

*Módszertani készségfejlesztés, többváltozós statisztikai eljárások - gyakorlat.

*Adatok előkészítése.

*FONTOS SZABÁLYOK SZINTAX-ÍRÁSBAN.

*1.

*A parancsírás a parancs nevével kezdődik (pl. VARIABLE LABELS), amit a SPSS kékre színez majd.

*Ezt követik a változók, melyeken a parancsot végre kell hajtani, (pl. szulev).

*Majd jönnek a beállítások. Ezek vagy zárójelben, vagy a következő sorban felsorolva vannak.

*FONTOS: ha a beállítások a következő sorba kerülnek, akkor egy SPACE kell minden sor elejére, ezzel jelölve, hogy ez még a fölötte lévő parancs alá tartozik'

*Végül a parancsot egy ponttal zárjuk le, így adva keretet az egésznek.

*2.

*A kommenteket csillaggal kell kezdeni, és ponttal kell zárni.

*kommentelni mindig fontos, hogy később is tudd, mit csináltál.

*3.

*A parancs neveket nagybetűvel szokás írni, így könnyebben átlátható kód készül.

*4.

*Fontosabb speciális szavak:

*TO - tartomány kijelölésére - pl. 1 TO 4 tartalmazza az 1 2 3 4-et. Változók megadásánál a változók dataeditorban megadott sorrendje számít.

*BY - valami mentén. Pl. a MEANS szulev BY nem. megadja a születési évek átlagát nemek mentén bontva.

*IF - ha feltétel.

*EZ A RÉSZT KÉZZEL ÁLLÍTOTTUK BE AZ ÓRÁN GYAKORLÁSKÉNT, DE HOGY KÉSŐBB TUDJUI VELE DOLGOZNI, ITT VAN SZINTAXBAN IS MIT CSINÁLTUNK.

* FORMATS nem(f1.0).

* VARIABLE LABELS nem

'A kérdezett neme'

.

* VALUE LABELS nem

1 'Férfi'

2 'Nő'

-6 'Nem kérdéztük'

.

* MISSING VALUES nem(-6).

* VARIABLE LEVEL nem(nominal).

*VÁLTOZÓK TULAJDONSÁGAINAK BEÁLLÍTÁSA.

*A FORMATS paranccsal a Variable view TYPE, WIDTH és DECIMALS oszlopa állítható.

*Az f4.0-ban az f betű jelentése, hogy standard számformátumot akarunk használni, a 4 a teljes hossz, és a 0 a tizedesjegyek száma.

*Ez így egy 4 karakter hosszúságú számot jelent tizedesjegyek nélkül.
FORMATS szulev(f4.0).

*A VARIABLE LABELS a címkéket állítja.

*Figyelj a végén a pontra!

VARIABLE LABELS szulev

'A kért szünetési év'

.

*A VALUE LABELS a különböző kódok jelentését állítja.

*Itt egyetlen számnak van külön jelentése, a -6-osnak, de pl. így állíthatnánk be, hogy 1 a férfit, 2 a nőt jelenti a nemnél.

VALUE LABELS szulev

-6 'Nem kértük'

.

*Hiányzó érték kódok adhatók meg. Ha több kódot akarunk megadni, a zárójelbe szóközzel elválasztva tehetjük meg, de működik a -6 TO 0 formátum is.

MISSING VALUES szulev(-6).

*A mérési szint adható meg. A három lehetséges érték: nominal, ordinal, scale.

VARIABLE LEVEL szulev(scale).

*ÚJRAKÓDOLÁS ÉS EGYÉB ADATTRASZFORMÁCIÓK.

*A településtípusok nem túl szerencsés sorrendben vannak megadva. Hiszen 2-es a megyeszékhely, 3-as a város, 4-es a falu, és 9-es Bp. Kódoljuk át őket úgy, hogy növekvő sorrendben legyenek!.

*Ez menüsorból is meghívható - Transform / Recode into Same Variables paranccsal, ahol meg kell adni, melyik kódból melyet szeretnénk csinálni.

*Az adatmanipulációs parancsok után ki kell adni az EXECUTE. parancsot, csak ezt követően kezd lefutni az adatmanipuláció.

RECODE teltip (2=3) (3=2) (4=1) (9=4).

EXECUTE.

*Az előző lépésben megváltoztattuk magukat a kódokat (pl a Budapest eddig 9-essel volt jelölve, most pedig 4-es lesz, de attól még a kódok elnevezései nem változtak. Ezeket is át kell írni a helyes értékekre.

VALUE LABELS teltip

1 'község'

2 'város'

3 'megyeszékhely'

4 'Budapest'

.

*Az adattraszformáció egy másik módja, amikor új változót szeretnénk számolni. Például a kérdőív az adatbázisban csak a születési év van megadva, de mi a korral szeretnénk dolgozni.

*Tudjuk, hogy a felmérés 2014-ben volt, tehát a kort úgy lehet kiszámolni, hogy 2014-ből kivonjuk a születési évet.

*Menüül a Transform / Compute Variable ablakban érhető el.

*Mivel adattraszformáció, itt is kell a végére EXECUTE.

*FIGYELJÜK MEG, mi történik a hiányzó értékekkel!

COMPUTE kor=2014-szulev.

EXECUTE.

*A könnyebb megtalálás érdekében a létrejött kor változót húzzuk a változók sorában előre!

*Beállítjuk a kor változó tulajdonságait is.

FORMATS kor(f3.0).

VARIABLE LABELS kor

'A kérdezett kora'

.

VALUE LABELS kor

-6 'Nem kérdezettük'

.

MISSING VALUES kor(-6).

VARIABLE LEVEL szulev(scale).

*Észrevesszük, hogy az oi változók helytelenül nominálisra vannak állítva. Az összes ilyen változót egyszerre átállíthatjuk ordinálisra.

*A beírást segíti, ha használjuk a Variables menüpontot a változónevek kimásolására.

VARIABLE LEVEL oi1a TO oi5c(ordinal).

*Észrevesszük azt is, hogy a legfontosabb személyek jellegénél a -6 nincs hiányzó értéknek beállítva. Tegyük meg ezt is.

*LEÍRÓ STATISZTIKÁK - TÁBLÁZATOS MEGJELENÍTÉS.

*Több formában is elérhető nagyjából ugyanaz az információ.

*Itt négy kikérési módot nézünk meg: Descriptive, Frequencies, Explore (Examine) és Crosstabs.

*Descriptives - Menüsorból az Analyze / Descriptive Statistics / Descriptive-ben található.

*Egyszerű, jól rendezett leíró statisztikák.

*Plusz funkció, itt lehetne egy változót z-értékekre konvertálni.

DESCRIPTIVES VARIABLES=nem kor teltip

/STATISTICS=MEAN STDDEV MIN MAX SEMEAN KURTOSIS SKEWNESS.

*Frequencies - Menüsorból az Analyze / Descriptive Statistics / Frequencies-ben található.

*Plusz funkció - gyakoriságtáblák - minden egyes érték előfordulási valószínűsége - nominális

változóknál jó.

*Plusz funkció - kvartilisek, percentilisek, és más osztók itt kérhetőek ki.

*grafikonok is kérhetőek.

FREQUENCIES VARIABLES=nem kor teltip

/NTILES=4

/STATISTICS=STDDEV MINIMUM MAXIMUM SEMEAN MEAN MEDIAN SKEWNESS SESKEW KURTOSIS SEKURT

/PIECHART PERCENT

/FORMAT=LIMIT(10)

/ORDER=ANALYSIS.

*Explore (Examine)- Menüsorból az Analyze / Descriptive Statistics / Explore-ban található.

*Plusz funkció - részletesebb statisztikák.

*Statisztikák valamilyen bontás alapján BY paranccsal (következő parancsban látunk rá példát..

*Hiányzó értékek kezelése többféleképp.

EXAMINE VARIABLES=nem kor teltip

/PLOT NONE

/PERCENTILES(5,10,25,50,75,90,95) HAVERAGE

/STATISTICS DESCRIPTIVES

/CINTERVAL 95

/MISSING LISTWISE

/NOTOTAL.

*A leíró statisztikákat megkaptuk három menüpontból is (Explore, Descriptive és Frequencies). Kicsit máshogy néznek ki, de ugyanazt az információt tartalmazzák.

*A táblázatok szerkezetét utólag is megváltoztathatjuk, hogy a számunkra legkönnyebben legyen átlátható.

*Ehhez kétszer kattintsuk a táblázatra, majd jobb gomb megnyomása után válasszuk a Pivoting-ot. Itt a táblázat egységeit (statisztikák, változók, stb.) szabadon húzhatjuk oszlopba, sorba vagy layerbe.

*Statisztikák nem szerint bontva.

EXAMINE VARIABLES=szulev teltip BY nem

/PLOT NONE

/PERCENTILES(5,10,25,50,75,90,95) HAVERAGE

/STATISTICS DESCRIPTIVES

/CINTERVAL 95

/MISSING LISTWISE

/NOTOTAL.

*Nem csak egy, de akármennyi bontást kérhetünk, ezek faszerkezet szerint fognak szétbomlani.

EXAMINE VARIABLES=szulev BY nem BY teltip

/PLOT NONE

/PERCENTILES(5,10,25,50,75,90,95) HAVERAGE

/STATISTICS DESCRIPTIVES

/CINTERVAL 95

/MISSING LISTWISE

/NOTOTAL.

*Kereszt táblák. Nominális és néha ordinális változóknál van értelme. Előfordulási gyakoriságot

tartalmaznak a cellák.

CROSSTABS

```
/TABLES=laptop BY nem  
/FORMAT=AVALUE TABLES  
/CELLS=COUNT  
/COUNT ROUND CELL.
```

*De néha százalékosan is lehet értelme megjeleníteni.

*Százalékot kérhetsz csak a sorra (ekkor azt tudod meg, hogy a laptophasználók és laptopot nem használók hány százaléka férfi, hány százaléka nő).

*Kérheted oszlopra (ekkor azt tudod meg, hogy a férfiak és nők hány százaléka használ laptopot).

*Kérheted totalra (ekkor azt tudod meg, hogy a minta hány százaléka laptophasználó férfi, hány százaléka laptophasználó nő, laptopot nem használó férfi, és nem használó nő).

*Gondold végig, a kutatói kérdésednek melyik megjelenítés felel meg jobban!.

CROSSTABS

```
/TABLES=laptop BY nem  
/FORMAT=AVALUE TABLES  
/CELLS=COLUMN  
/COUNT ROUND CELL.
```

*A 100%-okat tartalmazó oszlopra vagy sorra nincs szükség, ezt el is rejthetjük.

*Kattinsunk a táblázatra kétszer, hogy szerkeszthessük, majd a kérdéses oszlop vagy sor kijelölése után jobb gomb megnyomásával válasszuk a Hide category-t!.

*LEÍRÓ STATISZTIKÁK - GRAFIKUS MEGJELENÍTÉS.

*Egy minta megjelenítése.

*Nominális (kategorialis) változó - pie chart vagy bar chart használható. Mikor melyik?.

*Ha minden ember értékét ismered, lehet pie chart, ha vannak hiányzó értékek a mintádban, akkor vagy olyan pie chart, ami őket is megjeleníti, vagy bar chart.

*Például nézzük meg frekvencia táblával az első és második legfontosabb személy jellegét!.

```
FREQUENCIES VARIABLES=legfontcsop1 legfontcsop2  
/ORDER=ANALYSIS.
```

*Láthatod, hogy az első legfontosabb embernél mind a 1007 személy válaszolt. A második legfontosabbnál viszont van 365, aki nem válaszolt.

*Tehát az első legfontosabb mehet pie-ra - ezt a Graphs / Legacy Dialogs / Pie-ban találod.

*Pie-chartnál gondold át, tényleg kell-e, nem elég csak szövegesen megadni az arányokat.

*Bár nagy a kísértés rá, a kördiagrammot NEM SZABAD 3D-ben használni, mert félrevezető, és nem illik fellobbantani sem.

GRAPH

```
/PIE=COUNT BY legfontcsop1.
```

*A második legfontosabb is mehet pie-ra , ha a hiányzókat is megjeleníted (ezt az Options-ban tudod beállítani).

GRAPH

```
/PIE=COUNT BY legfontcsop2  
/MISSING=REPORT.
```

*Vagy rakd Bar chart-ra. *Menüsorból a Graphs / Legacy dialogs / Bar-ban találjátok.

GRAPH

/BAR(SIMPLE)=COUNT BY legfontcsop2.

*Ordinális változót Bar charton ábrázoljuk.

*Azt hogy ordinális a változó, azzal jelezzük, hogy az oszlopok között nullára állítjuk a térközt.

*Ezt úgy érjük el, hogy a grafikonra kétszer kattintva szerkesztjük, majd az oszlopokra kattintva felhozzuk a beállítási lehetőségeket, és a Bar options-ban a Bars-t 100%-ra állítjuk.

GRAPH

/BAR(SIMPLE)=COUNT BY teltip.

*Skála típusú változót hisztogrammon ábrázoljuk.

*Menüsorból a Graphs / Legacy dialogs / Histogram.

GRAPH

/HISTOGRAM=baratsag.

*A hisztogram

*****.

*Több minta megjelenítése.

*független mintákat valamely folytonos tulajdonságát oszlopdiagrammon szokás ábrázolni.

*Az átlag mellé ábrázolni kell valamelyik hibasávot is, és hogy melyik hibasávot használtad, a függőleges tengelyen a változó után kell írni zárójelben.

*Menüsorból a Graphs / Legacy dialogs / Bar-ban találjátok.

GRAPH

/BAR(SIMPLE)=MEAN(kor) BY teltip

/INTERVAL CI(95.0).

*Ha több szempont szerint szeretnétek csoportosítani, akkor Clustered Bar chartot kell választani, amit a Graphs / Legacy dialogs / Bar-ban a Clustered kiválasztásával tehettek meg.

*Syntaxban egy újabb BY hozzáadásával érhető el ugyanaz.

*3D grafikont NE használjatok, mert nem lehet róla az értékeket jól leolvasni!.

GRAPH

/BAR(GROUPED)=MEAN(kor) BY teltip BY nem

/INTERVAL CI(95.0).

*Ha ordinális változót szeretnénk megjeleníteni, arra egyfajta megoldás lehet a boxplot, mely a medián és a negyedelő pontokat ábrázolja.

*Menüből a Graphs / Legacy dialogs / Boxplot-ban lehet kiválasztani, ha több csoportot akarunk ábrázolni, akkor groups of cases, ha több változót, akkor separate variables választásával.

*ha még egy csoportosító tényezőt hozzá akarnánk adni, azt ugyanúgy plusz egy BY segítségével megtehetjük.

EXAMINE VARIABLES=iskola BY teltip

/PLOT=BOXPLOT

/STATISTICS=NONE

/NOTOTAL.

*Összefüggő mintákat vonaldiagrammal szokás ábrázolni, az összekötő vonallal jelezve, hogy ezek

összefüggő adatok. Természetesen erre is ki kell tenni a hibamutatót.

*Menüből a Graphs / Legacy dialogs / Line-ban a Summaries for separate variables kiválasztásával.

*Azt hogy hibásávként mit használtunk, az SPSS alulra írja ki, de APA formátum szerint a függőleges tengelyen a helye.

GRAPH

```
/LINE(SIMPLE)=MEAN(oi1a) MEAN(oi1b) MEAN(oi1c)
/MISSING=LISTWISE
/INTERVAL CI(95.0).
```

*Ha több csoport ismételt teljesítményét akarjuk ábrázolni, akkor a csoportokat külön vonallal ábrázoljuk, ezt a BY használatával érhetjük el.

GRAPH

```
/LINE(MULTIPLE)= MEAN(oi1a) MEAN(oi1b) MEAN(oi1c) BY nem
/MISSING=LISTWISE
/INTERVAL CI(95.0).
```


*Változók közötti kapcsolat.

*Folytonos változó - scatter dot-on ábrázolható.

*Menüből Graphs / Legacy dialogs / Scatter/Dot. Bár - mint a korrelációnál sem, itt sem következtethetünk ok-okozati kapcsolatra, az Y tengelyre az kerül, amit inkább a függő változóként tudunk elképzelni.

GRAPH

```
/SCATTERPLOT(BIVAR)=baratsag WITH elegendettseg
/MISSING=LISTWISE.
```

*A grafikont meg is színezhajjuk, ha Marker-ként megadunk egy harmadik változót - itt csak kategóriálisat lehet.

GRAPH

```
/SCATTERPLOT(BIVAR)=baratsag WITH elegendettseg BY nem
/MISSING=LISTWISE.
```

*Ha folytonos változót szeretnénk színeként használni, akkor a Legacy dialogs helyett az azt lassan felváltó GGRAPH-t használjuk.

*Menüből a Graphs / Graphboard Template Chooser -ből érjük el, ahol a két vizsgálandó változót kijelölve a felajánlott grafikonok közül a scatterplot mátrix választjuk.

*A harmadik, színezést adó változót a Detailed fülön adhatjuk meg.

GGRAPH

```
/GRAPHDATASET NAME="graphdataset"
  VARIABLES=optimizmus[LEVEL=scale] baratsag[LEVEL=scale] elegendettseg[LEVEL=scale]
  MISSING=LISTWISE REPORTMISSING=NO
/GRAPHSPEC SOURCE=VIZTEMPLATE(NAME="Scatterplot"[LOCATION=LOCAL]
  MAPPING( "color"="optimizmus"[DATASET="graphdataset"] "x"="baratsag"[DATASET="graphdatas
t"]
  "y"="elegendettseg"[DATASET="graphdataset"]))
VIZSTYLESHEET="Traditional"[LOCATION=LOCAL]
LABEL='SCATTERPLOT: baratsag-elegendettseg'
DEFAULTTEMPLATE=NO.
```

*A grafikont érdemes átszínezni - a fekete-fehér átmenet általában jól működik nyomtatásban.

*Több változó kapcsolatát is ábrázolhatjuk, ha Simple Scatter helyett Matrixot kérünk.

GRAPH

```
/SCATTERPLOT(MATRIX)=baratsag elegendettseg optimizmus
```

```
/MISSING=LISTWISE.
```

*Ennek egy kiegészített változatát kapjuk, ha a Legacy dialogs helyett az azt lassan felváltó GGRAPH-t használjuk.

*Menüből a Graphs / Graphboard Template Chooser -ből érjük el, ahol a három vizsgálandó változót kijelölve a felajánlott grafikonok közül a scatterplot mátrix választjuk.

*Ha színezni is szeretnénk, azt a Detailed fülön tehetjük meg.

*A szerkesztőfelülete kicsit más, de minden megtalálható rajta, ami a chartokon ott volt - meg kell szokni.

GGRAPH

```
/GRAPHDATASET NAME="graphdataset"
```

```
VARIABLES=baratsag[LEVEL=scale] optimizmus[LEVEL=scale] nem[LEVEL=nominal]
```

```
elegendettseg[LEVEL=scale]
```

```
MISSING=LISTWISE REPORTMISSING=NO
```

```
/GRAPHSPEC SOURCE=VIZTEMPLATE(NAME="Scatterplot Matrix (SPLOM)"[LOCATION=LOCAL]
```

```
MAPPING( "all"="baratsag"[DATASET="graphdataset"]
```

```
"all"="elegendettseg"[DATASET="graphdataset"] "all"="optimizmus"[DATASET="graphdataset"]
```

```
"color"="nem"[DATASET="graphdataset"]))
```

```
VIZSTYLESHEET="Traditional"[LOCATION=LOCAL]
```

```
LABEL='SCATTERPLOT MATRIX (SPLOM): optimizmus-baratsag-elegendettseg'
```

```
DEFAULTTEMPLATE=NO.
```

*Van, hogy a scatter/dot nem jól látható, mert nem elég finom egyik vagy akár mindkettő skála.

Lássunk erre egy példát.

*Kíváncsiak vagyunk, hogy a technikai eszközökkel való ellátottság és a kapcsolati háló nagysága között milyen az összefüggés.

*A kapcsolati háló nagyságát tudjuk a fontos személyek számából - ez egy 0-5ig terjedő érték lehet - ez egy bevett módja a kapcsolati háló mérésének.

*A technikai eszközök számát azonban még ki kell számolnunk. Hogy tehetjük meg? Gondoljuk végig, hogy minden esetben jól működik ez a módszer?.

```
COMPUTE infokomm=telefon+mob+pc+mail+face+laptop+jatekkonzol.
```

```
EXECUTE.
```

*A kapott változót próbáljuk meg scatterdoton ábrázolni! - Használjuk a Graphboard template choosert

GGRAPH

```
/GRAPHDATASET NAME="graphdataset"
```

```
VARIABLES=infokomm[LEVEL=nominal] legfontszam[LEVEL=scale]
```

```
MISSING=LISTWISE REPORTMISSING=NO
```

```
/GRAPHSPEC SOURCE=VIZTEMPLATE(NAME="Scatterplot"[LOCATION=LOCAL]
```

```
MAPPING( "x"="legfontszam"[DATASET="graphdataset"] "y"="infokomm"[DATASET="graphdataset"
```

```
))
```

```
VIZSTYLESHEET="Traditional"[LOCATION=LOCAL]
```

```
LABEL='SCATTERPLOT: legfontszam-infokomm'
```


DEFAULTTEMPLATE=NO.

*A probléma az, hogy az értékek átfednek egymással, így nem látjuk, hol hány személy van - nem látjuk a pontfelhő alakját.

*Ha kicsit megmozgatnánk a pontokat, csak annyira, hogy ne legyenek egymással tökéletes átfedésben, láthatóvá tehetnénk az összefüggést. A pontok megmozgatását jitternek nevezzük.

*Lépünk be a szerkesztőbe, jelöljük ki a pontfelhőt, és a modifier-ben válasszuk ki valamelyik jittert. Végezzük el a többi - láthatóan szükséges - szerkesztést is.

*Itt hagyok egy olyan verziót is, amely menüből nem érhető el, de utána nem graph-ot, hanem chartot eredményez, és így könnyebben szerkeszthető.

*Használni úgy kell, hogy X és Y helyére beírod azt a változót, aminek az összefüggését nézed. Figyelj mert két helyen kell átírni X-et és Y-ot!.

*X-labelje és Y-labelje helyére beírod a tengelyfeliratokat.

*Ha akarsz színeznit, azt a SZÍN átírásával teheted meg. Ha nem akarsz, akkor ki kell törölnöd három dolgot:.

*a VARIABLES közül a SZIN-t.

*a következő sort: DATA: szin1 = col(source(s), name("SZIN"), unit.category()) .

*az utolsó előtti sorból a color(szin1), részt.

*Ha akarsz feliratokat, azt a SZOVEG helyére kell beírnod. Ha nem akarsz, a színhez hasonlóan három helyről kell a szövegre vonatkozó parancsokat törölni.

*Meg persze vedd ki a GGRAPH és a BEGIN GPL elől a csillagokat, hogy ne komment legyen:).

*GGRAPH

```
/GRAPHDATASET NAME="graphdataset" VARIABLES=X Y SZIN SZOVEG
```

```
/GRAPHSPEC SOURCE=INLINE.
```

*BEGIN GPL

```
SOURCE: s=userSource(id("graphdataset"))
```

```
DATA: x1=col(source(s), name("X"))
```

```
DATA: y1=col(source(s), name("Y"))
```

```
DATA: szin1 = col( source(s), name( "SZIN" ), unit.category() )
```

```
DATA: szoveg1=col(source(s), name("SZOVEG"), unit.category())
```

```
GUIDE: axis(dim(1), label("X-labelje"))
```

```
GUIDE: axis(dim(2), label("Y-labelje"))
```

```
ELEMENT: point.jitter(position(x1*y1), color(szin1), label(szoveg1))
```

```
END GPL.
```

*Végül két ordinális vagy kategoriális változó kapcsolatát hődiagrammon is ábrázolhatjuk. Ugyan van a Graphbuilderben is heat-map, de nem teljesen az, ami nekünk kell, ezért az Excelhez fogunk folyamodni.

*Kérjünk ki egy keresztábrát a végzettség és településtípus összefüggésére. A cellákba gyakoriságokat megjelenítve.

CROSSTABS

```
/TABLES=telTip BY iskola
```

```
/FORMAT=AVALUE TABLES
```

```
/CELLS=COUNT
```

```
/COUNT ROUND CELL.
```

*kattintsunk a táblázatra jobb gombbal, válasszuk a Copy Specialt, és a kimásolt táblázatot illesszük be excelbe. Formázzuk!.

*A hődiagrammot a Kezdőlap / feltételes formázásában találjuk, ahol válasszuk színskálákat és a további beállításokat.

*Ha a számokat szeretnénk eltüntetni, akkor jelöljük ki a hődiagrammot, jobb gomb, cellaformázásban

válasszuk az egyéni formátumot, és írjuk be három pontosvesszőt!.