizsgálható formálisan is, ettől azonban most eltekintünk és a nagy elemszámra tekintettel elfogadjuk a feltételt.

A MODELL SZIGNFIKANCIÁJÁNAK TESZTELÉSE (F PRÓBA)

A modell létjogosultságát tesztelő *F*-próbát az ANOVA táblában találjuk.

Az *F*-próbához tartozó *p*-értéket tartalmazó cella.
p = 0,000

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1552,429	2	776,215	70,428	,000[b]
	Residual	8938,290	811	11,021		
	Total	10490,719	813			

a. Dependent Variable: Egy átlagos héten körülbelül hány pohár bort iszik meg? (12 cl vagy 125 ml = 1 pohár)
b. Predictors: (Constant), Válaszadó kora, Átlagos alkolos költés egy héten

Mivel $p = 0{,}000$, így H0 hipotézist elvethetjük, H1-et fogadjuk el, azaz szignifikáns a modellünk, fennáll az összefüggés a függő és a független változók között.

A BECSLŐFÜGGVÉNY FELÍRÁSA

Mivel a vizsgálható feltételeink teljesültek (kivétel a homoszkedaszticitás, de azt megtűrjük), ezért a becslőfüggvény felírható. (Az exogenitás szinte biztos, hogy nem teljesül a modellben, de ezzel jelen körülmények között és e segédanyag keretében nem tudunk mit tenni.) Ehhez a koefficiens táblázatot kell segítségül hívnunk.

Ebben az oszlopban találhatók a konstans és a regressziós együtthatók. Az első mindig a konstans (β_0), és alatta sorban az egyes magyarázó változókhoz tartozó együtthatók (β_1, β_2 stb.)

Ebben az oszlopban találhatók a konstans és a regressziós együtthatók *t*-próbáihoz tartozó *p*-értékek.

Coefficientsa

Model		U… Co…	…dardized …cients	Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	**-,291**	,368		**-,791**	**,429**			
	Átlagos alkoholos költés egy héten	**,159**	,016	,332	**10,251**	**,000**	,339	,339	,332
	Válaszadó kora	**,041**	,007	,183	**5,632**	**,000**	,194	,194	,183

a. Dependent Variable: Egy átlagos héten körülbelül hány pohár bort iszik meg? (12 cl vagy 125 ml = 1 pohár)

A táblázat alapján felírható a regressziós függvény:

Átl. heti borfogyasztás = -0,291 + 0,159·átl. alkoholos költés + 0,041·válaszadó kora + hiba

Mivel mindkét független változóhoz tartó *t*-próba alapján a regressziós együtthatók szignifikánsnak tekinthetők (Sig = 0,000 → H0-t elvetjük, H1-et fogadjuk el, azaz mindkét együttható szignifikáns), ezért a felírt becslőfüggvény megbízható és érvényes. *Amennyiben lett volna olyan együttható, amelynek t-próbájához tartozó p-értéke 0,05-nél nagyobb, akkor a hozzá tartozó változót el kellett volna hagynunk a modellből, és nélküle újrafuttatni a regressziós becslést.*
Habár a konstanshoz tartozó *t*-próba szignifikanciaszintje (Sig = 0,429) alapján el kellene távolítanunk a modellből, mégis a konstanst célszerű a modellben mindig benntartani.

Regressziós együtthatók értelmezése:

konstans - β_0:
Általánosan:
Amennyiben a magyarázó változók (X_1, X_2) mindegyikének értéke 0, az eredményváltozó értéke β_0 értékét veszi fel (Y tengellyel való metszéspont!).
Jelen feladatban:
Ez lenne: Amennyiben valaki nulla forintot költ alkoholra egy héten, és az életkora 0 év, abban az esetben a várható heti borfogyasztás -0,291 pohár.
DE! A nulla éves nem fogyaszt még bort, így jelen esetben nem értelmezhető az állítás. Ez

gyakran előfordul a konstans esetében. Ilyenkor egyszerűen azt mondhatjuk, hogy a konstanst önállóan nem értelmezzük a magyarázó változók ismeretében. *UGYANAKKOR*, ha az életkor helyett például az lenne a magyarázó változó, hogy havonta hány alkalommal látogat szórakozóhelyeket, akkor a konstans azt mutatná, hogy aki nem jár el bulizni és nem költ alkoholra, az hány pohár bort fogyaszt (pl. ajándékba kapott palack, vendégségben, tehát ilyenkor lenne értelme).

együtthatók – β_1 (heti átlagos alkohol italköltés együtthatója) és β_2 (életkor):
Általánosan:
Amennyiben a magyarázó változó értéke egy egységgel (a saját mértékegységének megfelelően!) növekszik, az az eredményváltozó átlagosan várhatóan β_1-nyi (az eredményváltozó mértékegységében kifejezve!) változásával jár együtt (amennyiben β pozitív, akkor növekedésével, ha negatív, akkor csökkenésével), minden más változatlansága mellett (*ceteris paribus*).
Jelen feladatban:
β_1: Az átlagos heti alkoholos italköltés 1 *euróval* történő növekedése esetében az egyének átlagosan várhatóan 0,159 *pohár* borral isznak meg többet, minden más változatlansága mellett.
β_2: Az életkor egy *évvel* való növekedése esetén az átlagos heti borfogyasztás átlagosan várhatóan 0,041 *pohárral* nő meg, minden egyéb változó változatlansága mellett.
Ezek alapján egy átlagos 30 éves fiatal, aki átlagosan 20 eurót költ egy héten alkoholos italra átlagosan várhatóan 4,12 pohár bort fog elfogyasztani egy héten, mert:

$$-0{,}291 + 0{,}159 \cdot 20 + 0{,}041 \cdot 30 = 4{,}12$$

A modell magyarázóereje – a determinációs együttható (R^2)

Az R^2 mutatót kétféleképpen szokás értelmezni.
1) A modell illeszkedése: Minél közelebb van a mutató értéke 1-hez, annál jobban illeszkedik a felírt regressziós egyenes a vizsgált ponthalmazhoz. Esetünkben gyengén illeszkedik a regressziós egyenes a vizsgált ponthalmazhoz.
2) A modell magyarázóereje: A modell magyarázóerejének százalékos formában való kifejezése. A független változók varianciái hány százalékban magyarázzák a függő változó varianciáját. Esetünkben az R^2 mutató alapján a független változók a függő változó varianciájának 14,8%-át magyarázzák. Ez gyenge magyarázóerőt jelent a modellünk esetében, vagyis sok egyéb tényező befolyásolja még az egyének heti borfogyasztását.

Model Summaryb

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,385a	**,148**	**,146**	3,320	1,893

a. Predictors: (Constant), Válaszadó kora, Átlagos alkolos költés egy héten
b. Dependent Variable: Egy átlagos héten körülbelül hány pohár bort iszik meg? (12 cl vagy 125 ml = 1 pohár)

SYNTAX

```
*Regresszió-elemzés.
REGRESSION
/DESCRIPTIVES MEAN STDDEV CORR SIG N
/MISSING LISTWISE
/STATISTICS COEFF OUTS CI(95) R ANOVA ZPP
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT V06_1
/METHOD=ENTER V10_SZUM1 V18
/SCATTERPLOT=(*ZRESID ,*ZPRED)
/RESIDUALS HISTOGRAM(ZRESID) NORMPROB(ZRESID).
```

GYAKORLÓ FELADATOK

1. Hogyan befolyásolja az alkoholos költést (v10_szum1) a heti szinten elfogyasztott alkoholos italok mennyisége (v06_1, v06_2, v06_5)? Mennyit költ átlagosan várhatóan alkoholra (v10_szum1) az, aki hetente átlagosan 1 pohár bort, 3 pohár sört, 2 pohár rövidet (v06_1, v06_2, v06_5) szokott meginni?
2. Hogyan befolyásolja az alkoholmentes költést (v10_szum2) a heti szinten elfogyasztott alkoholmentes italok mennyisége (v06_3, v06_4, v06_7, v06_8)? Mennyit költ átlagosan várhatóan alkoholmentes italra (v10_szum2) az, aki hetente átlagosan 7 palack vizet, 3 pohár üdítőt, 1 csésze teát és 7 csésze kávét (v06_3, v06_4, v06_7, v06_8) szokott meginni?
3. Hogyan befolyásolja az alkoholmentes költést a válaszadó kora és a háztartás mérete? Mennyit költ várhatóan alkoholmentes italra egy átlagos héten egy 35 éves 3 fős háztartásban élő személy (v10_szum2, v18, v23)?
4. A „Jobb mint otthon” kiskocsma szeretné megérteni, hogyan befolyásolja az elfogyasztott alkoholos és alkoholmentes ital mennyisége az teljes italköltést (v10_szum1 + v10_szum2), mivel bízik benne, hogy ez alapján jól tudja tervezni a bevételeit. A kapcsolatot a teljes minta esetében, valamint külön a férfiak és nők esetében is szeretnék felírni. Felfedezhetők-e jelentős különbségek a változók befolyásoló hatásában a két nemnél? Melyik esetben, mely változó bír a legerősebb, és melyik a leggyengébb hatással a teljes költésre? Ugyanakkora fogyasztás esetében átlagosan várhatóan a férfiak vagy nők fognak többet költeni?

8. FAKTORELEMZÉS

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

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

A faktorelemzés **látens**, nem direktbe mért, változókat eredményez, melyet a **manifeszt** – megfigyelt, mért – változók határoznak meg.
A létrehozott faktorok **standardizált változók**, melyek jellemzői, hogy átlaguk=0 és szórásuk=1.
A faktorelemzés **feltétele**, hogy legalább háromszor annyi válaszadónk legyen, mint bevont változónk, valamint legyen jelen multikollinearitás a változók között. A kiinduló változóinknak metrikusnak kell lennie és közöttük elvárt a viszonylag magas korreláció.

Matematikailag minden egyes kiinduló változó felírható új változók, azaz a faktorok lineáris kombinációjaként, ahol a faktorok közös és egyedi faktorokra oszthatók:

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

ahol:

X_i = i-edik standardizált változó
A_{ij} = az i-edik változó j-edik közös faktorra vonatkozó többszörös standardizált parciális regressziós együtthatója
F = a közös faktor
V_i = az i-edik változó j-edik közös faktorra vonatkozó többszörös standardizált parciális regressziós együtthatója
U_i = az i-edik változó egyedi faktora
m = a közös faktorok száma

Az egyedi faktorok (U_i) korrelálatlanok egymással és a közös faktorokkal (F_i).
A közös faktorok is kifejezhetők a megfigyelt, kiindulási változók lineáris kombinációjaként. $F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \ldots + W_{ik}X_k$

ahol:

F_i =az i-ik faktor becslése
W_i = súly vagy faktorérték együtthatója
k = a változók száma

A faktorok meghatározásának módjai:

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

1. **A Priori meghatározás.** Előfordulhat, hogy előzetes kutatásokból tudjuk, hány faktort akarunk előállítani.

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

A faktorelemzés során a következő legfontosabb fogalmakkal foglalkozunk:

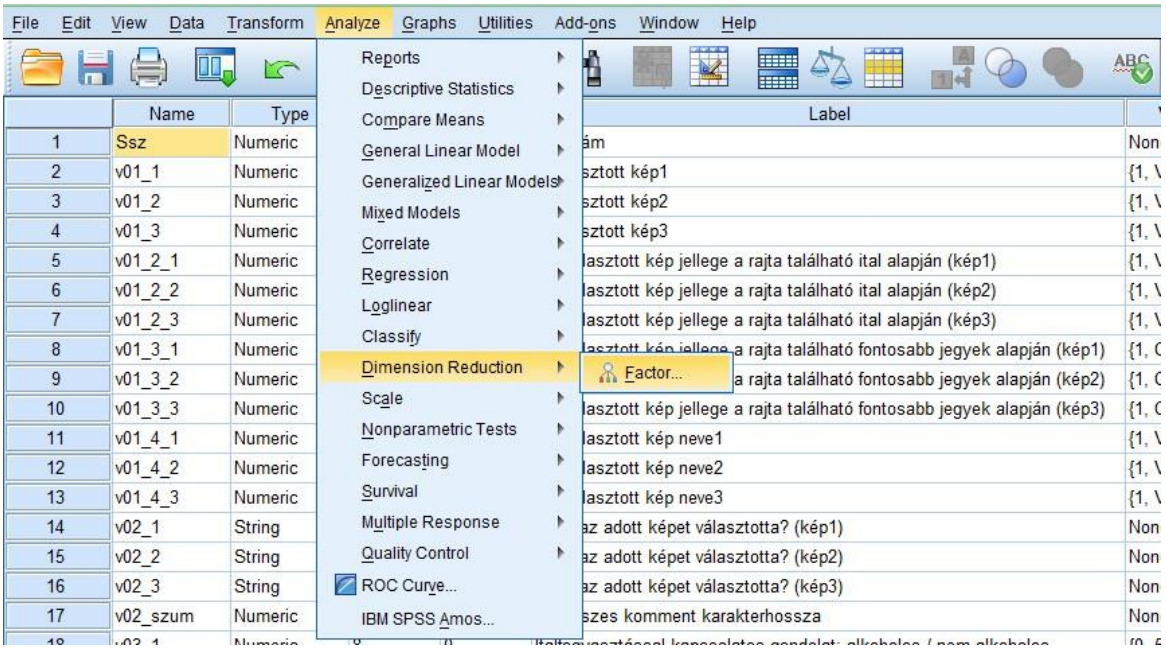
- ✓ **Kommunalitás**: A variancia azon hányada, amelyet egy változó magyaráz. Ez egyben a közös faktorok által magyarázott variancia aránya is. *Minimum elvárt érték 0,25, amennyiben nem teljesül, az adott változót törölni kell a faktorelemzésből.*
- ✓ **Sajátérték**: A variancia azon hányada, amelyet egy-egy faktor magyaráz, vagyis azt mutatja, hogy egy faktor mennyit magyaráz az összes változó varianciájából.
- ✓ **Faktorsúly**: A változók és a faktorok közötti korrelációs együtthatók, ezek a lineáris kombináció paraméterei. A cél, hogy abszolút értékben minél nagyobb legyen. *Minimum elvárt értéke 0,4, amennyiben nem teljesül, az adott változót törölni kell a faktorelemzésből.*
- ✓ **Faktorsúly-ábra**: Az eredeti változók ábrázolása, ahol a faktorokat mint tengelyeket használjuk.
- ✓ **A faktor mátrix:** Az előállított faktoroknak az egyes megkérdezettekre vonatkozóan becsült értékei.
- ✓ **Kaiser-Meyer-Olkin (KMO)mutató**: A korrelációs mátrix elemeiből számítódik és azt méri, hogy az adatbázis alkalmas-e a faktorelemzésre. Ha az érték kisebb, mint 0,5, akkor az adatbázis alkalmatlan a faktorelemzésre. *Értékhatárok: KMO≥0,9 - kiváló, KMO≥0,8 – nagyon jó, KMO≥0,7 - megfelelő, KMO≥0,6 - közepes, KMO≥0,5 – gyenge, KMO≤0,5 - elfogadhatatlan*
- ✓ **Bartlett-féle gömbölyűségi teszt**: A faktorelemzés elvégezhetőségének próbája. H0: a megfigyelt változók korrelációs mátrixa egységmátrix, vagyis a változók páronként korrelálatlanok
- ✓ **Varianciahányad**: a varianciának az a hányada, amelyet egy-egy faktor magyaráz.
- ✓ **Reziduum**: A megfigyelt, eredeti korrelációs mátrixban található korrelációs együtthatók és a reprodukált, azaz a faktormátrixból becsült korrelációs együtthatók közötti különbségek.
- ✓ **Sajátérték-ábra**: A sajátértékek ábrázolása az előállított faktorok sorszámának függvényében.
- ✓ **Rotáció**: a faktorok tengelyeit elforgatjuk úgy, hogy könnyebben értelmezhetőek legyenek a faktorok.

FELADAT

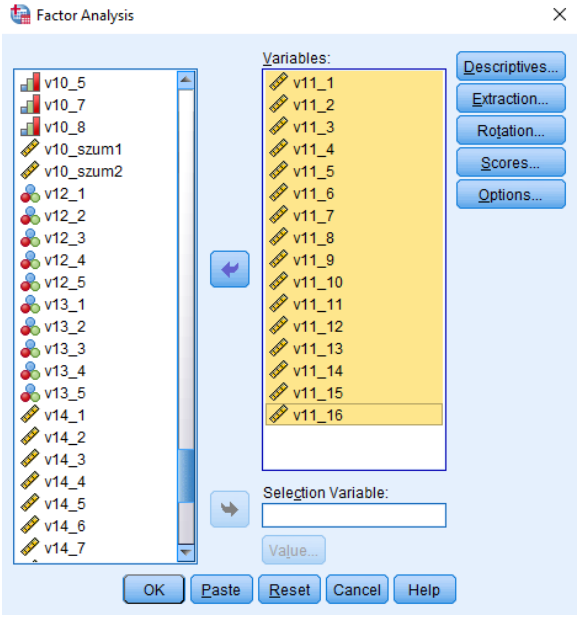
A válaszadók vásárlási attitűdjeit leíró változók (k11_1-16) esetében végezzünk faktorelemzést, és válasszuk ki a legjobb megoldást. A rotálatlan vagy a Varimax rotált megoldást alkalmazzuk inkább?
***Felhasznált fájlok**: Italfogyasztási szokások.sav*

MEGOLDÁS

Elérés útvonala: Analyze/ Dimension Reduction/ Factor



Változók bevitele: Variables: v11_1 – v11_16



Lekért adatok:

Descrpitives

Statistics:

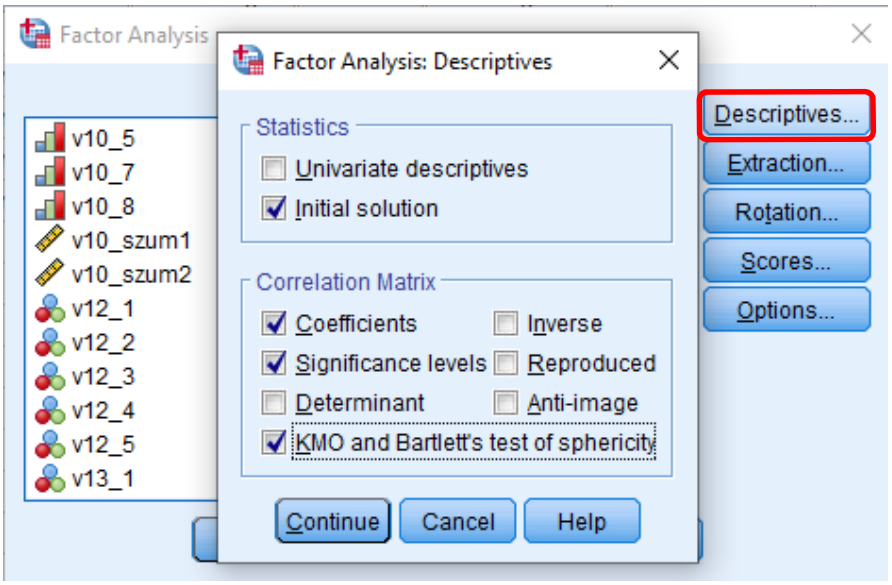
✓ Initial solution – *kiinduló adatok a kommunalitásnál megjelenjenek (értéke 1)*

Correlation Matrix:

✓ Coeffitiens – *elemzésbe bevont változók közötti korrelációt mutatja*

✓ Significance levels – e*lemzésbe bevont változók közötti korreláció szignifikanciaszintjét mutatja*

✓ KMO and Bartlett's test – *a faktorelemzés előfeltételeként szolgáló mutatók és próbák eredményét adja*



Extraction

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

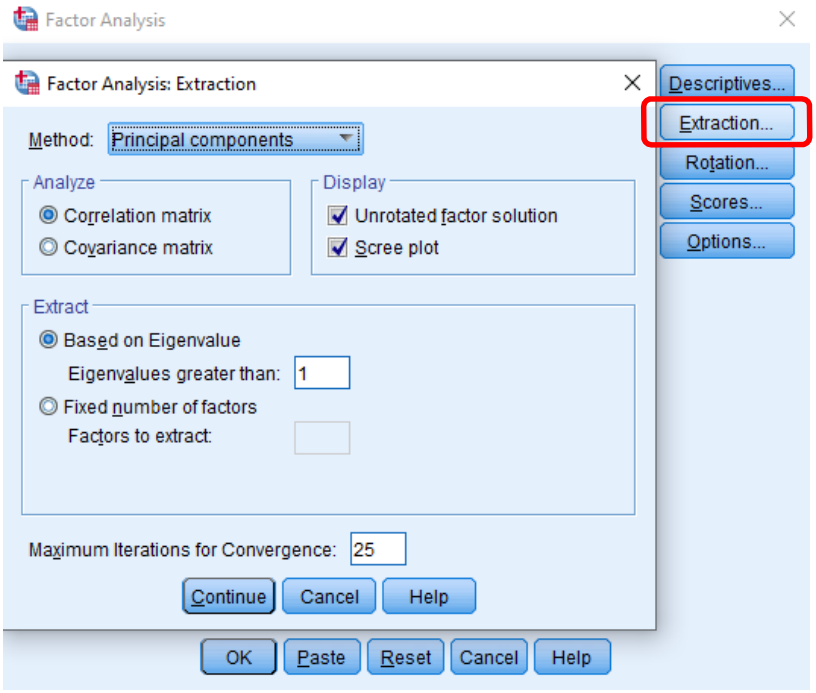
Analyze

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

Extract – Ebben a pontban tudjuk megjelölni, hogy mi alapján szeretnénk a megoldást megjeleníteni. Mint láthattuk, több szempont alapján is meghatározható az ideális faktorszám. Az SPSS-ben automatikusan be tudjuk állítani, hogy a kívánt sajátérték alapján – jelen esetben nagyobb mint 1 (Kaiser kritérium) – alakítsa ki a faktorokat a program. Amennyiben más módszer alapján szeretnénk a kívánt faktorszámot kiválasztani és elemezni - a'priori, magyarázott variancia -, a *„Fixed number of factors"* pontban kell megadnunk, hogy az eredmények tükrében, milyen faktorszámú megoldással kívánunk dolgozni.

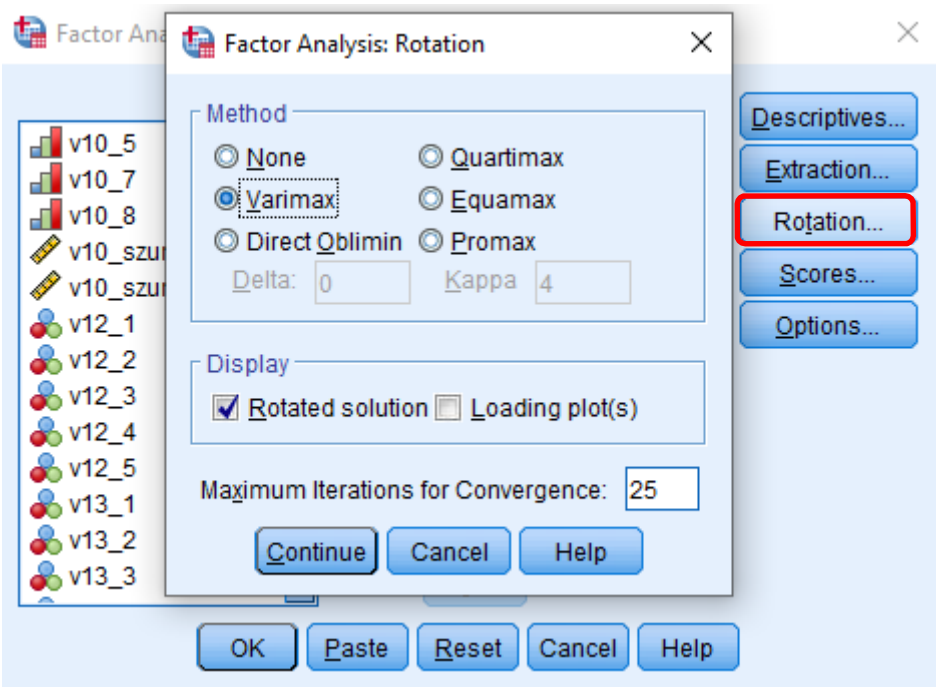
- ✓ *Eigenvalues greatger than: 1 (*az első futtatásnál jellemzően ezzel a beállítással kezdünk, és ha szükséges, akkor később áttérünk az alábbi, fix számú faktort kínáló megoldásra)

  VAGY
- ✓ Fixed n*umber of factors (pl. a'priori meghatározott, magyarázott variancia)*



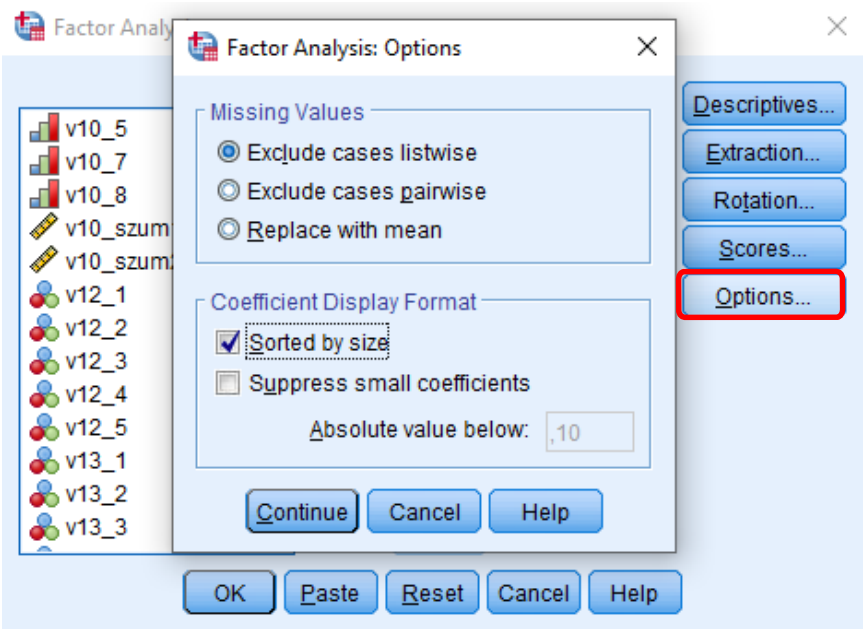
Rotation – itt állítható be, hogy rotált vagy nem rotált megoldással szeretnénk dolgozni. Amennyiben csak a nem rotált megoldás érdekel minket a *„None”* pontot jelöljük meg. Amennyiben bármilyen rotált megoldás érdekes lehet, válasszuk ki a kívánt módszert. Mivel az Extraction menüpontban jelöltük a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

- ✓ *Varimax*



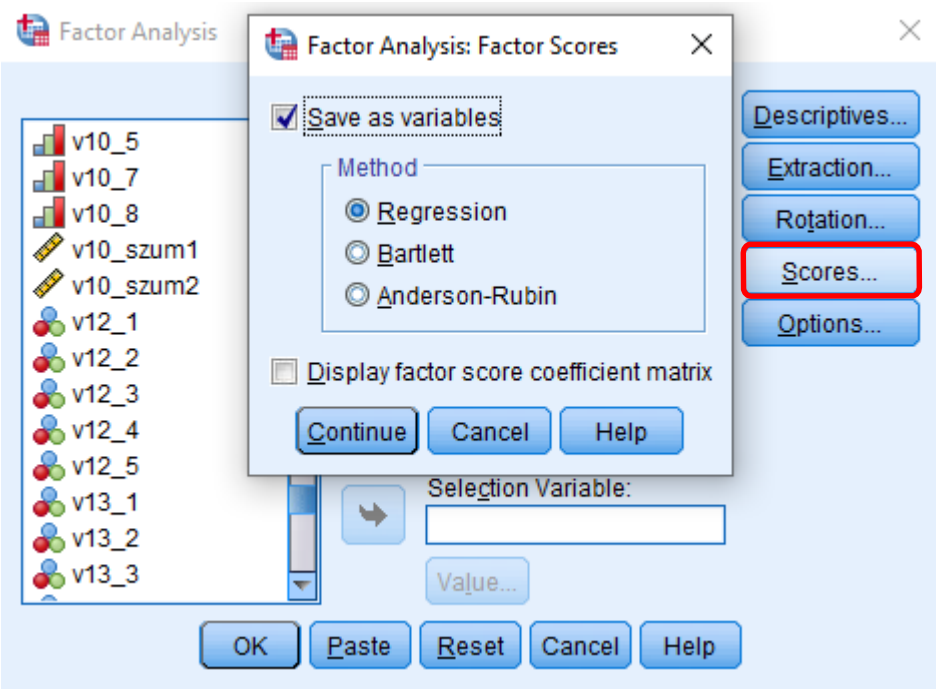
Options

- ✓ . *Sorted by size* - a változók faktorsúly szerinti sorba rendezése. Érdemes mindig választani a könnyebb értelmezhetőség miatt.



Scores

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



ÉRTÉKELÉS

1. Korrelációs tábla:

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

2. Faktorelemzés feltétele

KMO értéke 0,6 feletti, tehát az elemzés elvégezhető, mivel a főkomponens jól illeszkedik az adatokhoz. A KMO pontos értéke 0,736, tehát megfelelő illeszkedésünk van, a változók alkalmasok faktorelemzésre.

Bartlett-féle gömbölyűségi teszthez tartozó szignifikanciaszint 0,000, vagyis a H0t, mely szerint a megfigyelt változók korrelációs mátrixa egységmátrix, vagyis a változók páronként korrelálatlanok, elvethetjük.

Ezeknek az eredményeknek a tükrében az adatbázisunk alkalmas a faktorelemzésre.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,736
Bartlett's Test of Sphericity	Approx. Chi-Square	807,593
	df	120
	Sig.	,000

KMO mutató értéke 0,736

A gömbölyűségi Bartlett teszthez tartozó p érték

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

Kaiser kritériumot a „Total Variance Explained" táblában tudjuk ellenőrizni. Ennek alapján, vagyis, hogy a faktorokhoz tartozó sajátérték magasabb legyen, mint 1, az ideális megoldásnak az 5 faktoros megoldás mutatkozik.
A magyarázott varianciahányad alapján, vagyis, hogy a faktorok által magyarázott variancia legalább 60% legyen, az ideális faktorszámnak a 6 faktoros megoldás tűnhet.

KAISER kritériumhoz: Ebben az oszlopban látjuk magukat az egyes faktorszámokhoz tartozó sajátértékeket

60% kritériumhoz: Ebben az oszlopban látjuk, hogy az adott faktorszám a variancia hány %-át magyarázza meg.

Total Variance Explained

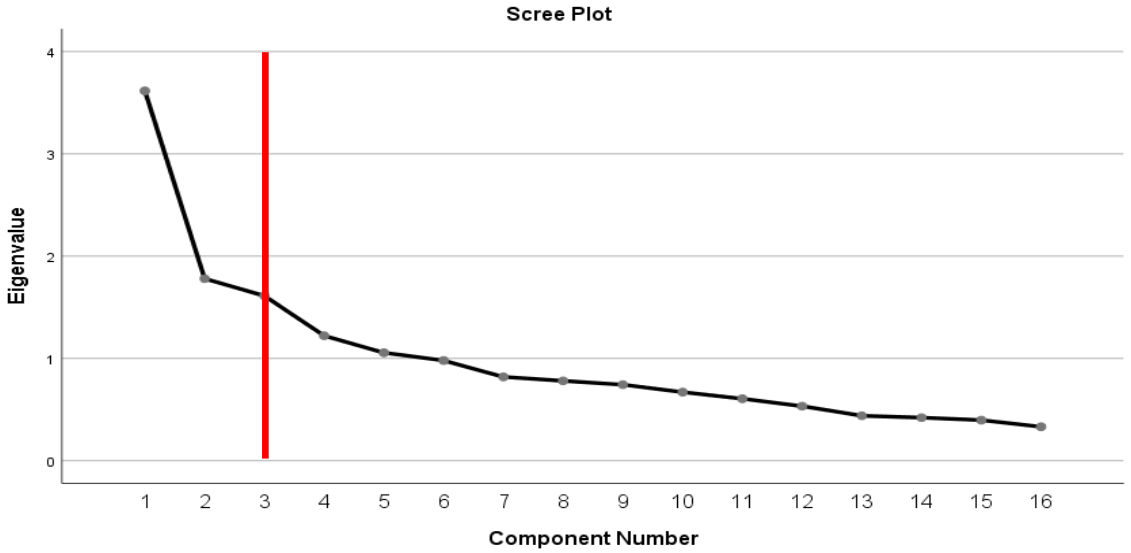
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,615	22,596	22,596	3,615	22,596	22,596	2,834	17,713	17,713
2	1,780	11,122	33,718	1,780	11,122	33,718	1,805	11,282	28,995
3	1,611	10,072	43,790	1,611	10,072	43,790	1,690	10,565	39,560
4	1,221	7,633	51,423	1,221	7,633	51,423	1,640	10,247	49,807
5	1,055	6,595	58,018	1,055	6,595	58,018	1,314	8,211	58,018
6	,979	6,121	64,139						
7	,819	5,119	69,258						
8	,78	4,875	74,134						
9	,7	4,643	78,777						
10		4,187	82,964						
11		3,784	86,748						
12		3,331	90,079						
13			22						
14			52						
15			28						
16			00						

Extracti

6 faktornál lépi át a 60% megmagyarázott varianciahányadot (64,139%)

Eddig a faktorszámig (5) a sajátérték 1 felett van. Ha eggyel több faktort választunk (6), akkor a sajátérték 1 alá csökken (0,979-re).

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



Mivel az Extraction menü Extract pontjában a *„Based in Eigenvalue”*-t jelöltük meg, ezért az SPSS azt a megoldást mutatja most számunkra, amikor a faktorok sajátértéke magasabb, mint 1 (Kaiser kritérium), tehát az 5 faktoros megoldást. Nézzük meg a rotálatlan faktor mátrix és rotált faktormátrix esetében a faktorok és faktorsúlyok alakulását, valamint a kommunalitások értékét ebben az esetben.

Vegyük először a rotálatlan faktor mátrixot szemügyre. A táblázat soraiban az elemzésbe bevont változók (kérdések) találhatóak, a *Component* részben megjelenő oszlopok pedig a kiválasztott megoldás esetében mutatja a faktorszámokat, jelen esetben 5 faktoros megoldással találkozunk. A táblázat belsejében található számok a faktorsúlyok (piros körvonal). Ezek mutatják, hogy az adott változó az adott faktorhoz, milyen erősen „tartozik”, vagyis az adott változó és faktor korrelációs együtthatója ez az érték. Minél magasabb a faktorsúly, annál erősseb kapcsolat áll fenn a változó és az adott sorszámú faktor között.
Első lépésben *vizsgáljuk meg minden egyes változó esetében, hogy abszolút értékben melyik sorszámú faktorhoz tartozik a legerősebb faktorsúlya.* Ezt célszerű valamilyen színnel külön jelölni (ehhez excelbe praktikus átvinni a táblázatot).
A faktorsúlyokat megvizsgálva megállapítható, hogy *a legmagasabb faktorsúly egyetlen esetben sem alacsonyabb, mint 0,4* (ez a minimálisan elvárt faktorsúly). Az első 9 változó esetében az 1-es faktorhoz tartozó faktorsúly a legmagasabb, a következő 3 változó esetében a 2-hez, 2-2 esetében a 3 és 4eshez, míg az 5-ös faktorhoz egyetlen egy változó sem tartozik. Mivel van egy faktorunk, amelyhez egyetlen egy változó sem tartozik a legerősebb súllyal, ezért azt a megoldást nem használhatjuk a későbbi elemzések során.
Vizsgáljuk meg a **Varimax rotált** megoldást is. a ’Rotated Component Matrix’ táblában.
Ebben az esetben az *egyes faktorokhoz tartozó változók száma is kiegyensúlyozottabb* képet mutat, ugyancsak nincs 0,4 alatti faktorsúlyunk, és minden egyes faktorhoz tartozik legalább egy változó. A kommunialitások is rendben vannak a megoldás esetében, mivel *egyetlen egy esetben sem találkozunk 0,25 alatti kommunalitással.*
Amennyiben találkoznánk abszolút értékben 0,4 alatti legnagyobb faktorsúllyal vagy 0,25 alatti kommunalitással, akkor azokat a változókat ki kellene zárnunk az elemzésből, és úgy újra lefuttatni a faktorelemzést

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
Környezetbarát termékeket próbálok vásárolni, amikor csak lehetséges.	,802	,042	,207	-,069	,050
Fair Trade termékeket vásárolok, amikor csak lehet.	,744	,145	,011	,127	,010
Olyan terméket igyekszem vásárolni, amit helyi termelők gyártanak.	,691	,111	,182	-,022	,024
A fogyasztóknak bojkottálniuk azokat a termékeket, amelyeket felelőtlen vállalatok készítenek.	,671	-,108	,143	,003	,155
Egészségtudatos fogyasztónak tartom magam.	,653	,051	,045	-,161	,020
Azok a dolgok, amiket megvásárolok sokat mondanak rólam.	,202	,722	,056	,113	-,038
Gyakran vásárolok szórakozásból.	-,134	,679	-,046	,276	,318
Nagy figyelmet fordítok a márkákra, amelyeket ismerek	,199	,603	,253	-,278	,041
Hajlamos vagyok ismertebb márkákat vásárolni.	,073	-,557	-,258	,476	,110
Fontos számomra, hogy nagyon jó minőséget kapjak.	,154	,119	,824	-,066	,045
Magasak az elvásáraim azokkal a dolgokkal kapcsolatban, amit megvásárolok.	,276	,132	,756	-,010	,095
A vásárlás unalmas számomra.	-,177	,231	-,010	,724	-,085
Nem nagyon figyelek oda, hogy mennyibe kerülnek a dolgok, amiket vásárolok.	,049	-,089	-,034	,673	,123
Amennyiben lehetséges, leárazásokon vásárolok.	,048	,041	,087	,085	,765
Szerintem az egyik márka olyan, mint a másik.	-,021	-,068	,274	,452	-,549
Ha elégedetlen vagyok azzal, amit vettem, panaszt teszek.	,206	,033	,346	,090	,500

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Communalities

	Initial	Extraction
Nagy figyelmet fordítok a márkákra, amelyeket ismerek	1,000	,546
Hajlamos vagyok ismertebb márkákat vásárolni.	1,000	,620
Fair Trade termékeket vásárolok, amikor csak lehet.	1,000	,591
Ha elégedetlen vagyok azzal, amit vettem, panaszt teszek.	1,000	,421
Azok a dolgok, amiket megvásárolok sokat mondanak rólam.	1,000	,579
A vásárlás unalmas számomra.	1,000	,616
Szerintem az egyik márka olyan, mint a másik.	1,000	,586
Olyan terméket igyekszem vásárolni, amit helyi termelők gyártanak.	1,000	,524
Magasak az elvásáraim azokkal a dolgokkal kapcsolatban, amit megvásárolok.	1,000	,675
Fontos számomra, hogy nagyon jó minőséget kapjak.	1,000	,724
Nem nagyon figyelek oda, hogy mennyibe kerülnek a dolgok, amiket vásárolok.	1,000	,479
Amennyiben lehetséges, leárazásokon vásárolok.	1,000	,604
Egészségtudatos fogyasztónak tartom magam.	1,000	,457
Gyakran vásárolok szórakozásból.	1,000	,658
A fogyasztóknak bojkottálniuk azokat a termékeket, amelyeket felelőtlen vállalatok készítenek.	1,000	,506
Környezetbarát termékeket próbálok vásárolni, amikor csak lehetséges.	1,000	,695

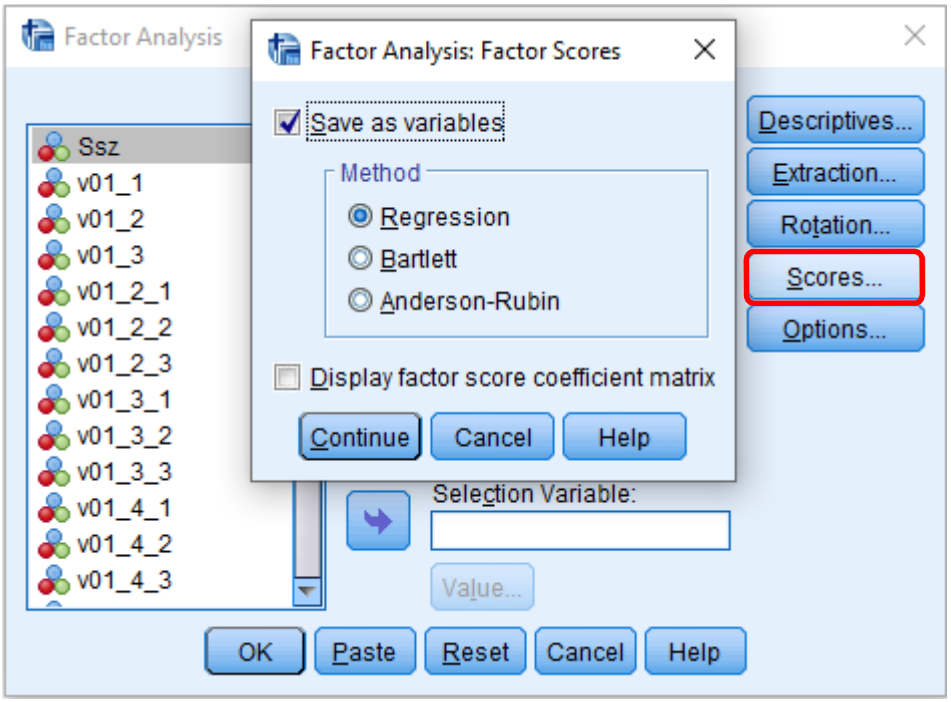
Extraction Method: Principal Component Analysis.

Mivel minden kritériumnak megfelelünk, ezért a megoldás ideálisnak tűnik, tehát nevezzük el a faktorokat. A faktorok nevét a hozzájuk tartozó változók alapján adjuk meg, mivel azoknak a változók jelentését tömörítik legnagyobb súllyal.
Amennyiben egy változó a faktorhoz negatív súllyal tartozik, akkor annak a változónak az ellentétes jelentését kell figyelembe vennünk.[2]
Ezek alapján a faktorainkat a következőképpen nevezhetjük el:

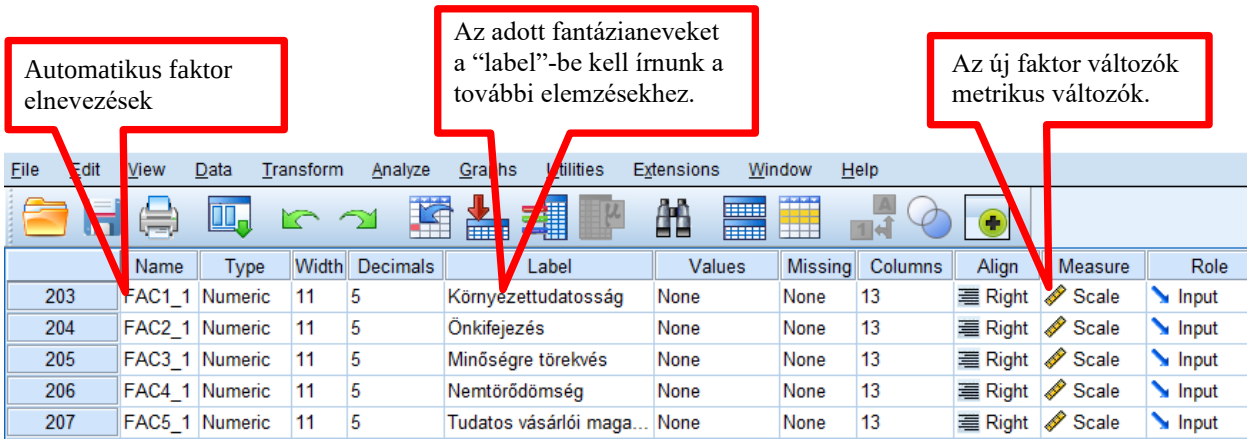
Környezetbarát termékeket próbálok vásárolni, amikor csak lehetséges.	**,802**					***Környezet-tudatosság***
Fair Trade termékeket vásárolok, amikor csak lehet.	**,744**					
Olyan terméket igyekszem vásárolni, amit helyi termelők gyártanak.	**,691**					
A fogyasztóknak bojkottálniuk azokat a termékeket, amelyeket felelőtlen vállalatok készítenek.	**,671**					
Egészségtudatos fogyasztónak tartom magam.	**,653**					
Azok a dolgok, amiket megvásárolok sokat mondanak rólam.		**,722**				***Önkifejezés***
Gyakran vásárolok szórakozásból.		**,679**				
Nagy figyelmet fordítok a márkákra, amelyeket ismerek		**,603**				
Hajlamos vagyok ismertebb márkákat vásárolni.		**-,557**				
Fontos számomra, hogy nagyon jó minőséget kapjak.			**,824**			***Minőségre törekvés***
Magasak az elvásásaim azokkal a dolgokkal kapcsolatban, amit megvásárolok.			**,756**			
A vásárlás unalmas számomra.				**,724**		***Nem törődömség***
Nem nagyon figyelek oda, hogy mennyibe kerülnek a dolgok, amiket vásárolok.				**,673**		
Amennyiben lehetséges, leárazásokon vásárolok.					**,765**	***Tudatos vásárlói magatartás***
Szerintem az egyik márka olyan, mint a másik.					**-,549**	
Ha elégedetlen vagyok azzal, amit vettem, panaszt teszek.					**,500**	

Amennyiben nehezen értelmezhető a megoldás (azaz az egy faktorba sorolt változók között tartalmi összefüggés nincs), vagy nehezen elnevezhetően érezzük őket, érdemes megnézni más megoldási opciókat is (pl. 60%-is magyarázott variancia, azaz esetünkben a 6 faktor). Ezt úgy tehetjük meg, hogy újrafuttatjuk a faktoranalízist, de az ’extraction’ menüben sajátérték/eigenvalues helyett azt választjuk, hogy „fix number of factors”, majd megadjuk a kívánt faktorszámot. Ez nem változtatja a KMO és a Bartlett eredményeit, mivel ugyanazt a változócsoportot elemezzük, ellenben előfordulhat, hogy könnyebben értelmezhető megoldásra jutunk.

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



A mentés eredményeként az adatbázisban megjelennek az új változók, ahol a Label cellába a korábbi elnevezéseket be tudjuk gépelni. Ettől kezdve ezekkel a *standardizált, metrikus változókkal* bármilyen elemzést el tudunk végezni.
A Name oszlopba az SPSS automatikusan ad egy kódot a faktoroknak. A FAC jelenti, hogy faktorelemzés eredményeként jelentek meg az új változók, az „_" utáni szám jelöli, hogy az adott adatbázison belül hányadik elmentett megoldásunk van. Azok a változók, ahol az „_" utáni szám megegyezik egy megoldáshoz tartoznak. Az „_" előtti szám mutatja, hogy az adott változók, az adott, elmentett faktorelemzés hányas számú faktora.
Pl. FAC1_2 – a másodszorra elmentett faktorelemzés első számú faktora
FAC2_2 – ugyanezen faktorelemzés – másodszorra elmentett – második számú faktora.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
203	FAC1_1	Numeric	11	5	Környezettudatosság	None	None	13	Right	Scale	Input
204	FAC2_1	Numeric	11	5	Önkifejezés	None	None	13	Right	Scale	Input
205	FAC3_1	Numeric	11	5	Minőségre törekvés	None	None	13	Right	Scale	Input
206	FAC4_1	Numeric	11	5	Nemtörődömség	None	None	13	Right	Scale	Input
207	FAC5_1	Numeric	11	5	Tudatos vásárlói maga...	None	None	13	Right	Scale	Input

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

		CLU3_3	FAC1_2	FAC2_2	FAC3_2	FAC4_2	FAC5_2
1	1	1	,61137	,09608	-,69638	1,18178	-,19606
2	1	1	.	.	.	.	.
3	2	2	.	.	.	.	.
4	1	1	1,38293	-,62557	,33600	,60939	-,79549
5	1	1	.	.	.	.	.
6	1	1	.	.	.	.	.
7	1	1	,56409	,11183	-,67716	-,08943	-,64740
8	1	1	.	.	.	.	.
9	.	.	.	.	.	.	.
10	2	2	.	.	.	.	.
11	3	3	.	.	.	.	.
12	1	1	,02492	-,48220	-,12011	-,65567	-1,49683
13	1	1	,34730	-,17076	-,27267	1,01917	-,23713
14	2	2	.	.	.	.	.
15	1	1	.	.	.	.	.
16	.	.	.	.	.	.	.
17	1	1	.	.	.	.	.
18	.	.	,10851	1,03671	-,48685	,57279	-,62069
19	.	.	.	.	.	.	.
20	3	3	2,21790	-,16311	-1,05025	,23928	-1,86022
21	1	1	.	.	.	.	.
22	.	.	1,76678	-,39228	1,21818	-,53384	-1,29490
23	.	.	,74777	-1,89800	-,49596	-,59201	-,06010

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

Például: az első válaszadó a környezettudatosság (0,611) és nem törődömség (1,181) jelzőkkel átlag feletti szinten ért egyet (átlag feletti szinten jellemző saját viselkedésére az adott tulajdonság), míg a további három faktor átlag alatti szinten. Legkevésbé a minőségre törekvés jellemzi az adott válaszadót (-0,69)

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

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

SYNTAX

```
*Faktorelemzés – már a mentés is benne van.
FACTOR
/VARIABLES V11_1 V11_2 V11_3 V11_4 V11_5 V11_6 V11_7 V11_8 V11_9 V11_10
V11_11
V11_12 V11_13
V11_14 V11_15 V11_16
/MISSING LISTWISE
/ANALYSIS V11_1 V11_2 V11_3 V11_4 V11_5 V11_6 V11_7 V11_8 V11_9 V11_10
V11_11
V11_12 V11_13
V11_14 V11_15 V11_16
/PRINT INITIAL CORRELATION SIG KMO EXTRACTION ROTATION
/FORMAT SORT
/PLOT EIGEN
/CRITERIA MINEIGEN(1) ITERATE(25)
/EXTRACTION PC
/CRITERIA ITERATE(25)
/ROTATION VARIMAX
/SAVE REG(ALL)
/METHOD=CORRELATION.
```

GYAKORLÓ FELADAT

1. A Macifröccs Kft. szeretné részletesen megvizsgálni az okokat, amelyek miatt az egyének egy italt kedvenc alkoholos italuknak választanak. A kérdőívben lekérdezett változók azonban nehezen interpretálhatóak a vállalat kommunikációja szempontjából. A menedzsmenten belül felmerült a kérdés, hogy leírhatóak-e valamilyen sűrített, absztrakt dimenziókkal a választási indokok (v7_4_1 – v7_4_17)? Amennyiben az adatok alkalmasak a módszer használatára, válasszuk ki a számunkra legszimpatikusabb megoldást (Kaiser kritérium, vagy 60%, vagy a'prioiri), melyet mentsük is el. A létrehozott új változók esetében megfigyelhető-e szignifikáns különbség a férfiak és a nők között (v17)? A származási ország (v20) valamint a különböző kedvenc italok (v07_0) esetében eltérően alakulnak az adatok? A válaszadók életkora és az új faktorok között van-e kapcsolat (v18)

9. KLASZTERELEMZÉS

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

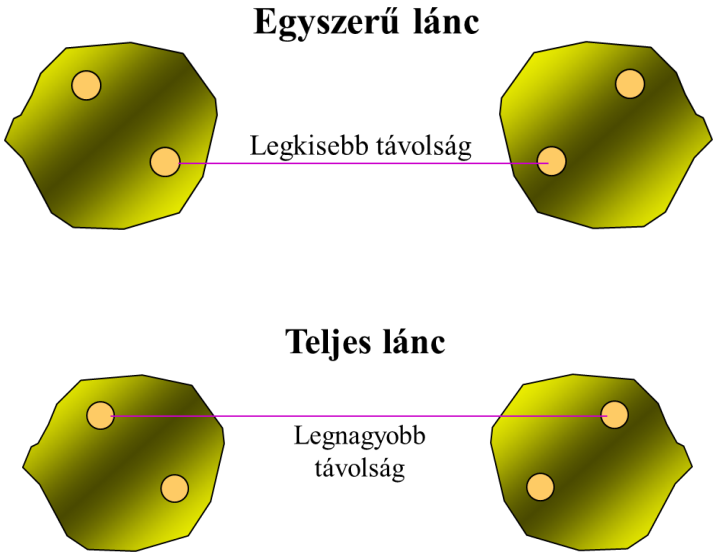
A klaszterelemzés nehézsége és szépsége hasonló a faktorelemzéshez. Itt sincs egyetlen megoldási mód, amelyet követhetünk, azonban számos segítség áll a rendelkezésünkre, hogy meghozzuk a döntésünket a végső csoportosításról.

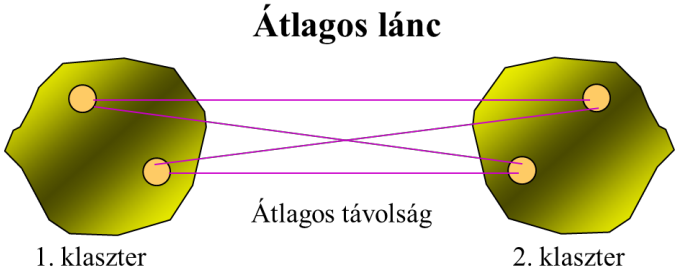
A felhasznált változók metrikusak és nem metrikusak is lehetnek, azonban *mi csak a metrikus* adatokkal foglalkozunk. Az elemzés során *nem teszünk különbséget függő és független változó között, nincs oksági kapcsolatra vonatkozó hipotézis*.

A klaszterelemzés feltétele, hogy a felhasznált változók között *ne legyen, vagy alacsony korreláció legyen*. (Éppen ezért lehet jó kiindulópont egy faktorelemzés során létrehozott változócsoport.)

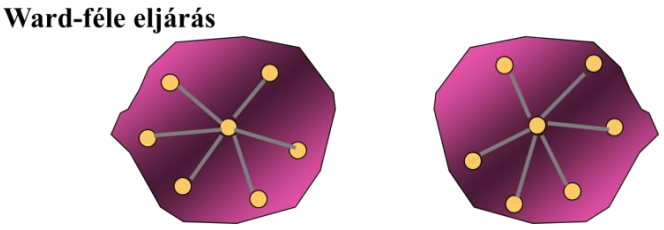
A módszerek között megkülönböztetünk hierarchikus és nem hierarchikus klaszterelemzési módszereket. A hierarchikus klaszterelemzés tovább bontható összevonó és felosztó módszerekre.
Mi csak az összevonó módszerekkel foglalkozunk, ahol az összevonás a következő összevonási módszerek alapján történhet:

1. Lánc módszer

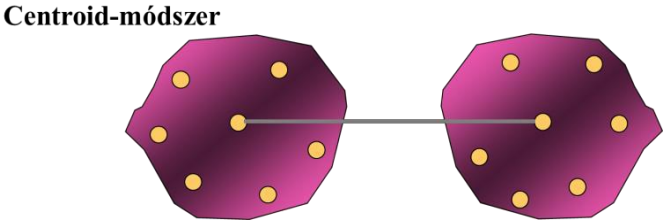




2. Variancia módszer



3. Centroid módszer



A téma során érintett fogalmak:

- ✓ hasonlósági/távolságmérték
- ✓ klaszterelemzési módszerek
- ✓ összevonási/csoportképzési módszerek
- ✓ összevonási séma
- ✓ klaszter-középérték
- ✓ dendrogram

FELADAT

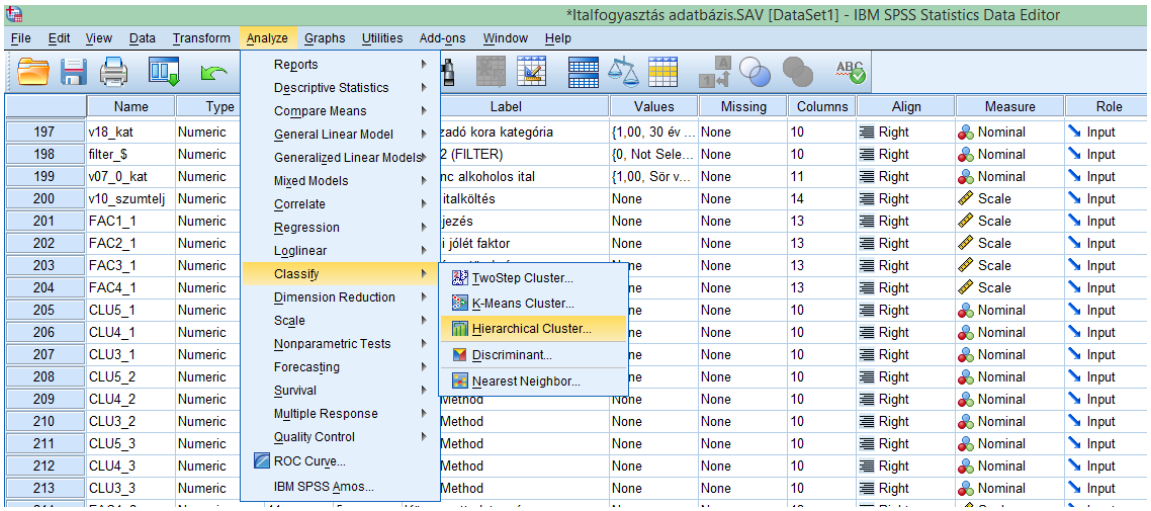
A különböző alkoholos italokat kedvencként-választás indokaiból (17 állítás) hozzunk létre szűkebb, összegző dimenziókat, és ezek alapján vizsgáljuk meg, hogy milyen fogyasztói csoportok azonosíthatóak a mintán.

***Felhasznált fájlok**: Italfogyasztási szokások.sav*

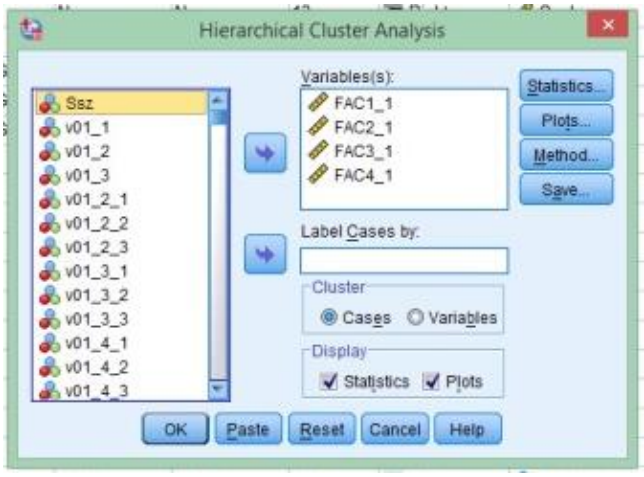
MEGOLDÁS

A főkomponens-elemzés módszerével Varimax forgatást alkalmazva 4 faktort azonosítottunk a választási indokokat megvizsgálva, melyeket a következő névvel láttuk el: Önkifejezés, Egyéni jólét faktor, Minőségre törekvés, Szórakozás faktor. Ezeket az újonnan létrehozott korrelálatlan változókat használjuk fel a klaszterelemzés során.[2]

Elérés útvonala: Analyze/Classify/Hierarchical Cluster



Változók bevitele: Variables: fac1_1, fac2_1, fac3_1, fac4_1



[2] A klaszterelemzés bármilyen metrikus változókkal elvégezhető, nem csak faktorokkal.

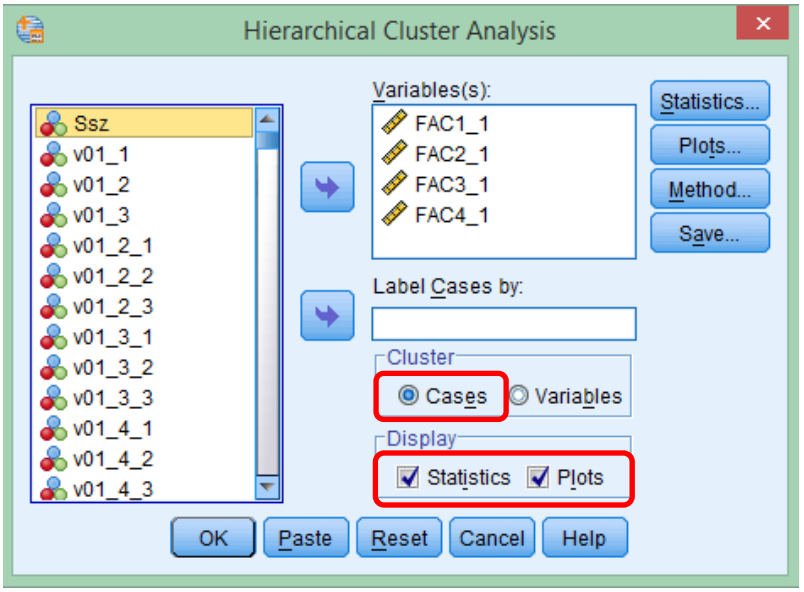
Lekért adatok:

Cluster

- ✓ cases → *az válaszadókra akarunk csoportosítást látni*

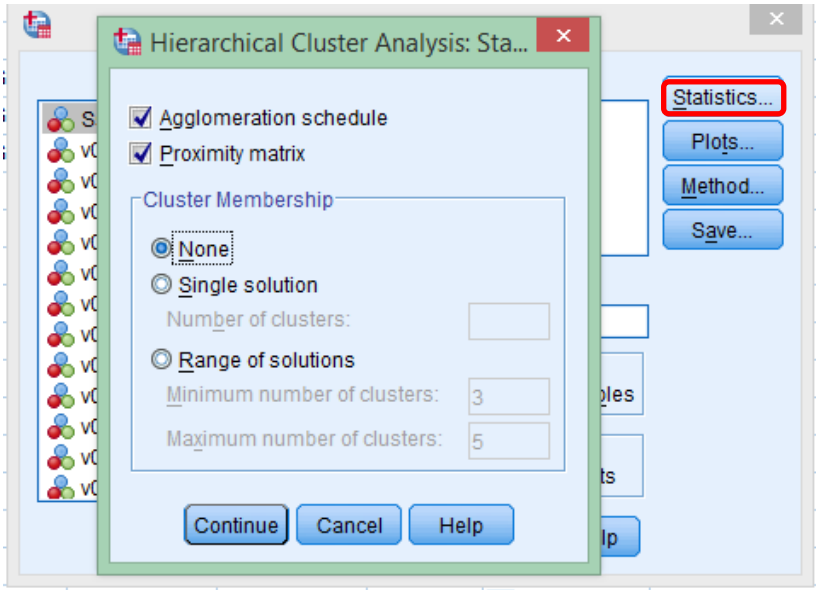
Display

- ✓ Statistics
- ✓ Plots



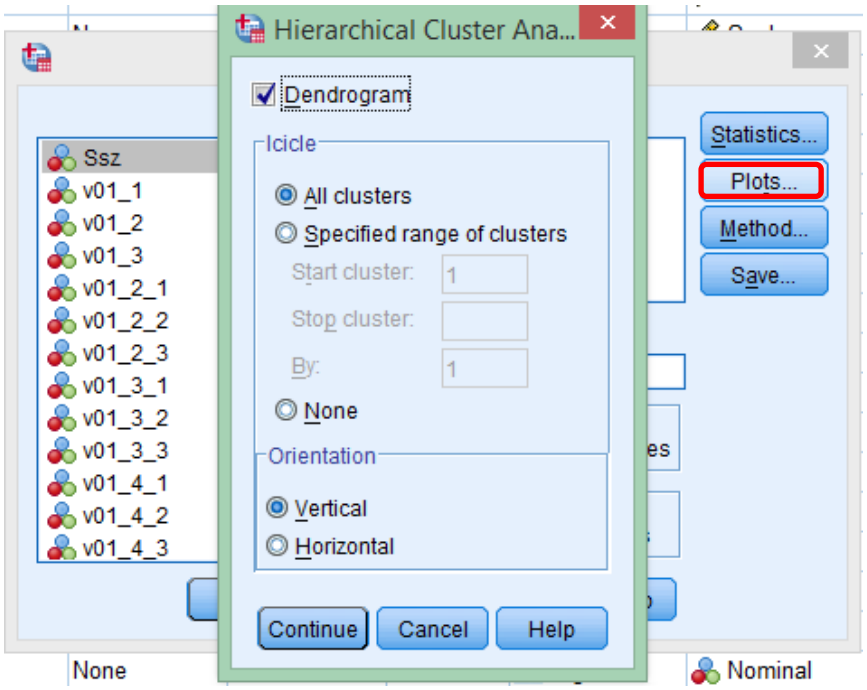
Statistics

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



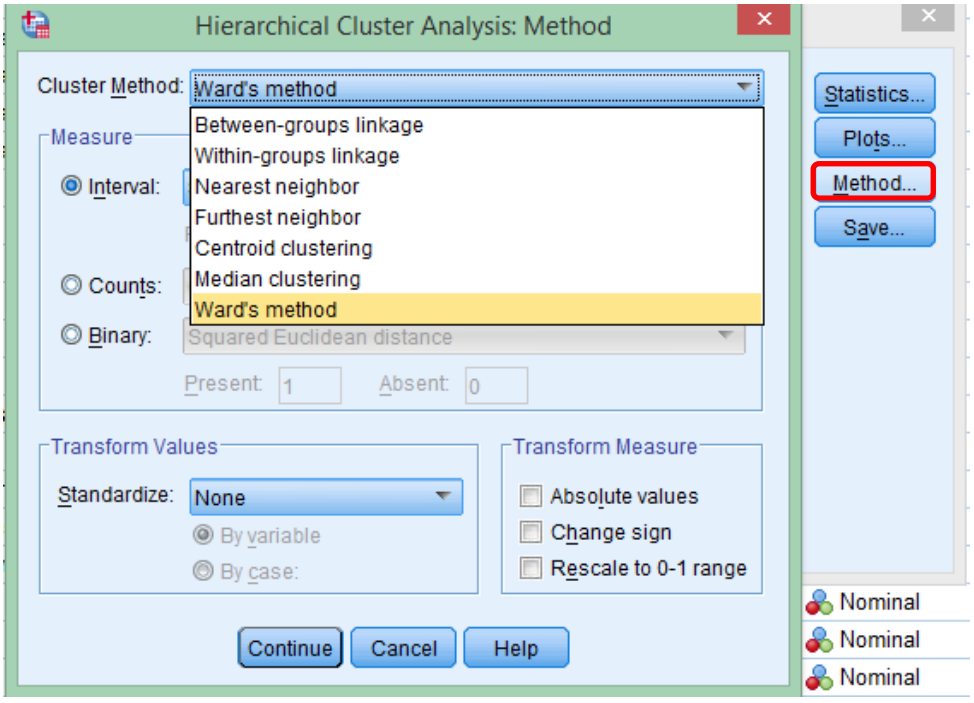
Plots

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



Methods – ideális megoldás megtalálása

- ✓ Ward's method



Save **–** mentés (csak jó eredmény esetén, tehát az első futtatásnál még nem érdemes választani.)

ÉRTÉKELÉS

1. **Proximity** – távolság mátrix

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

2. **Agglomaration schedule** – összevonási séma

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

A táblázatban található koefficiens értékek alapján lehetséges döntést hozni a klaszterek számáról. Az 50%-os szabály szerint (maximum koefficiens érték fele) jelen esetben a 6 klaszteres megoldás lenne célszerű (max. érték 2220 → fele 1110 → ahhoz közeli érték a 1149-es *értéktől hány csoportunk marad*).

Itt láthatjuk, hogy milyen távolságok alapján vonódtak a klaszterek össze.

Itt látjuk azt, hogy a két összevont klaszter melyik lépésben jelent meg korábban.

Láthatjuk, hogy az első lépésnél a 89. és a 762. megfigyelést vonta össze.

Az 50%-os szabály alapján 6 csoportunk van 2200/2=1100 felett.

Itt látjuk azt, hogy a 89-es és a 762-es közös klasztere hol jelenik meg legközelebb (215-ös lépésnél)

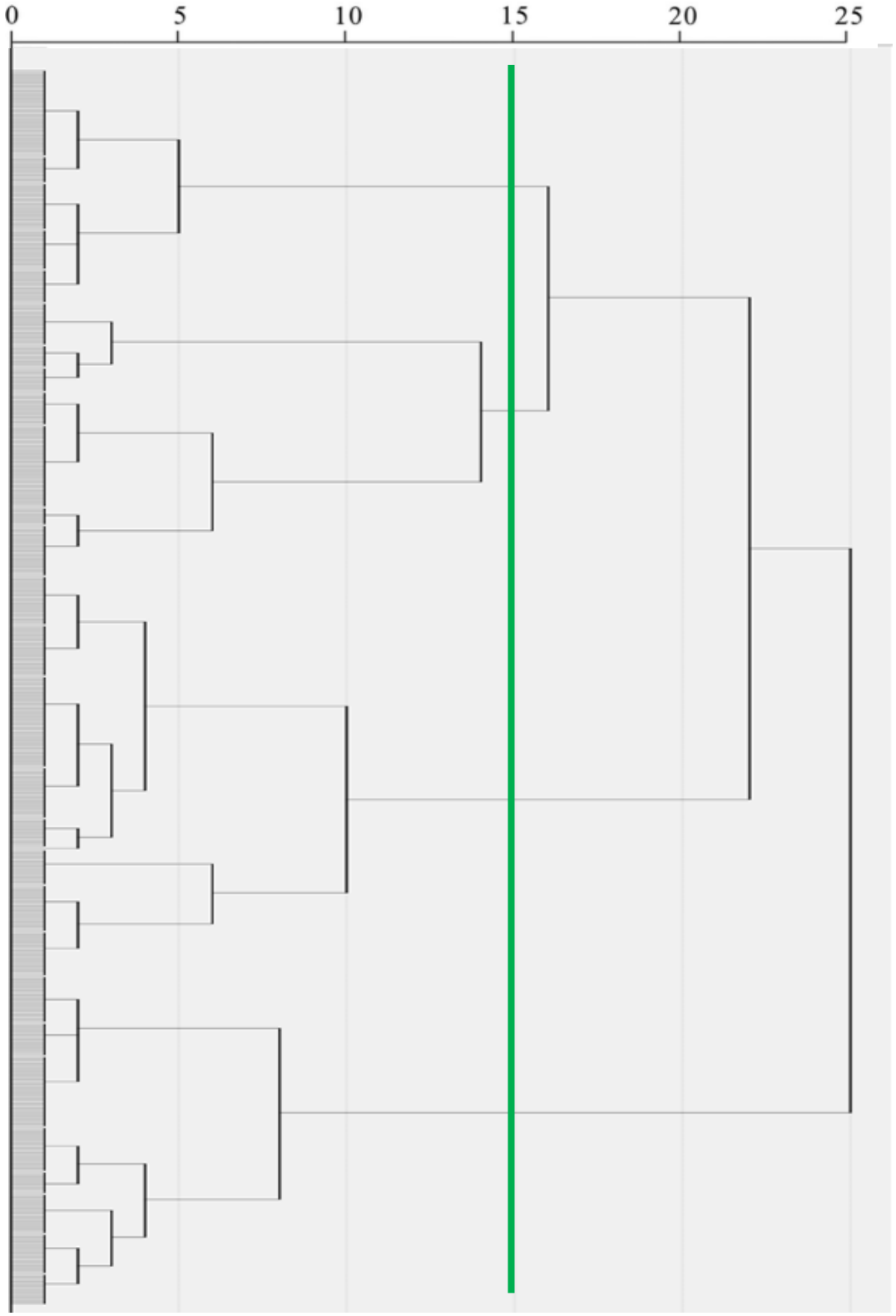
Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	89	762	0,000	0	0	215
2	712	742	0,000	0	0	4
3	700	733	0,000	0	0	6
4	572	712	0,000	0	2	9
…	…	…	…	…	…	…
548	100	264	979,817	538	517	551
549	4	10	1049,604	534	537	552
550	2	6	**1149,814**	546	541	555
551	21	100	1268,653	545	548	554
552	4	81	1439,170	549	542	553
553	4	22	1630,883	552	547	554
554	4	21	1906,088	553	551	555
555	2	4	2220,000	550	554	0

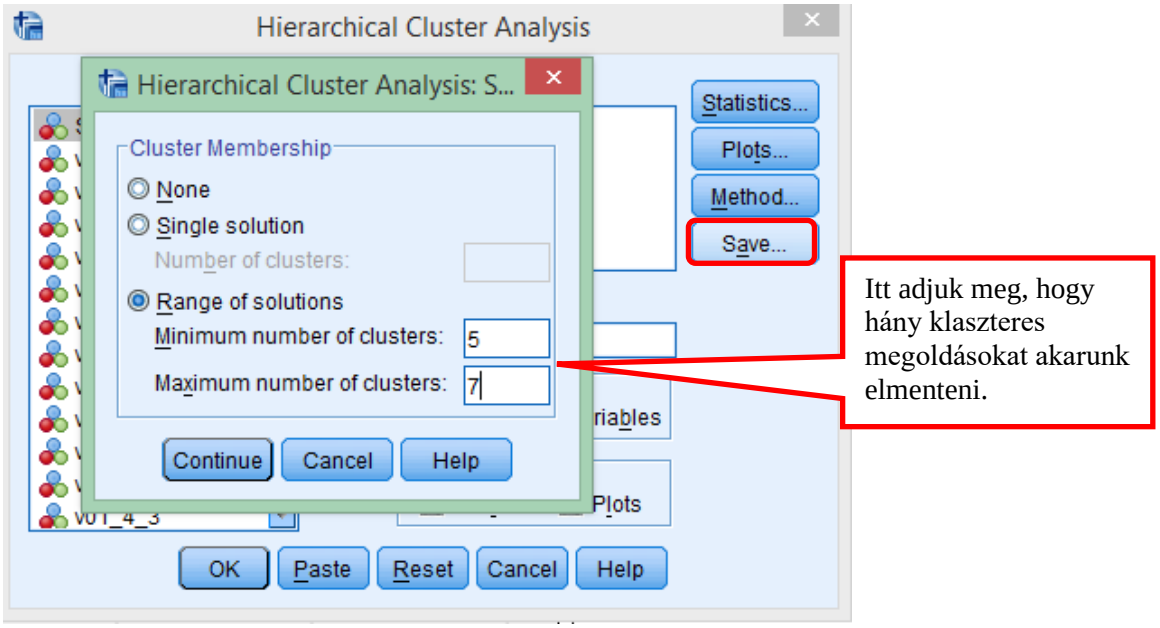
3. Dendrogram

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

Például, ha az összevonásnál maximum 15 távolságot engedünk meg, akkor a 4 klaszteres megoldást kell választanunk. (Ezt praktikusan úgy kapjuk meg, hogy a 15-ös távolságnál (zöld vonal) megnézzük, hogy hány vízszintes vonalat metszünk.)



Az 50%-os elv alapján a 6 klaszteres megoldás tűnik ideálisnak, most menjünk ezzel tovább. A tapasztalatok alapján ilyenkor érdemes a +/-1 „elvet” követve elmenteni (tehát az 5, 6, 7 klaszteres megoldást is) és megvizsgálni ezeket a megoldási opciókat is.
Ahhoz, hogy meg tudjuk vizsgálni, ezek közül melyik klaszterszám lesz számunkra a megfelelő, mind a 3 megoldást el kell mentenünk az *Analyze/Classify/Hierarchical Cluster* menü *Save* pontjában.



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

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

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

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

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

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

Célunk:

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

Analyze=>Compare Means => Means
dependent list: faktor változók
independent list: klaszter változók

Az eredményeinket megvizsgálva megállapítható, hogy minden esetben a csoportok átlagai között szignifikáns különbség van (sig=0,000), valamint a klaszterek elemszáma is hasonlóan alakul. Mivel az 5 klaszteres megoldás esetében csak egy esetben találkozunk 1-nél magasabb szórással, ezért ezzel a megoldással dolgozunk tovább.

A klaszterek elemszámai az öt klaszteres megoldás esetében rendre 148, 83, 180, 105, 40

Az első klaszter tagjainak kevésbé fontos az önkifejezés (-0,7619), míg az egyéni jólét számukra a legértékesebb (0,8997)

Az ötös klaszter tagjainak a legfontosabb a minőségre törekvés (1,1605), de a szórakozás faktor a számukra a legérdektelenebb (-1,5810)

Report

Ward	Method	Önkifejezés	Egyéni jólét faktor	Minőségre törekvés	Szórakozás faktor
1	Mean	-0,762	0,900	-0,401	0,075
	N	**148**	148	148	148
	Std. Deviation	0,671	0,699	0,908	0,827
2	Mean	-0,214	-1,224	-0,972	-0,370
	N	**83**	83	83	83
	Std. Deviation	0,351	0,564	0,345	1,127
3	Mean	0,949	0,296	0,266	-0,033
	N	**180**	180	180	180
	Std. Deviation	0,879	0,649	0,976	0,749
4		-0,349	-0,547	0,436	0,846
	N	**105**	105	105	105
	Std. Deviation	0,732	0,656	0,707	0,507
5	Mean	-0,089	-0,686	1,161	-1,581
	N	**40**	40	40	40
	Std. Deviation	0,713	0,829	0,575	0,954
Total	Mean	0,000	0,000	0,000	0,000

N		556	556	556	556
	Std. Deviation	1,000	1,000	1,000	1,000

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Önkifejezés * Ward Method	Between Groups	(Combined)	264,864	4	66,216	125,752	**,000**
	Within Groups		290,136	551	,527		
	Total		555,000	555			
Egyéni jólét faktor * Ward Method	Between Groups	(Combined)	310,284	4	77,571	174,658	**,000**
	Within Groups		244,716	551	,444		
	Total		555,000	555			
Minőségre törekvés * Ward Method	Between Groups	(Combined)	188,640	4	47,160	70,928	**,000**
	Within Groups		366,360	551	,665		
	Total		555,000	555			
Szórakozás faktor * Ward Method	Between Groups	(Combined)	187,558	4	46,890	70,314	**,000**
	Within Groups		367,442	551	,667		
	Total		555,000	555			

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

			Hippik	Otthon-ülők	Party arcok	Pedán-sak	Igénye-sek	Különbség értékelése
Klaszterelemzésbe bevont változók	N (fő)		148	83	105	180	50	
	Önkifejezés		-,762	-,214	-,349	,949	-,089	*Szig*
	Egyéni jólét faktor		,900	-1,224	-,547	,296	-,686	*Szig*
	Minőségre törekvés		-,401	-,9716	,436	,266	1,161	*Szig*
	Szórakozás faktor		,0753	-,370	,846	-,033	-1,581	*Szig*
Klaszterelemzésbe be nem vont változók	Válaszadó	Férfiak	47,3%	45,8%	60,0%	55,6%	47,5%	*Nem Szig.*
		Nők	52,7%	54,2%	40,0%	44,4%	52,5%	
	Átlagos életkor		**41,22**	**50,66**	47,32	45,93	47,25	***Szig.***
	Származási ország	Ausztria	20,9%	**31,3%**	27,6%	5,6%	10,0%	***Szig.***
		Németország	**26,4%**	10,8%	**30,5%**	17,8%	2,5%	
		Magyarország	17,6%	3,6%	4,8%	21,7%	**70,0%**	
		Írország	16,2%	**34,9%**	16,2%	25,6%	15,0%	
		Anglia	18,9%	19,3%	21,0%	**29,4%**	2,5%	
	Átlagos alkoholos költés		**5,78**	**5,48**	6,95	**9,29**	7,24	***Szig.***
	Átlagos alkoholmentes költés		**5,98**	**5,95**	6,80	7,91	**8,13**	***Szig.***

A fenti táblázatban összefoglaltuk, hogy a vizsgált változók kereszttábla elemzés vagy varianciaelemzés során szignifikáns kapcsolatot mutattak-e a klaszterváltozóval.
A klaszterelemzésbe bevont változók, jelen esetben a faktorok klasztereken belüli átlagai (ANOVA táblából kinyert érték) alapján fantázianeveket adunk az egyes klasztereknek. Ezeket a neveket az SPSS *Variable view* felületén is rögzítjük a „*Values*” oszlopban, hogy a későbbi elemzések során már ezek az elnevezések jelenjenek meg.
A klaszterelemzésbe be nem vont változók a klaszterek jellemzését segítik.

SYNTAX

```
*Klaszterelemzés.
CLUSTER   FAC1_1 FAC2_1 FAC3_1 FAC4_1
  /METHOD WARD
  /MEASURE=SEUCLID
  /PRINT SCHEDULE
  /PRINT DISTANCE
  /PLOT DENDROGRAM
  /SAVE CLUSTER(3,5).  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /SAVE REG(ALL)
  /METHOD=CORRELATION.
```

GYAKORLÓ FELADAT

1. Az AGYŐ szeretné megvizsgálni, hogy az egyének az italfogyasztási mennyiségeik alapján (v06) milyen szegmensekbe sorolhatóak. Az elemzésbe bevont változók mellett szeretné megismerni a létrehozott csoportok nemek szerinti összetételét, átlagos életkorát, származási országukat, valamint az átlagos alkoholos-, valamint alkoholmentes költésük összegét.
2. A Házi-TU Gyümölcslégyártó Kft. szeretné megvizsgálni, hogy milyen fogyasztói szegmensek léteznek a piacon a különböző italválasztási indokok alapján (v8_4_1-v8_4_17). Az előzetes feltáró kutatások 17 alapvető indokot azonosítottak, ami miatt kedvenc italnak választanak az egyének alkoholmentes italokat. A szegmensek kialakítása előtt a szegmensképző ismérveket szeretnék csoportokba sorolni. A végső megoldást a 60%-os magyarázott varianciahányad alapján szeretnék meghozni Varimax rotálással. A létrehozott új változók esetében megfigyelhető-e szignifikáns különbség a férfiak és a nők között (v17)? A különböző kedvenc alkoholmentes italok (v08_0) esetében eltérően alakulnak az adatok? A válaszadók életkora és az új faktorok között van-e kapcsolat (v18)? Értelmezze a faktorok alakulását a 18. sorszámú válaszadó esetében (Name: Ssz).

A létrehozott, látens változók alapján milyen 5 fogyasztói csoport azonosítható a válaszadók körében a Ward összevonást alkalmazva? Jellemezze a klasztereket az alábbi táblázat alapján.

			1. kl	2. kl	3. kl	4. kl	5. kl	Különbség értékelése
N (fő)								
Klaszterelemzésbe bevont változók								
Klaszterelemzésbe be nem vont változók	Válaszadó neme	Férfiak	%	%	%	%	%	
		Nők	%	%	%	%	%	
	Átlagos életkor							
	Származási ország	Ausztria	%	%	%	%	%	
		Németország	%	%	%	%	%	
		Magyarország	%	%	%	%	%	
		Írország	%	%	%	%	%	
		Anglia	%	%	%	%	%	
	Kedvenc alkoholmentes ital	*csak a TOP1 említése*	%	%	%	%	%	
	Átl. alkoholos költés							
	Átl. alkoholmentes költés							