

Többváltozós statisztika, ANOVA és adatredukciós módszerek (BMNPS07700M)

Soltész-Várhelyi Klára

Főkomponens elemzés
Faktor elemzés

SD = Strongly Disagree, D = Disagree, N = Neither, A = Agree, SA = Strongly Agree

Példa SPSS-ben

	SD	D	N	A	SA
1 Statistics make me cry	☐	☐	☐	☐	☐
2 My friends will think I'm stupid for not being able to cope with SPSS	☐	☐	☐	☐	☐
3 Standard deviations excite me	☐	☐	☐	☐	☐
4 I dream that Pearson is attacking me with correlation coefficients	☐	☐	☐	☐	☐
5 I don't understand statistics	☐	☐	☐	☐	☐
6 I have little experience of computers	☐	☐	☐	☐	☐
7 All computers hate me	☐	☐	☐	☐	☐
8 I have never been good at mathematics	☐	☐	☐	☐	☐
9 My friends are better at statistics than me	☐	☐	☐	☐	☐
10 Computers are useful only for playing games	☐	☐	☐	☐	☐
11 I did badly at mathematics at school	☐	☐	☐	☐	☐
12 People try to tell you that SPSS makes statistics easier to understand but it doesn't	☐	☐	☐	☐	☐
13 I worry that I will cause irreparable damage because of my incompetence with computers	☐	☐	☐	☐	☐
14 Computers have minds of their own and deliberately go wrong whenever I use them	☐	☐	☐	☐	☐
15 Computers are out to get me	☐	☐	☐	☐	☐
16 I weep openly at the mention of central tendency	☐	☐	☐	☐	☐
17 I slip into a coma whenever I see an equation	☐	☐	☐	☐	☐
18 SPSS always crashes when I try to use it	☐	☐	☐	☐	☐
19 Everybody looks at me when I use SPSS	☐	☐	☐	☐	☐
20 I can't sleep for thoughts of eigenvectors	☐	☐	☐	☐	☐
21 I wake up under my duvet thinking that I am trapped under a normal distribution	☐	☐	☐	☐	☐
22 My friends are better at SPSS than I am	☐	☐	☐	☐	☐
23 If I am good at statistics people will think I am a nerd	☐	☐	☐	☐	☐

- tobbval2_03_fokomponens02_SAQ
- Andy Field példafájlja
- „SPSS szorongást” mérő kérdőív

Feltételek

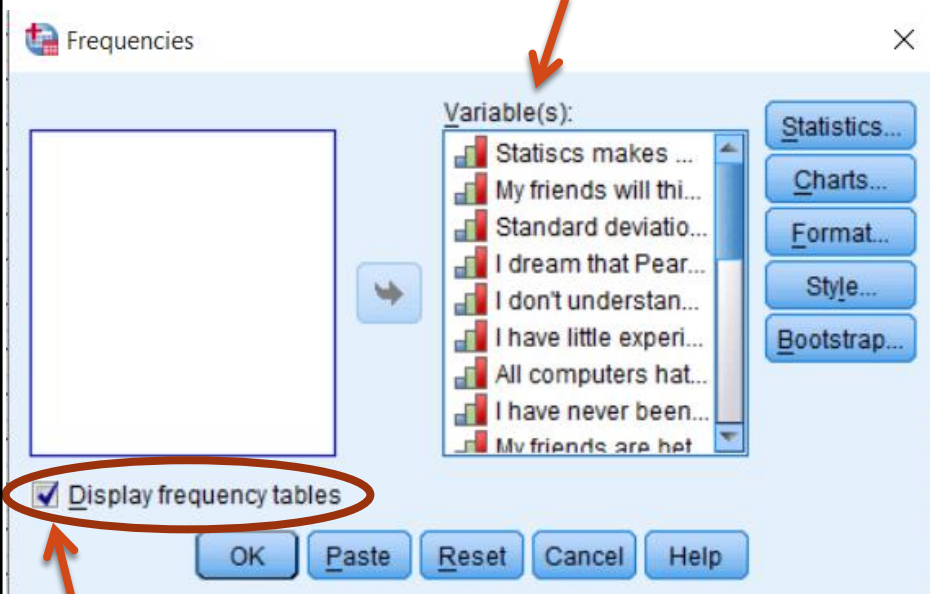
- Elemzésenként változik, de van néhány minden elemzésre igaz kitétel:
 - **Változók legalább intervallum típusúak (vagy legalább 5fokú Likert skálás adatok)**
 - **Változók normális eloszlásúak (min 5fokú Likertnél nem követeljük meg)**
 - A Likert skála annyira zárt tartományú, hogy jelentős ferdeség nem alakulhat ki. Azt azért nézzük meg, hogy olyan item, aminek nincs varianciája (az emberek legalább 90%-a ugyanazt a választ adta), ne kerüljön az elemzésbe.
 - **Korreláció**
 - Az itemek között közepes/közepesen erős korrelációkat várunk
 - **Túl gyenge korreláció** probléma
 - ha egyáltalán nincs korreláció az itemek között az baj, mert minden item külön faktort fog alkotni. Ellenőrizni a Bartlett teszttel fogjuk
 - **Túl erős korreláció** probléma
 - Ha túl erős a korreláció, szintén baj, mert redundáns információt viszünk be a modellbe.
 - Kollinearitás: két item között $r > .8$ korreláció van. Ellenőrizni a korrelációs táblával fogjuk.
 - Multikollinearitás: egy item nagyon sok másik itemmel korrelál erősen ($r > .5$). Ellenőrizni a korrelációs tábla és a korrelációs determinánssal fogjuk.

- **Elemsszám:**
 - A korrelációs együtthatók változnak mintáról mintára, ezért megbízható faktoranalízishez nagy elemsszám kell
- **Általános szabály:**
 - 10-15 fő/változó
 - Legalább 300
- **Másik megközelítés: melyik faktorok értelmezhetőek**
 - Azok a faktorok, melyre legalább 4 változó tölt legalább .6-os töltéssel, elemsszámtól függetlenül megbízhatóak
 - Azok a faktorok, melyekre legalább 10 változó tölt legalább .4-es töltéssel 150 fő felett megbízhatóak
- **Harmadik megközelítés: változók kapcsolatának függvényében**
 - Ha a változók között a kommunalitás 0.6-nál nagyobb, elég 100 fő
 - Ha átlagosan 0.5-ös kommunalitás van a változók között 150 fő kell
 - .5 alatt 500 fő ajánlott
- **Kaiser-Meyer-Olkin mérőszám / KMO: elemsszám jósága**
 - a változók közötti korreláció és a változók közötti parciális korreláció aránya (a változók közötti korrelációk mekkora része magyarázható lineáris regresszióval)
 - 0-1-ig terjed – 0 teljesen megbízhatatlanok a faktorok, 1 - tökéletes
 - 0.5-0.7 éppen elfogadható – a faktorokat óvatosan kell értelmezni
 - 0.7 felett jó

- Analyze / Descriptive Statistics / Frequencies
 - Ellenőrizzük, hogy minden itemnek van-e varianciája

SPSS-ben

Ide kerülnek az itemek



Ezzel kérjük ki a gyakoriságokat

Ellenőrizzük, hogy nincs-e olyan item, ahol az emberek több, mint 90%-a ugyanazt a választ adja!

Az ilyen itemeket ki kell venni az elemzésből.

Frequency Table

Statiscs makes me cry

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly agree	270	10,5	10,5	10,5
	Agree	1338	52,0	52,0	62,5
	Neither	735	28,6	28,6	91,1
	Disagree	187	7,3	7,3	98,4
	Strongly disagree	41	1,6	1,6	100,0
	Total	2571	100,0	100,0	

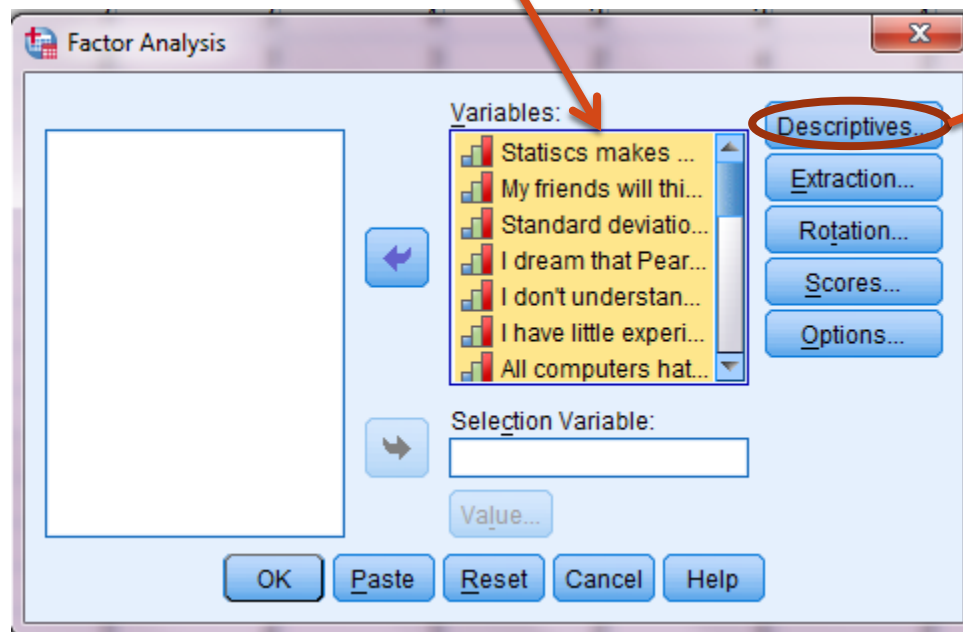
My friends will think I'm stupid for not being able to cope with SPSS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly agree	1436	55,9	55,9	55,9
	Agree	808	31,4	31,4	87,3
	Neither	206	8,0	8,0	95,3
	Disagree	101	3,9	3,9	99,2
	Strongly disagree	20	,8	,8	100,0
	Total	2571	100,0	100,0	

- Analyze / Dimension Reduction / Factor
 - Deskriptív: főleg mutatók annak ellenőrzésére, hogy jók-e az adatok FA-hoz, megbízható FA-t kapunk-e

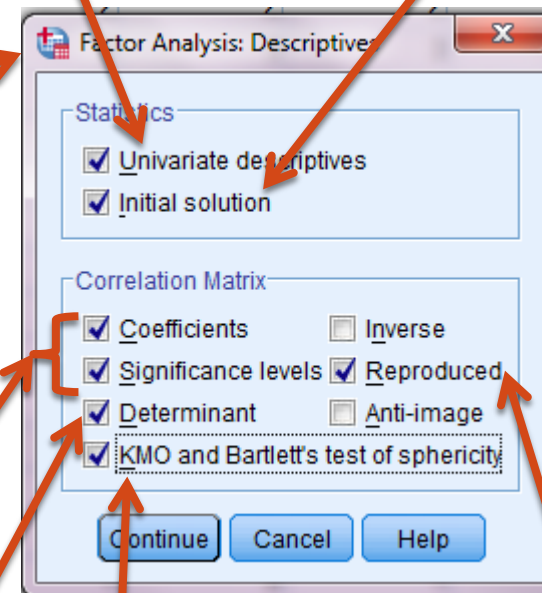
SPSS-ben

Ide kerülnek az itemek



Itemek egyszerű deskriptív statisztikája

Elemzés eredménye



Változók közötti korrelációs mátrix r és p értékei
Ellenőrzés túl alacsony & magas korrelációkra az adatban

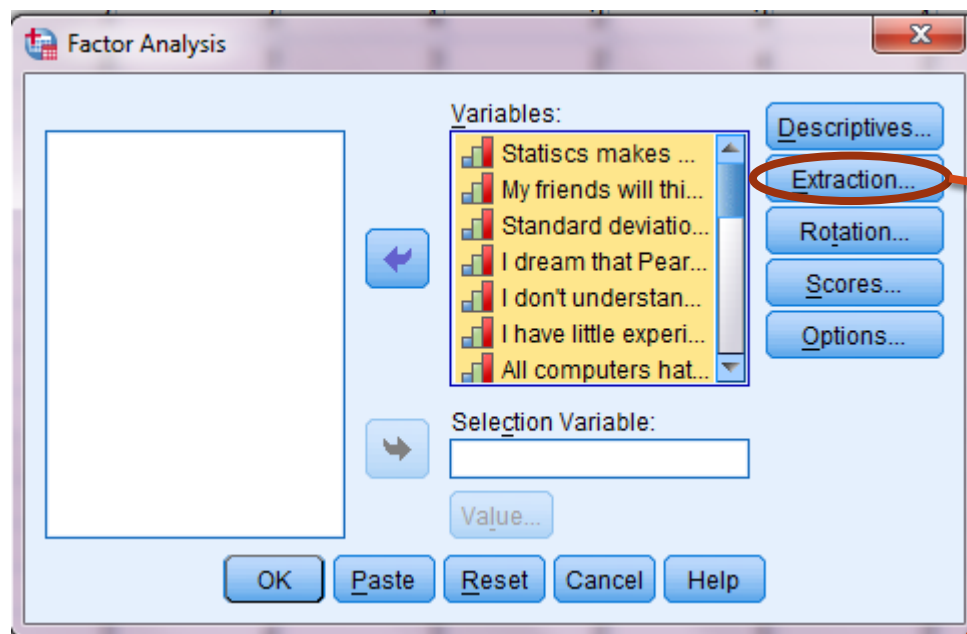
KMO

Korrelációs mátrix determinánsa – ha .00001 alatti érték, akkor kell a korr. táblát túl erős korrelációk után átnézni

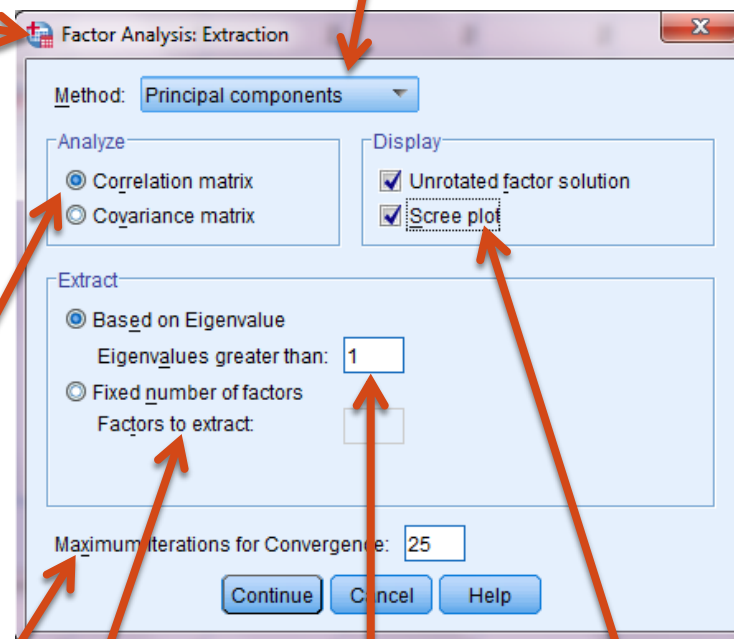
Mennyire tér el az elemzés alapján létrehozható korrelációs tábla az eredetitől. Legtöbb érték 0.05 alatt legyen

- Extraction: a FA menetéhez tartozó tulajdonságok beállítása

SPSS-ben



Itt választhatjuk ki, milyen módszert használunk



Min fusson az FA: korrelációs vagy kovariancia adatokon? A korrelációs mátrix a kovariancia mátrix „standardizált” formája – érdemes ezt használni

Hány kört fusson – átl. 25 elég – nagyon nagy adathalmaznál lehet több

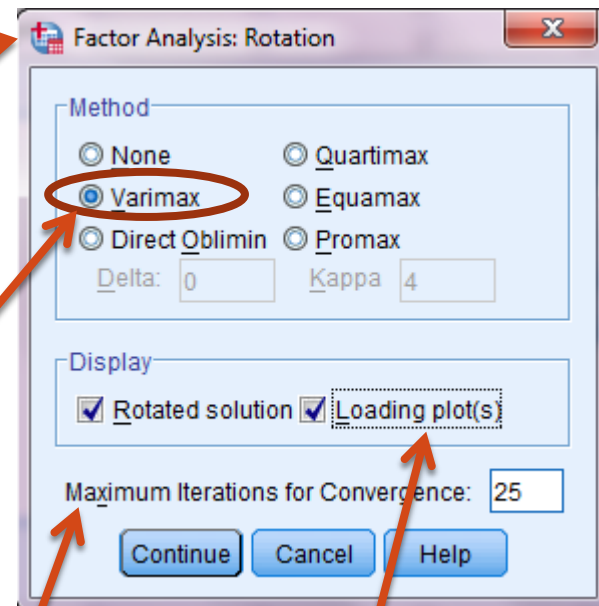
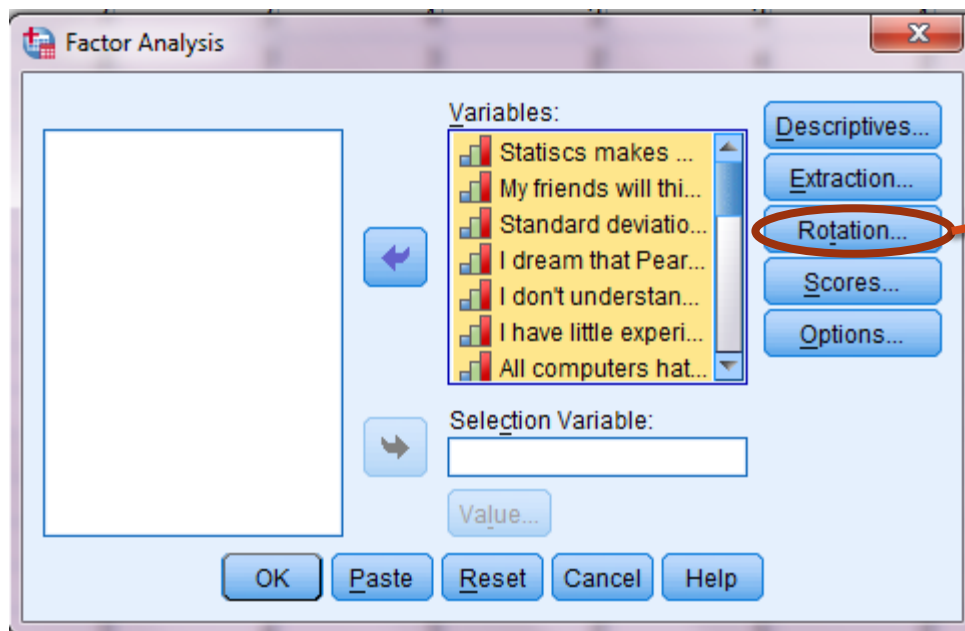
Scree plot

Melyik faktorokat tartsa meg:
Melyek legalább 1 sajátértékkel rendelkeznek (Kaiser kritérium)

X db faktort szeretnénk

- Rotation: forgatás tulajdonságainak beállítása

SPSS-ben



A faktorok korrelálhatnak-e? Legtöbbször a következő két rotációt használjuk:

Ha független faktorokat várok - varimax.

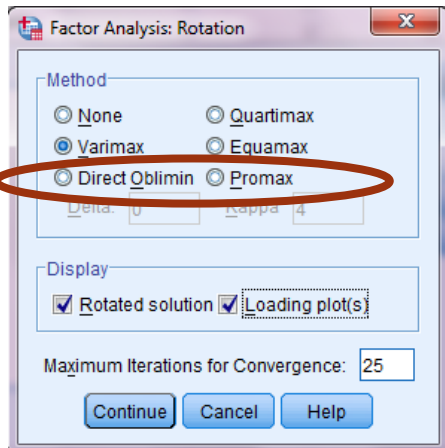
Ha korreláló faktorokat várok – obimin

(Mi most független forgatást kérünk ki, de nem azért mert az bármivel jobb lenne, mint a ferde. Ha vársz korrelációt a skálák között bátran használd a ferde forgatást!)

A faktorok mentén ábrázolja a változókat (max 3 faktorig)

Hány kört fusson – átl. 25 elég – nagyon nagy adathalmaznál kellhet több

SPSS-ben



- Ha ferde forgatást csinálsz, be kell állítani, milyen mértékű módosítást engedsz meg. Jól működő értékek:
 - Direct Obliminnél delta = 0 és Promaxnál Kappa = 4



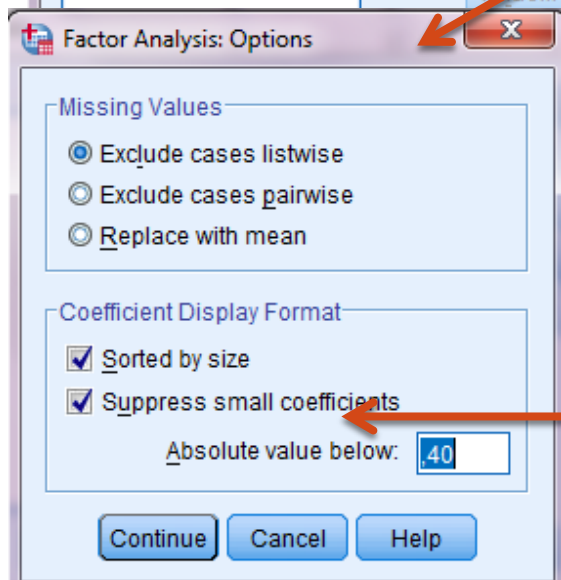
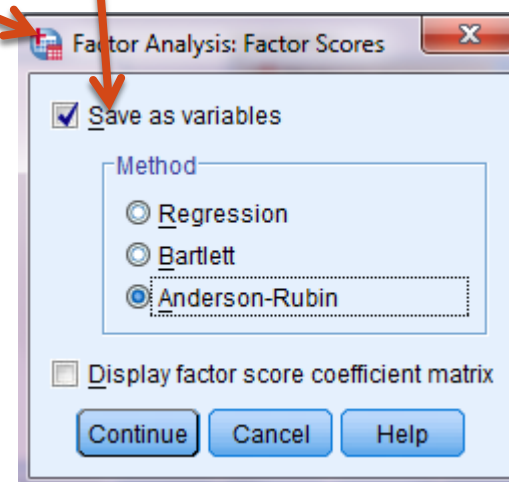
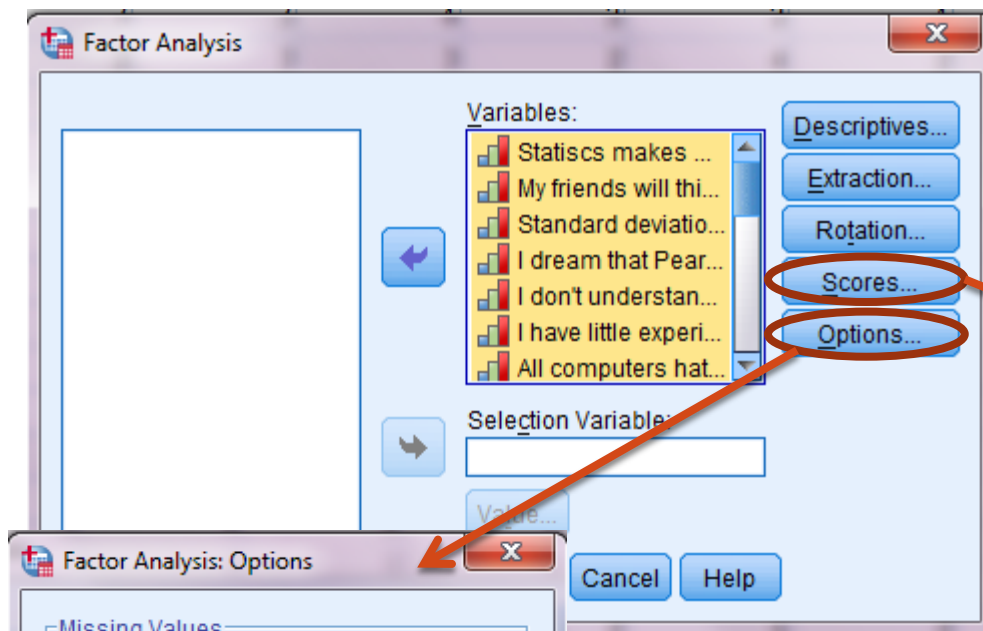
- A Promax azt csinálja, hogy varimax forgatással elforgatja a tengelyeket, majd a faktortöltéseket hatványozza kappával (kappadikra emeli őket). Ezáltal a kicsi töltések kicsik maradnak, a nagyok a hatványozás miatt sokkal nagyobbak lesznek, és kiemelődik a különbség a kicsi és nagy töltések között. Ezt követően visszaskálázza a tengelyt. A kappa azt adja meg, hányadikra emeljük a töltéseket.
- A Direkt oblimin úgy forgat, hogy a töltések között minél kisebb legyen az interakció. A delta értéke a tengelyek közötti maximálisan megengedhető korrelációt adja meg. Értéke $-\infty$ és 0.8 között lehet. Minél negatívabb a szám, annál kevésbé korrelálnak a tengelyek, minél nagyobb, annál erősebben. 0 felett annyira erős korreláció jön ki, hogy a tengelyeket nem lehet elkülöníteni, ezért a delta ajánlott maximális értéke 0 (ez erős korrelációkat, de értelmezhető tengelyeket fog eredményezni). Hogy negatívban meddig érdemes elmenni, az az adatoktól függ.



SPSS-ben

- Scores: kész faktorok változókká alakításához szükséges beállítások
- Options: itt állíthatjuk be az értelmezést segítő kinézetet

Létrehozza a faktoroknak megfelelő változót, és megadja minden személyhez, hogy milyen faktorérték tartozik hozzá. A módszer regresszió, ha megengedjük, hogy a faktorok korreláljanak. A módszer Anderson-Rubin, ha a faktorok függetlenségét szeretnénk



Értelmezést segíti

A változókat a hatásuk szerint rendezzi sorba, és azokat, melyek faktoregyütthatója .4-nél kisebb, nem mutatja

- Deskriptív statisztikák

Descriptive Statistics			
	Mean	Std. Deviation	Analysis N
Statistics makes me cry	2,37	,828	2571
My friends will think I'm stupid for not being able to cope with SPSS	1,62	,851	2571
Standard deviations excite me	2,59	1,075	2571
I dream that Pearson is attacking me with correlation coefficients	2,79	,949	2571
I don't understand statistics	2,72	,965	2571
I have little experience of computers	2,23	1,122	2571
All computers hate me	2,92	1,102	2571
I have never been good at mathematics	2,24	,873	2571
My friends are better at statistics than me	2,85	1,263	2571
Computers are useful only for playing games	2,28	,877	2571
I did badly at mathematics at school	2,26	,881	2571
People try to tell you that SPSS makes statistics easier to understand but it doesn't	3,16	,916	2571
I worry that I will cause irreparable damage because of my incompetence with	2,45	,949	2571

- **Korrelációs tábla** – (multi)kollinearitásra ellenőrzünk
 - Keressük a .8-nál nagyobb korrelációt
 - Az olyan változót, ami majdnem mindennel .5-nél erősebben korrelál – ha a táblázat alján determináns 0.00001-nél nagyobb, rendben

		Statiscs makes me cry	My friends will think I'm stupid for not being able to cope with SPSS	Standard deviations excite me	I dream that Pearson is attacking me with correlation coefficients	I don't understand statistics	I have little experience of computers
Correlation	Statiscs makes me cry	1,000	-,099	-,337	,436	,402	,217
	My friends will think I'm stupid for not being able to cope with SPSS	-,099	1,000	,318	-,112	-,119	-,074
	Standard deviations excite me	-,337	,318	1,000	-,380	-,310	-,227
	I dream that Pearson is attacking me with correlation coefficients	,436	-,112	-,380	1,000	,401	,278
	I don't understand						
	thinking that I am trapped under a normal distribution	,000	,000	,000	,000	,000	,000
	My friends are better at SPSS than I am	,000	,000	,000	,000	,000	,000
	If I'm good at statistics my friends will think I'm a nerd	,410	,000	,000	,043	,017	,000

a. Determinant = ,001

- **KMO és Bartlett's test**
 - **KMO érték** – a változók közötti korreláció és a változók közötti parciális korreláció aránya - 0-1-ig terjed – 0 teljesen megbízhatatlanok a faktorok, 1 - tökéletes
 - **Bartlett's test** – túl kicsi korrelációkra / függetlenségre ellenőrzünk
null hipotézis: a változók teljesen függetlenek egymástól (rossz, mert akkor minden változó külön komponens)
ha a teszt szignifikáns – feltételezhetjük, hogy valamennyi korreláció van a változók között, elvégezhető a főkomponens elemzés

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,930
Bartlett's Test of Sphericity	Approx. Chi-Square	19334,492
	df	253
	Sig.	,000

Értelmezés

- **Kommunalitás**

- A változók varianciája mennyiben közös más változókéval

- **Initial** – eredetileg : főkomponens elemzést végzünk, ahol feltételezzük, hogy a változók teljes varianciája közös, ezért itt az initial mindenhol 1

- **Extracted** - a faktorok kiválasztása után: pontosabb kép a közös varianciáról
Másképp nézve megadja, hogy a bennmaradó komponensek a változó varianciájának hány százalékát magyarázzák. Lényeges komponensek kiválasztása – adatvesztéssel is jár – már nem tudjuk tökéletesen megmagyarázni a változókat.

Communalities

	Initial	Extraction
Statistics makes me cry	1,000	,435
My friends will think I'm stupid for not being able to cope with SPSS	1,000	,414
Standard deviations excite me	1,000	,530
I dream that Pearson is attacking me with correlation coefficients	1,000	,469
I don't understand statistics	1,000	,343
I have little experience of computers	1,000	,654
All computers hate me	1,000	,545
I have never been good at mathematics	1,000	,739
My friends are better at statistics than me	1,000	,484
Computers are useful only for playing games	1,000	,335
I did badly at mathematics at school	1,000	,690

Értelmezés

- **Total variance Explained**

- Mivel 23 változónk volt, 23 komponens keletkezett
- Ezeket a sajátértékük szerint csökkenő sorrendben látjuk. A Kaiser kritérium alapján azok a komponensek maradnak, amiknek a sajátértéke 1 fölött van, most 4 komponens.

Initial Eigenvalues – a komponensek eredetileg

Rotated – elforgatott értékek

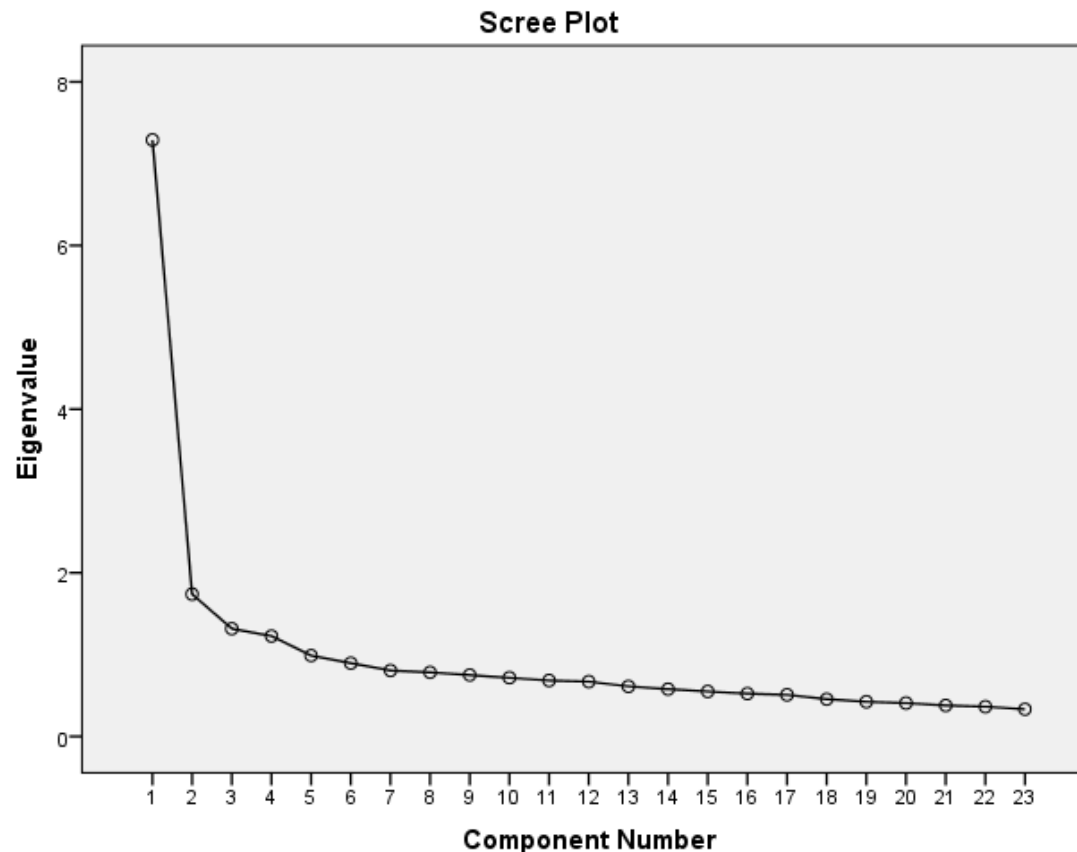
Sajátérték

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	7.290	31.696	31.696	7.290	31.696	31.696	3.730	16.219	16.219
2	1.739	7.560	39.256	1.739	7.560	39.256	3.340	14.523	30.742
3	1.317	5.725	44.981	1.317	5.725	44.981	2.553	11.099	41.842
4	1.227	5.336	50.317	1.227	5.336	50.317	1.949	8.475	50.317
5	.988	4.295	54.612						
6	.895	3.893	58.504						
7	.806	3.502	62.007						
8	.783	3.404	65.410						
9	.751	3.265	68.676						
10	.717	3.117	71.793						
11	.684	2.972	74.765						
12	.670	2.911	77.676						
13	.612	2.661	80.337						
14	.578	2.512	82.849						
15	.549	2.388	85.236						
16	.523	2.275	87.511						
17	.508	2.210	89.721						
18	.456	1.982	91.704						
19	.424	1.843	93.546						
20	.408	1.773	95.319						
21	.379	1.650	96.969						
22	.364	1.583	98.552						
23	.333	1.448	100.000						

Mennyit magyaráz a teljes varianciából
Látjuk, hogy a forgatás után
egyenletesebben oszlik el

- **Scree dot**

- Az elemzés a Kaiser kritérium alapján 4 komponenst tartott meg, a Scree dot alapján szemmel mi is ellenőrizhetjük ennek helyes voltát.
- A Scree dot inkább annak ellenőrzésére van, hogy nem siklott valami nagyon félre, ne ez alapján döntsünk a komponensek számáról
- Két törés is van, egy a 2. után, egy a 4. komponens után – a 4 komponens elfogadható



Értelmezés

- **Component Matrix**

- **VIGYÁZAT:** ez még az elforgatás előtti állapot – mi nem ezt akarjuk értelmezni
- Csak azokat a változókat látjuk, amelyek .4-nél nagyobb töltéssel töltene az adott komponensre (ezt mi kértük)
- Látszik, hogy a legtöbb változó az első komponensre tölt itt még

Component Matrix^a

	Component			
	1	2	3	4
SPSS always crashes when I try to use it	,701			
All computers hate me	,685			
I weep openly at the mention of central tendency	,679			
I worry that I will cause irreparable damage because of my incompetence with computers	,673			
People try to tell you that SPSS makes statistics easier to understand but it doesn't	,669			
I wake up under my duvet thinking that I am trapped under a normal distribution	,658			
Computers have minds of their own and deliberately go wrong whenever I use them	,656			
I did badly at mathematics at school	,652			-,400
I slip into a coma whenever I see an equation	,643			
I dream that Pearson is attacking me with	,634			

- **Reprodukált korrelációs mátrix**

- **Reproduced correlation:**

- A modell alapján a változók között milyen korreláció várható
- A modell jó, ha a modell alapján bejósolt korreláció hasonlít az eredetire, tehát különbségük 0 körül van

- **Residuals:**

- A különbségeket tartalmazza a táblázat alsó fele – a residuals megmutatja, mennyire tér el a valóság a bejósolt korrelációs értéktől
- Ha a legtöbb érték 0.05 alatt van, akkor jó a modell
- (csak az egyik háromszöget nézzük!)

	Statiscs makes me cry	My friends will think I'm stupid for not being able to cope with SPSS	Standard deviations excite me	I dream that Pearson is attacking me with correlation coefficients
Reproduced Correlation	,435 ^a	-,112	-,372	,447
Statiscs makes me cry		-,112	,414 ^a	,380
My friends will think I'm stupid for not being able to cope with SPSS			,380	-,399
Standard deviations excite me				-,399
I dream that Pearson is attacking me with correlation coefficients				

Residual ^b	Statiscs makes me cry	My friends will think I'm stupid for not being able to cope with SPSS	Standard deviations excite me	I dream that Pearson is attacking me with correlation coefficients
Statiscs makes me cry		,013	,035	-,011
My friends will think I'm stupid for not being able to cope with SPSS	,013		-,062	,022
Standard deviations excite me	,035	-,062		,019
I dream that Pearson is attacking me with correlation coefficients	-,011	,022	,019	

legalul hasznos infó

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 91 (35,0%) nonredundant residuals with absolute values greater than 0.05.

Értelmezés

- **Rotated component matrix**

- Elforgatott komponens mátrix
- A változóink mennyire töltene a kiválasztott 4 komponensre
- Azt kértük, hogy minden komponenshez csak azoknak a változóknak mutassa a töltését, melyek töltése .4 fölött van
- Láthatjuk, hogy a különböző komponensre különböző „témájú” itemek töltene első komponensre – számítógéppel kapcsolatos félelmek, másodikra statisztikával kapcsolatos félelmek és így tovább

- **Értelmezés**

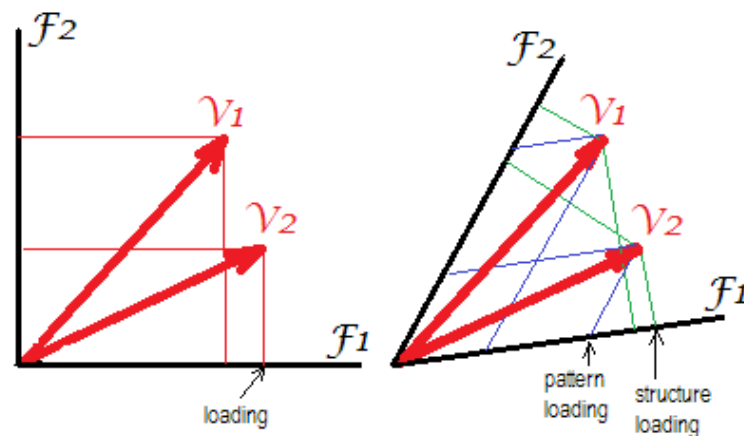
- A: a SAQ nem jó – SPSS szorongás helyett más kapcsolódó fogalmakat mér
- B: a négy komponens alskálája az SPSS szorongásnak

Rotated Component Matrix^a

	Component			
	1	2	3	4
I have little experience of computers	,800			
SPSS always crashes when I try to use it	,684			
I worry that I will cause irreparable damage because of my incompetence with computers	,647			
All computers hate me	,638			
Computers have minds of their own and deliberately go wrong whenever I use them	,579			
Computers are useful only for playing games	,550			
Computers are out to get me	,459			
I can't sleep for thoughts of eigen vectors		,677		
I wake up under my duvet thinking that I am trapped under a normal distribution		,661		
Standard deviations excite me		-,567		
People try to tell you that				

Értelmezés

- Ha **ferde forgatást** csinálsz, a Rotated component matrix helyett két mátrixot kapsz:
 - **Pattern Matrix:** a tengelyre való töltést regressziós együtthatókként kiszámítva (általában ezt értelmezzük inkább – mert csak az egyedi varianciákat tartalmazza)
 - **Structure Matrix:** a tengelyre való töltést korrelációs együtthatókkal kiszámítva
- Az eltérés oka:
 - Orthogonális forgatás esetén (ahol a tengelyek merőlegesek egymásra) e kettő egybe esik egymással és egyszerűen faktortöltésről beszélünk (ez a Rotated component matrix táblázat)
 - Ferde forgatás esetén a tengelyek között megengedett a korreláció, így nem lesznek merőlegesek, és a két számítás eltérő eredményt hoz.
 - Minél erősebb a korreláció:
 - annál eltérőbb a két táblázat (annál nehezebb értelmezni)
 - annál kevésbé értelmezhetőek a tengelyek önálló tengelyként



Értelmezés

- **Component Transformation Matrix**

- Mennyire lettek a komponensek elforgatva, de értelmezés nehézkes

- **Component Plot in Rotated Space**

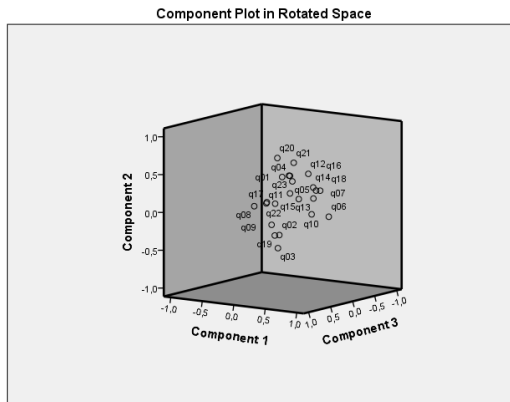
- Az itemek hogyan helyezkednek el a három legerősebb komponens által kifeszített térben –nehezen látható, mit is ábrázol
- Ha meg akarjuk nézni több nézőpontból, a grafikonra kattintva a szerkesztőben beállíthatjuk, melyik két komponens szeretnénk 2D-ben látni

Component Transformation Matrix

Component	1	2	3	4
1	,635	,585	,443	-,242
2	,137	-,167	,488	,846
3	,758	-,513	-,403	,008
4	,067	,605	-,635	,476

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.



Properties

Chart Size Fill & Border Variables

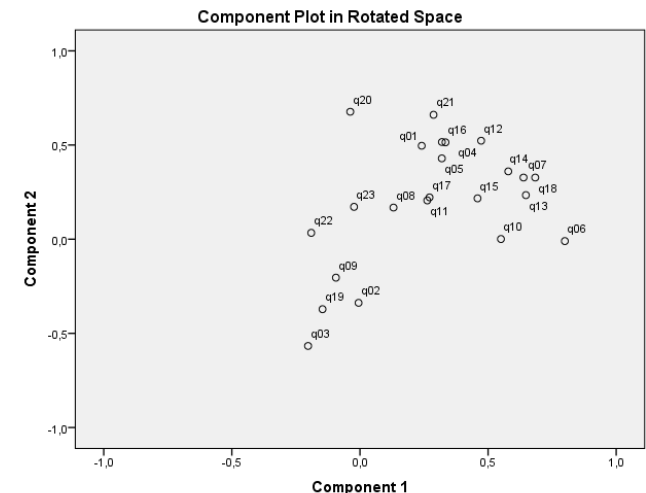
Element Type: ☐ Marker

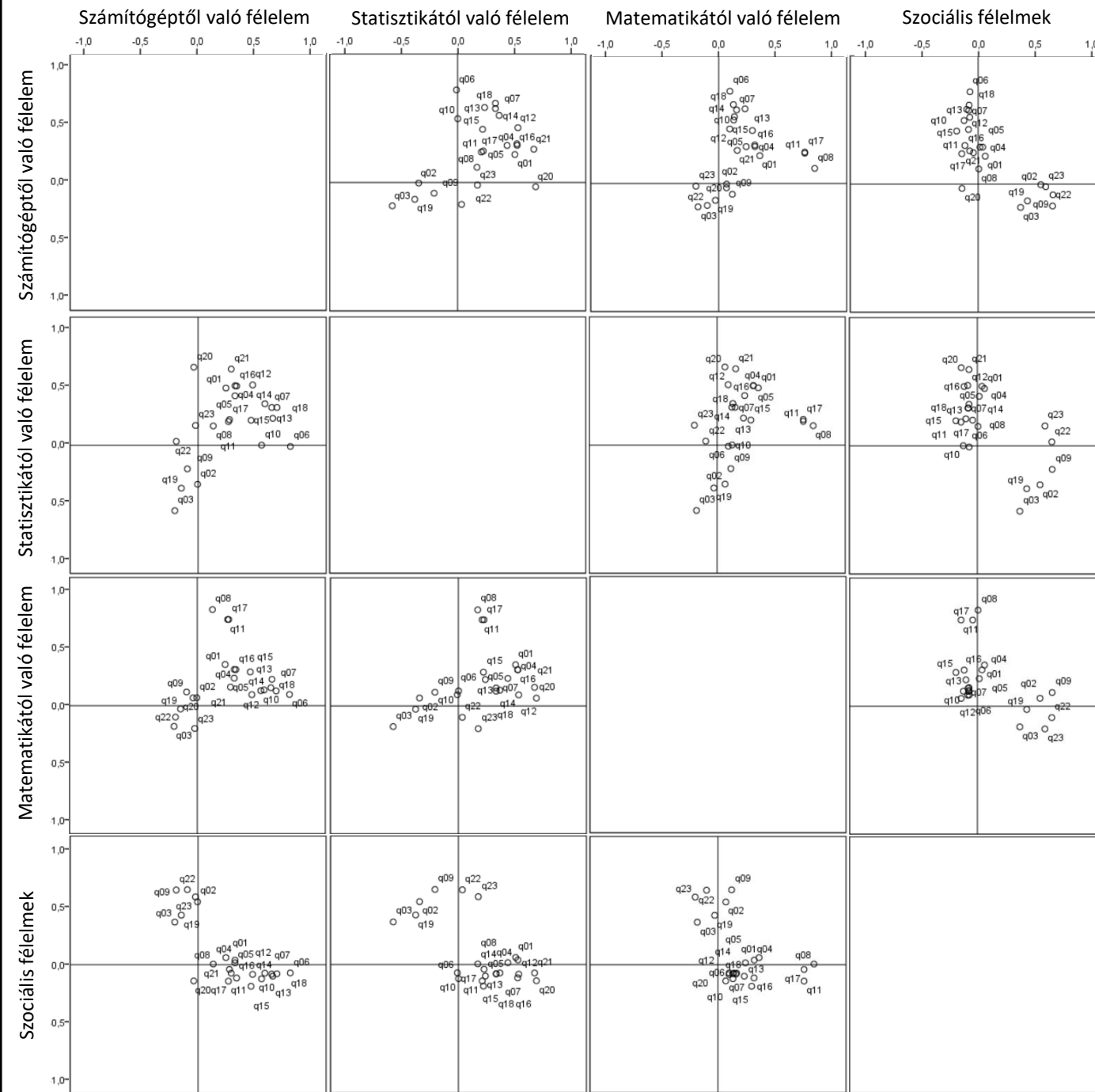
Component 1	X Axis
Component 2	Y Axis
Component 3	
Component 4	
Frequency (\$frequency)	
Percent (\$percent)	

Element (1 of 1)

Previous Delete New Next

Apply Close Help





Elegáns megjelenítési módja az előző dián látható 3D loading plotnak

Az a jó item, mely egy tengelyre erősen tölt (a tengelyen erősen pozitív vagy negatív értéket vesz fel) a többi tengelyre viszont gyengén tölt (a többi tengelyen 0 körüli értéket vesz fel)

Végül mindezt miért csináltuk?

- **Ha faktoranalízisként dolgozom fel:**

- Az adott jelenség a következő négy nagy struktúrára bomlik
- Pl. Big5 – leegyszerűsítve: vették az összes melléknevet (ezek az itemeink) – leskáláztatták, az emberek mennyire érzik magukra/másokra jellemzőnek. Az adatokon faktorelemzést végeztek. Azt kapták, hogy az összes melléknév 5 nagy faktorral magyarázható

- **Ha főkomponens elemzésként értelmezem:**

- Keletkezett 4 új változóm, mely magában foglalja annak az információnak a nagy részét, melyet 23 itemmel mértem. A további elemzésekbe már csak ezt a négy faktort kell betennem.

q22	q23	FAC1_1	FAC2_1	FAC3_1	FAC4_1	v
2	5	,10587	-,92816	-1,82760	-,45953	
4	2	-,58279	-,18932	-,04137	,29332	
2	2	-,54761	,02953	,19918	-,97142	
4	3	,74661	,72268	-,68654	-,18547	
4	4	,25169	-,51489	-,63354	,68301	
1	4	1,91614	-,27356	-,68201	-,52709	
4	4	-,26051	-1,40947	-,00431	,91008	
4	4	-,28572	-,91761	-,08947	1,03746	
3	3	1,71750	1,15102	3,15659	,81082	
4	4	-,69151	-,73318	,18376	1,49541	
5	5	-,91102	-,43355	-,67959	2,30844	
4	4	,00265	,49479	-1,00502	,63617	
3	4	,60378	,87496	,61359	,11381	
3	1	,30849	,89540	-,68141	-1,40898	
4	4	,19016	,66551	-,58721	,74920	
3	4	-,08220	-,20762	-,54816	-,46101	
4	4	-1,42215	-1,73068	-,63872	1,02073	
3	4	,96681	-,69147	,13719	,85616	
4	4	-,69702	-,75636	-,93962	1,88110	
5	5	1,40505	2,63449	-2,12809	1,59810	
2	5	-1,55143	,87996	-2,04499	-,29319	
2	4	1,71846	,66685	,18026	,51247	

Megjegyzések

- Nem baj, ha egy item több tengelyre is tölt. Ez főleg ferde forgatásnál gyakori. Csak akkor szokás az ilyen itemeket kihagyni, ha kérdőívet készítesz.
- A skálákat nem neked kell az itemek összeadásával létrehozni, hanem a Save-ben elmentett változó már maga a skála.
- A faktorelemzés nem érzékeny a fordított itemekre. Az egyenes itemek egyik irányba (pl. pozitívan), a fordítottak a másik irányba töltenek majd a tengelyre.
- A faktorelemzéssel létrehozott skálák reliabilitását mindig ellenőrizni kell. A reliabilitásvizsgálat viszont már érzékeny a fordított tételekre.
- A faktorelemzésben a tengely iránya véletlenszerűen alakul ki, így ha nincs is fordított tételek, akkor is előfordulhat, hogy minden itemed negatívan tölt egy tengelyre. A negatív és pozitív előjelek egy tengelyen bármikor felcserélhetőek.
- Ferde forgatásnál ellenőrizd a tengelyek korrelációs mutatóit!

Megjegyzések

- Azok a faktorok, melyekbe 3-nál kevesebb item tölt, nem vehetők komolyan, mert gyakran nem stabilak. Ezek a faktor elemzésben benne maradhatnak, de további elemzést nem végzünk a skálával (Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7).)
- Ha az volt a célod a főkomponens elemzéssel, hogy két vagy több változót, melyek között kollinearitás vagy multikollinearitás van „egybegyűrj”, akkor Kaiser kritérium helyett egy komponenses modellt kell használni, viszont nem kell törődni a kollinearitás és a minimum itemszám feltételével.
- Egyszerre tanultunk faktor- és főkomponens elemzésről, de ezek különböző eljárások. A faktor elemzés végeredményei faktorok, melyeknek faktortöltése van stb., a főkomponens elemzés végeredményei komponensek, melyeknek komponenstöltése van stb.

- Főkomponens elemzéssel vizsgáltuk, hogy az általunk használt Statisztika Szorongás Kérdőív milyen alskálákra bontható. Az elemzés célja a kérdőív struktúrájának feltárása, az egyes alskálákra jelentéstartalmon töltő itemek azonosítása, és az elemzés eredményeként kimenthető alskálák további elemzésbe való bevonása.
- A kérdőív 23 itemek tartalmaz, melyek ötfokú Likert skálán mérődnek. A vizsgálatban 2571 személy vett részt. A főkomponens elemzést megelőzően ellenőriztük, hogy az itemek egyike sem szélsőségesen ferde, nincs olyan item, ahol a kitöltők több mint 90%-a ugyanazt a választ adta volna.
- A dimenzióredukció módszereként főkomponens elemzést választottunk varimax forgatással, mely biztosítja, hogy a létrejövő tengelyek orthogonálisak, függetlenek egymástól. A kialakult alskálákat a orthogonalitást biztosító Anderson-Rubin módszerrel mentettük ki.
- A Kaiser-Mayer Olkin érték alapján az elemszám a megfelelő, a főkomponens elemzés eredményei értelmezhetőek (KMO=.930).

- A főkomponens elemzés szempontjából problémát jelent, ha az itemek között túl gyengék vagy túl erősek a korrelációk, mert az első esetben nem alakul ki komponens struktúra, az utóbbi esetben pedig minden item egyetlen komponensbe szerveződik.

A Bartlett teszt alapján, mely az itemek teljes függetlenségét teszteli, feltételezhetjük, hogy az itemek között megfelelően erősek a korrelációk $\chi^2(df = 253 \ N = 2571) = 19334.492 \ p < .001$.

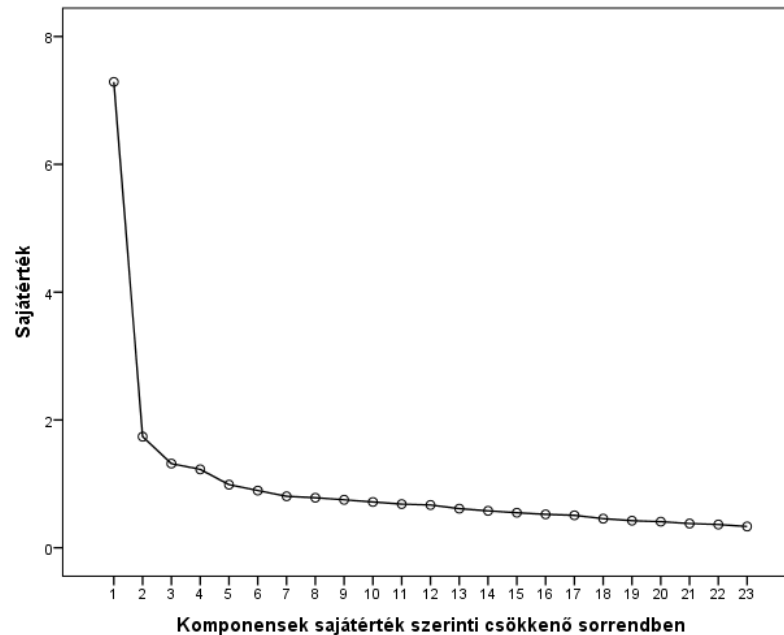
A korrelációs tábla alapján láthatjuk, hogy nincs kollinearitás (a korrelációk erőssége $r = .004$ és $r = .629$ között van), a determináns értéke alapján nincs multikollinearitás ($D = .001$).

Az itemek közötti korrelációk megfelelő erősségűek.

- A létrejövő főkomponensek közül a Kaiser-kritérium alapján azokat tartottuk meg, melyeknek sajátértéke egy fölött van. Az így kapott négy komponens az itemek teljes variáciájának 50.317%-át magyarázza. A komponensek forgatás előtti és azt követő sajátértékeit valamint az általuk megmagyarázott variáciát az alábbi táblázat tartalmazza. A választott kritériumszint helyességét a Scree-plot is alátámasztja.

Komponens	Forgatás előtt		Forgatást követően	
	Sajátérték	Magyarázott variancia (%)	Sajátérték	Magyarázott variancia (%)
1	7.290	31.696	3.730	16.219
2	1.739	7.560	3.340	14.523
3	1.317	5.725	2.553	11.099
4	1.227	5.336	1.949	8.475

Kaiser kritérium alapján megtartott komponensek. A kumulált megmagyarázott variancia 50.317%



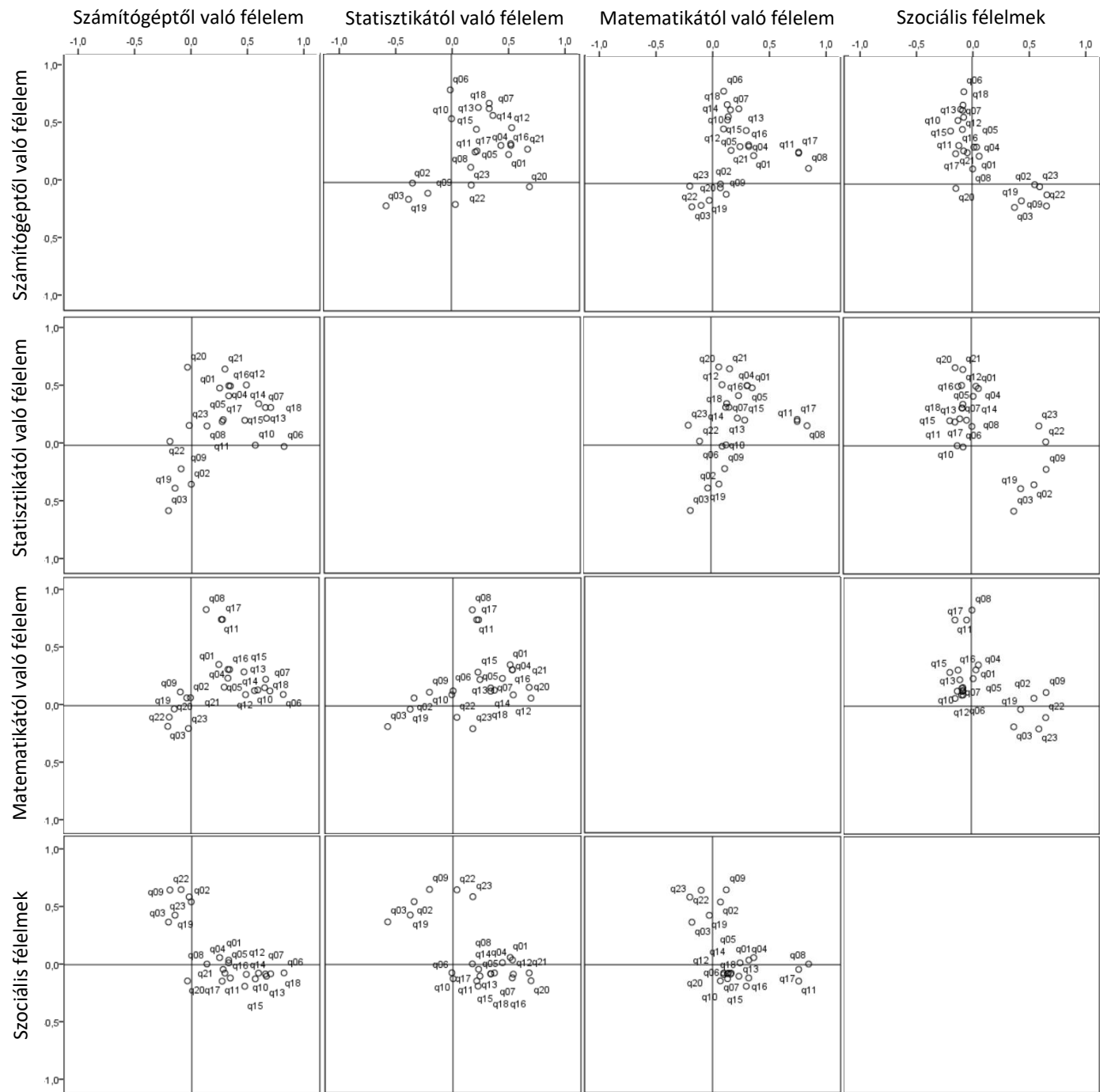
- Az elemzés során fontos, hogy az egynél kisebb sajátértékkel rendelkező komponensek elhagyásából adódó információveszteség ne változtassa meg jelentős mértékben az itemek összefüggésének mintázatát. Az eredeti és reprodukált korrelációs mátrix közötti különbségek .00012 és .177 között vannak (abszolútértékben), és a különbségeknek összesen 35%-a haladja meg a .05-ös szintet. Tehát az eredeti és reprodukált korrelációs mátrix közötti eltérés mértéke az elfogadható tartományon belül van.
- Az orthogonális forgatást követő komponensstruktúrát a következő táblázatban mutatjuk be. Bár a nagy elemszám miatt egy item már $b_s = .0162$ -es értéktől szignifikánsan tölt egy tengelyre $\alpha = .01$ szinten, az értelmezhetőség érdekében egy tengelyhez csak azokat az itemeket jelenítettük meg, melyek az adott tengelyre $b_s = .400$ -es érték fölött töltenek (Stevens, 2002). Ez 16%-os magyarázóerőt jelent.

A 23 elemű kérdőív itemeinek komponenstöltés értékei orthogonális forgatást követően*

Komponens	1	2	3	4
I have little experiment of computers	.800			
SPSS always crashes when i try to use it	.684			
I worry that i will cause irreparable damage because of my incompetence with computers	.647			
All computers hate me	.638			
Computers have minds of their own and deliberately go wrong when i use them	.579			
Computers are useful only for playing games	.550			
Computers are out to get me	.459			
I can't sleep for thoughts of eigen vectors		.677		
I wake up under my duvet thinking that i am trapped under a normal distribution		.661		

* Csak a bs >= .400 fölötti értékeket megjelenítve

- Nem írom ki az összest, el tudjátok képzelni, hogyan folytatódna a táblázat
- Az komponenstöltések grafikusán a következőképpen néznek ki.



- Nyolc item tölt az első komponensre jelentős mértékben. Az itemek mindegyike a számítógéphasználattal kapcsolatos inkompetencia érzéssel kapcsolatos. Olyan itemek tartoznak ide, mint „All computers hate me” vagy „I have little experiment of computers”. A komponens *Félelem a számítógépektől*-nek neveztük el.
- A második komponens nyolc itemet tartalmaz, melyek tartalmilag a statisztikától való félelmet írják le. Példa az itemekre: „I dream that Pearson is attacking me with correlation coefficients” vagy „I don't understand statistics”. A komponens *Félelem a statisztikától*-nak neveztük el.
- A harmadik komponens három itemet tartalmaz, melyek a matematikától való félelem körül forognak. Példa az itemre: „I have never been good at mathematics” és „I slip into a coma whenever I see an equation”. A komponens *Félelem a matematikától*-nak neveztük el.
- A negyedik komponens öt itemet tartalmaz, melyek társas összehasonlítással, szociális félelmekkel kapcsolatosak. Érdekes, hogy az előző skáláktól eltérően ez a tengely nem feltételezi a kitöltő reál tudományokbeli inkompetenciáját, csupán a társaktól való bármely irányú eltérést. Pozitívan töltene a skálára az „If I'm good at statistics my friends will think I'm a nerd” és „My friends will think I'm stupid for not being able to cope with SPSS” itemek egyaránt. A tengelyt *Szociális félelmek*nek neveztük el.

- A létrejött négy komponenshez Anderson-Rubin módszert használva elmentettük a komponensértékeket. A hisztogram alapján látható, hogy mind a négy skála megközelítőleg normál eloszlást követ.

