

Az IBM SPSS használatának alapjai

Az IBM SPSS egy széles körben elterjedt statisztikai elemző szoftver. A szoftver működése közben több ablakot/fájlt használ, ebből számunkra kettő fontos:

1. az adatok tárolása egy **adatmátrixban, adattáblában** történik. Ennek kiterjesztése sav. Emellett az SPSS nagyon sokféle adatállományt tud kezelni.

Az SPSS adattáblájának kétféle nézete van. Ezeket úgy lehet elképzelni, mintha két munkalapunk lenne. A **Data View** nézetben láthatjuk a tárolt elemi adatokat, törölhetünk, szerkeszthetünk adatokat. Ez gyakorlatilag egy olyan táblázatot jelent, ahol az egyes **rekordokban (sorokban)** az egyes megfigyelések, azaz az **egyedek** találhatóak, míg az mindegyik **változót (ismérvet)** egy-egy **oszlop** reprezentál, míg. Ha új változót szeretnénk létrehozni, akkor elég egy üres oszlopba begépelni az adatokat. Azonban, ügyeljünk arra, hogy a változó tulajdonságait megadjuk. Ha egy változót törölni szeretnénk, akkor kijelöljük az oszlopot és kitoröljük.

	Azonosító	KezdőfizetésUSD	Nem	Életkor	Jelenlegi fizetésUSD	Iskolábneltöltöttidőév	var	var	var	var	var	var	var	var	var
1	626	5400	férfi	20-29	10680	12									
2	627	5100	nő	50-59	8940	15									
3	628	8400	férfi	20-29	16080	16									
4	629	5400	férfi	30-39	11640	12									
5	630	24000	férfi	40-49	41400	16									
6	631	4800	nő	60-69	8580	12									
7	632	10200	férfi	30-39	21960	15									
8	633	8700	férfi	30-39	19200	16									
9	634	6996	férfi	40-49	13320	16									
10	635	17400	férfi	40-49	28350	19									
11	636	5100	férfi	40-49	7860	12									
12	637	12996	férfi	20-29	27250	18									
13	638	6420	férfi	50-59	10500	8									
14	639	4800	nő	20-29	9900	12									
15	640	4800	férfi	20-29	11340	8									
16	641	6900	férfi	20-29	16080	15									
17	642	5700	nő	30-39	10620	15									
18	643	6600	férfi	30-39	11220	15									
19	644	4500	nő	20-29	7860	12									
20	645	4500	nő	20-29	8700	12									
21	646	5280	nő	40-49	8760	8									
22	647	4080	nő	40-49	6960	12									
23	648	4800	nő	20-29	10980	8									
24	649	5400	férfi	20-29	14100	15									

A **Variable View** nézetben adhatjuk meg a változók tulajdonságait. Amennyiben új változót hozunk létre, meg kell adnunk, illetve meglévő változók esetében be kell állítanunk a **változó fontosabb tulajdonságait**, így például:

- a változó **nevét (name)**, amivel azonosítani tudjuk.
- A változó tárolási formáját (**Type**). Ez nagyon sokféle lehet, például szöveg, szám, pénznem, stb. Elemzések során a változók lehetséges értékeit számkódokkal tárolják: például a nem esetében nem férfi és nő, hanem ezek helyett egy-egy számkód (pld. 1-2 kerül) rögzítésre.
- A változó formai beállításait. Például hány tizedes jegy pontosságú értékekkel dolgozunk, vagy maximum hány karakter hosszúságú szöveget tárolunk, stb.
- A változó **mérési szintjét**. A mérési szint határozza meg azt, hogy a változót milyen elemzésekbe, és ezeken belül milyen szerepkörbe lehet bevonni. Az SPSS a mérési szintek jelölésére grafikus jelet alkalmaz: nominális esetben halmazok, ordinális esetben oszlopok, míg skálák esetében egy vonalzó szimbolizálja a változót. Amennyiben a nominális/ordinális változó szöveggként van tárolva, úgy az ikonján megjelenik egy „a” betű is.

- A **számkódok (value)** jelentését, ha van. Például, ha az elemzésekben szerepel változóként a nem, akkor beállíthatjuk azt, hogy a 0 szám(jegy) a nőket, míg az 1 szám(jegy) a férfiakat jelentse.
- A **hiányzó adatok (missing value)** helyettesítési körét. Ha egy megfigyelés során hiányzik egy adatunk és ezt jelölni szeretnénk, akkor ezt egy általunk meghatározható kóddal lehet helyettesíteni. Például, ha egy kérdőíven kérdésként az szerepel, hogy „Értékelje 1-4 között 4 fokú skálán főnökét!”, és a válaszok között meg akarjuk különböztetni azt, hogy ki miért nem válaszol, akkor lehetőségünk olyan alternatívák felvételére is: 5: Nem tudom megítélni; 6: Megtagadom a választ.

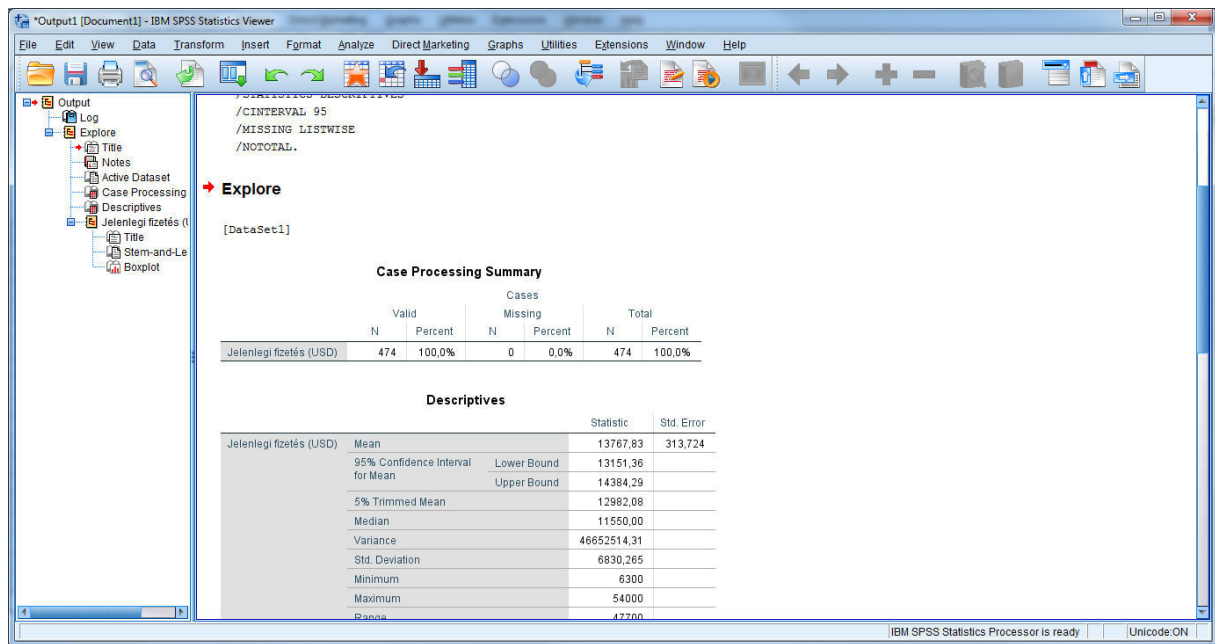
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Azonosító	Numeric	11	0		None	None	11	Right	Nominal	Input
2	Kezdfizeté...	Numeric	11	0	Kezdő fizetés (USD)	None	None	11	Right	Scale	Input
3	Nem	String	6	0		None	None	6	Left	Nominal	Input
4	Életkor	String	5	0		None	None	5	Left	Ordinal	Input
5	Jelenlegi fizeté...	Numeric	11	0	Jelenlegi fizetés (USD)	None	None	11	Right	Scale	Input
6	Iskolábanelt...	Numeric	11	0	Iskolában eltöltött idő (év)	None	None	11	Right	Scale	Input
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											

Az SPSS alkalmazásakor leggyakrabban az alábbi menüket használjuk.

- **File:** adatállományok betöltése, mentése.
- **Transform:** a változók transzformálása, átkódolása.
- **Analyze:** az elemzések tárháza.
- **Graphs:** a grafikus ábrázolási lehetőségek tárháza.

Az SPSS moduláris szerkezetű, ami azt jelenti, hogy az egyes menüpontokban csak azok az almenük érhetőek el, amelyekhez hozzáférési jogosultsággal rendelkezünk.

2. Az SPSS elemzések kimeneteit egy külön **output ablakban** kapjuk meg. Ennek alatermezett kiterjesztése spv. Ezeket a kimeneteket akár Word, akár Excel, akár Html dokumentumként, akár képként is exportálhatjuk.



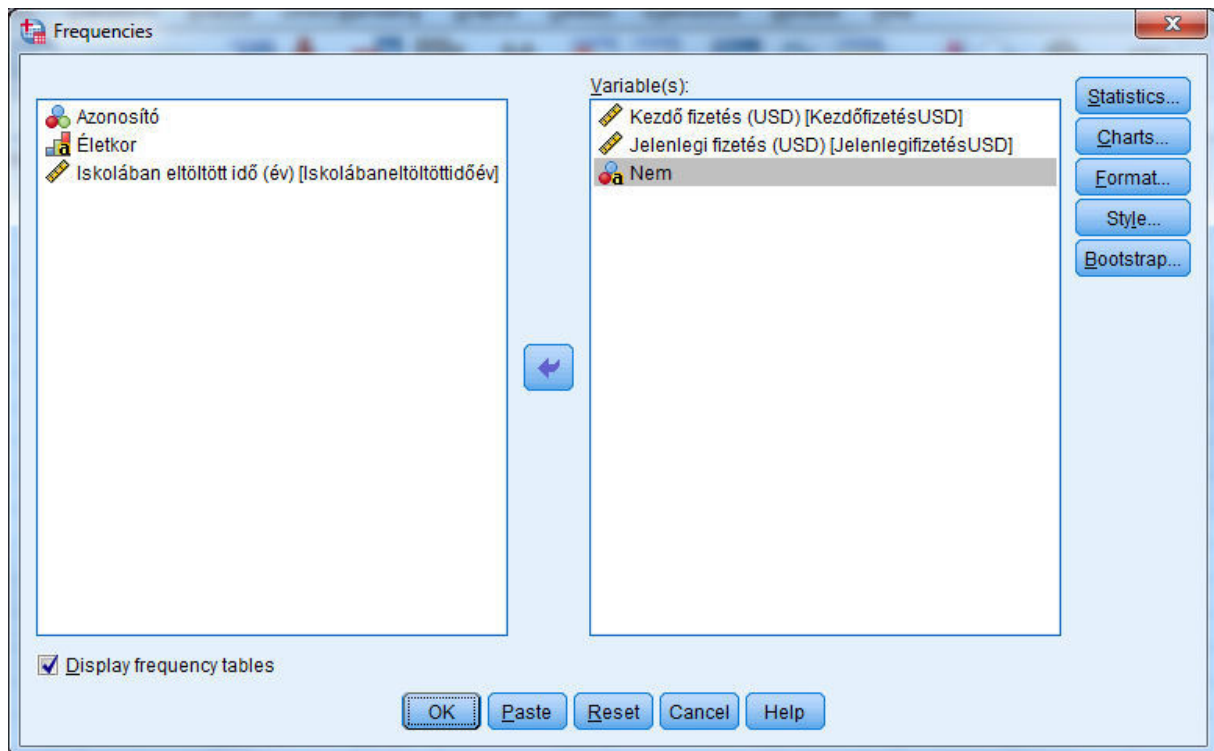
Fontos megjegyezni, hogy az SPSS azt feltételezi, hogy FAE-mintával dolgozunk! A kimenetek elemzésekor ügyelnünk kell arra, hogy egyszerre jelenhetnek meg leíró és induktív statisztikai elemek is.

Leíró statisztika eljárások

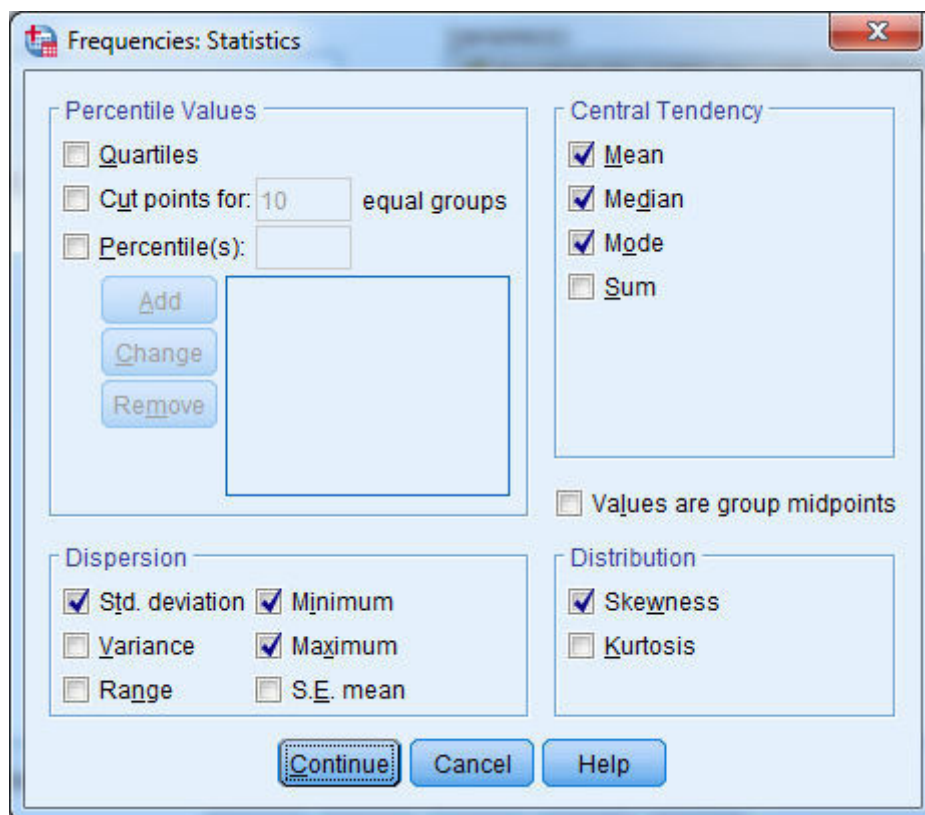
Leíró statisztikai eljárások végrehajtására az *Analyze/Descriptives Statistics* menüpont alatt található lehetőségek közül választhatunk.

A menüpont kiválasztása után először baloldalon ki kell választanunk azokat a változókat, amelyeken leíró statisztikai elemzést szeretnénk végrehajtani és a jobboldalra kell őket áthelyeznünk. Ha egyszerre több változót szeretnénk kijelölni, akkor ezt az egérrel megtehetjük, a CTRL gomb folyamatos nyomva tartása mellett.

1. **Frequencies** menüpont kiválasztásával is végezhetünk leíró statisztikai elemzést. Az adott változó gyakorisági táblázatát is itt szoktuk elkészíteni.



Az *Statistics* gomb megnyomása után adhatjuk meg a kiszámítandó mutatószámok körét.



Ezek közül a fontosabbak:

- Sum: a változó értékeinek összege.

- Mean, várható érték: ez a mintaátlag, mely párhuzamosan a sokasági átlag pontbecslése.
- Median: a medián.
- Modus: a módusz.
- Minimum, maximum: a változó legkisebb és a legnagyobb értéke.
- Std. Deviation: ez az SPSS-ben a korrigált tapasztalati szórás értékét jelenti, azaz a sokasági szórás pontbecslését kapjuk meg eredményként.
- Kurtosis, csúcsosság: egy eloszlás csúcsosságának megállapítása. Az alapértelmezésként használt mutató pozitív értéke csúcsosabb, míg negatív értéke lapultabb eloszlást jelez.
- Skewness: egy eloszlás aszimmetriájának megállapítása. Az alapértelmezésként használt mutató pozitív értéke baloldali, míg negatív értéke jobboldali aszimmetriát mutat.

		Statistics		
		Kezdő fizetés (USD)	Jelenlegi fizetés (USD)	Nem
N	Valid	474	474	474
	Missing	0	0	0
Mean		6806,43	13767,83	
Median		6000,00	11550,00	
Mode		6000	12300	
Std. Deviation		3148,255	6830,265	
Skewness		2,853	2,125	
Std. Error of Skewness		,112	,112	
Minimum		3600	6300	
Maximum		31992	54000	

A kimeneten látható, hogy szöveggként tárolt változó esetében az alapstatisztikákat nem számította ki a program. Ugyanakkor, ha számkódokat használjuk, akkor ezeket kiszámítja a program, annak ellenére, hogy nincs értelme.

A kimenet második táblázata a gyakorisági tábla.

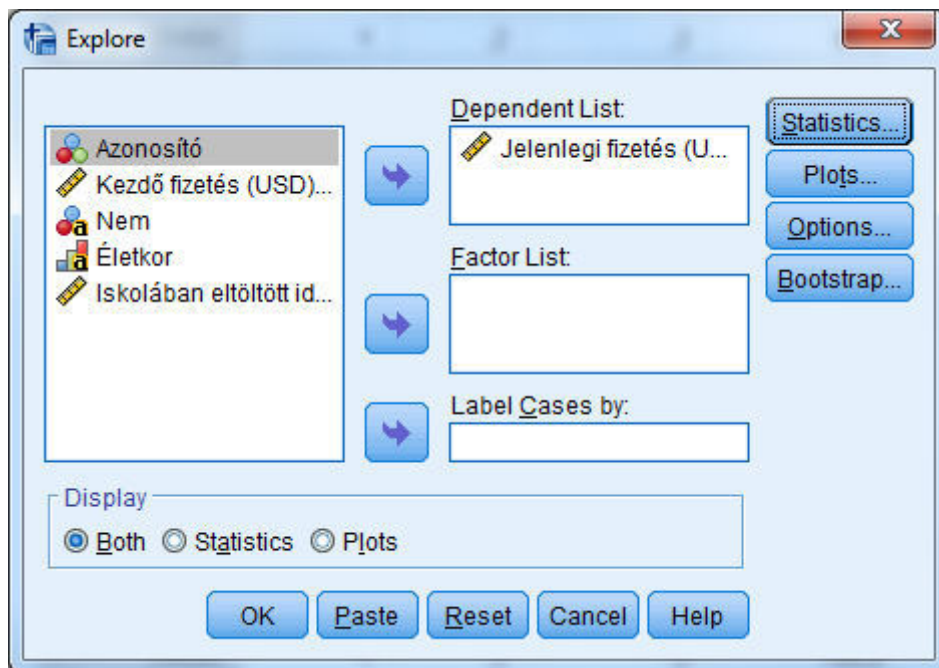
		Nem			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	férfi	258	54,4	54,4	54,4
	nő	216	45,6	45,6	100,0
Total		474	100,0	100,0	

Amennyiben van hiányzó értékünk a percent és a Valid percent oszlop elválík. A percent az összes megkérdezett, míg a Valid percent az érvényes válaszok százalékában mutatja a megoszlást.

Előadó szakmai felkészütsége

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4	5	5,9	6,0
	5	78	91,8	100,0
	Total	83	97,6	100,0
Missing	9	2	2,4	
Total	85	100,0		

2. Az **Explore elemzési módszer** segítségével szintén egy változó leíró statisztikai elemzését végezhetjük el, de az előzőektől eltérően – ha *Factor List* változót is megadunk – ezt már részsokaságokként is megtehetjük.



Például, ha a jelenlegi fizetéseket vizsgáljuk, akkor az eljárás outputja az alábbi.

Descriptives

		Statistic	Std. Error
Jelenlegi fizetés (USD)	Mean	13767,83	313,724
	95% Confidence Interval for Mean	Lower Bound	13151,36
		Upper Bound	14384,29
	5% Trimmed Mean	12982,08	
	Median	11550,00	

Variance	46652514,31	
	0	
Std. Deviation	6830,265	
Minimum	6300	
Maximum	54000	
Range	47700	
Interquartile Range	5265	
Skewness	2,125	,112
Kurtosis	5,378	,224

A kimenet fontos eleme, hogy a jelenlegi fizetés átlagának pontbecslése (13767,83) megkapjuk ennek standard hibáját (313,724) a $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ képlet alapján, és az alapsokaság átlagának 95 százalékos (ez az érték a statistics... opció alatt módosítható) intervallumbecslését is a $(\text{pontbecslés} - \text{hibahatár}; \text{pontbecslés} + \text{hibahatár})$, ahol a $\Delta = t_{0,975}(n-1) \cdot s_{\bar{x}}$ összefüggés alapján. Ahogyan tárgyaltuk ennek a képletnek van alkalmazási feltétele. Emlékeztetőül:

Mintaátlagok eloszlása

Mintanagyság	Eloszlás
Nagyminta ($n \geq 100$)	Megközelítőleg normális eloszlás (Centrális határeloszlás tétel alapján), tehát k értékét <ul style="list-style-type: none"> - ha ismerjük az alapsokaság szórását normális eloszlás alapján - ha nem ismerjük az alapsokaság szórását t-eloszlás alapján számítjuk ki.
$30 \leq n < 100$	<ul style="list-style-type: none"> - Ha nincs erős baloldali aszimmetria (mutatók értéke < 1), akkor megközelítőleg normális eloszlás, tehát k értékét ha ismerjük az alapsokaság szórását normális eloszlás alapján - ha nem ismerjük az alapsokaság szórását t-eloszlás alapján számítjuk ki.
$30 < n$	Megegyezik a vizsgált változó eloszlásával. k értékét az eloszlásnak megfelelően számítjuk ki.

Mivel 474 elemű mintánk van, így az alkalmazási feltétel teljesül, így értelmezhetjük a program által kiírt eredményt. E szerint a sokaságban (és nem a 474 elemű mintában!) a jelenlegi fizetések átlaga 95 százalékos valószínűséggel 13151,36 USD és 14384,29 USD közé esik.

A fejezet zárásaként javasoljuk a feltöltött bank.xls fájl alapján a Frequency és az Explore funkciót kipróbálását. Eredményként a fejezetet szemléltető minta kimenetet kell megkapni.

SZEGEDI TUDOMÁNYEGYETEM
GAZDASÁGTUDOMÁNYI KAR
KÖZGAZDÁSZ KÉPZÉS
TÁVOKTATÁSI TAGOZAT
LECKESOROZAT
COPYRIGHT © SZTE GTK 2017/2018

A LECKE TARTALMA, ILLETVE ALKOTÓ ELEMEI ELŐZETES,
ÍRÁSBELI ENGEDÉLY MELLETT HASZNÁLHATÓK FEL.

JELLEN TÁNYAG
A SZEGEDI TUDOMÁNYEGYETEMEN KÉSZÜLT
AZ EURÓPAI UNIÓ TÁMOGATÁSÁVAL.
PROJEKT AZONOSÍTÓ: EFOP-3.4.3-16-2016-00014

SZÉCHENYI 



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE