Többváltozós statisztika, ANOVA és adatredukciós módszerek (BMNPS07700M)

Előadó: Soltész-Várhelyi Klára

Főkomponens elemzés & faktor elemzés

Első rész

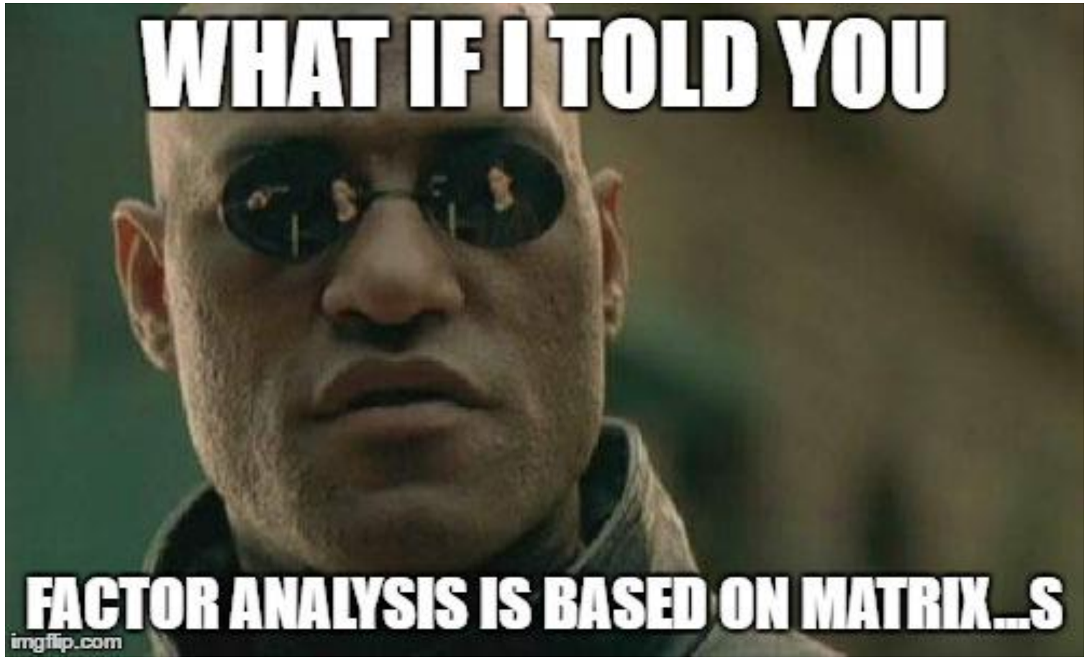
Elmélet

Mire használjuk?

- **Látens változó:**
  - olyan változó, melyet közvetlenül nem tudunk mérni
  - Több általunk mérhető változóra van hatással / fejeződik ki benne
    - Extraverziót nem tudjuk mérni, de azt igen, hogy mennyire szereti a társaságot, mennyire szereti a szereplést, mennyire keresi az izgalmakat…

- **Három használati mód:**
  - 1) megtalálni több tulajdonság mögött meghúzódó látens változót
    - Pl. BigFive dimenziók megtalálása
  - 2) kérdőívkészítés & kérdőívadaptálás
    - Pl. ellenőrizhető, hogy a kialakított kérdőívben tényleg úgy csoportosulnak-e az itemek, ahogy az elmélet alapján gondolnánk vagy adaptálás esetén ahogy az eredeti nyelven csoportosultak
  - 3) adatredukció
    - Sok változó helyett statisztika a kevés faktoron
    - Regresszióanalízisben multikollinearitás problémájára megoldás (bizonyos beállításokkal elérhető, hogy a létrejövő faktorok korrelálatlanok legyenek egymással, így a regresszióba egymástól teljesen független prediktorok kerülhetnek)

Általános elméleti bevezető



Konceptuális alap

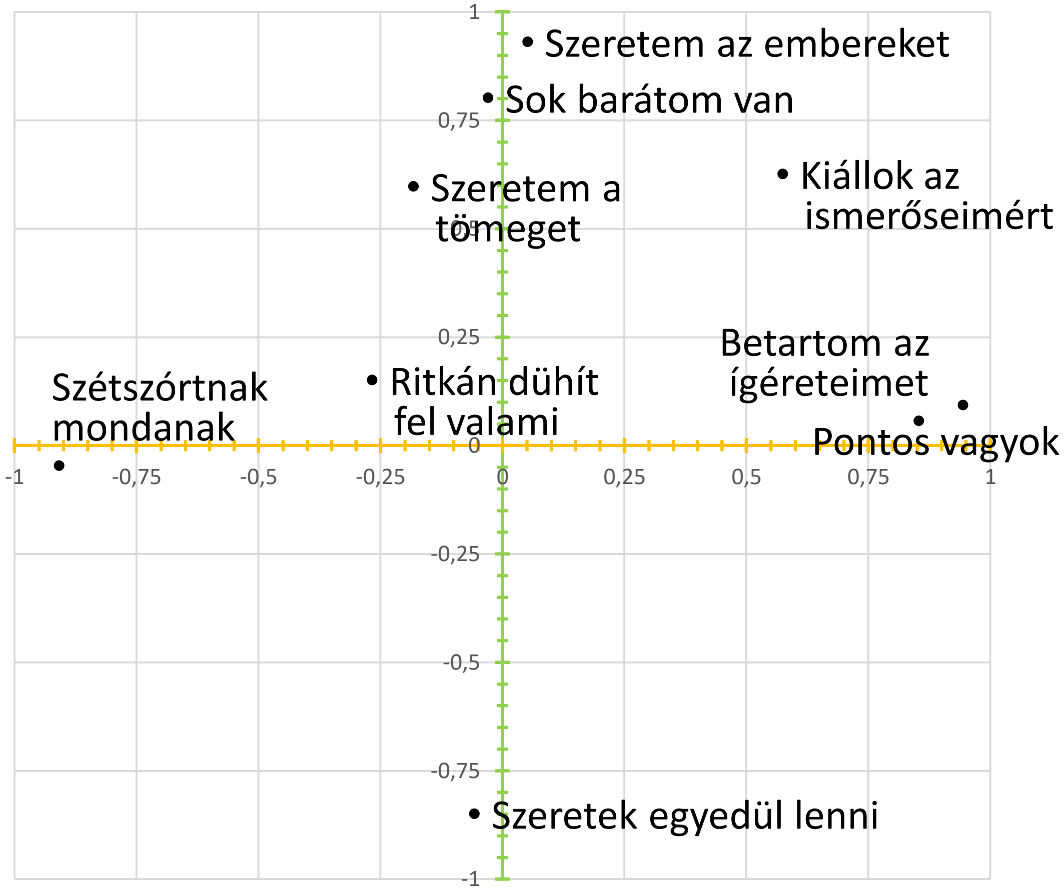
- **Korrelációs mátrix**
  - Korrelációs tábla az itemek között.
  - Azok a változók, melyek között nagy a korreláció, alegységet, csoportot alkotnak.
  - Megkeressük az itemek olyan lehető legkisebb számú csoportosítását, hogy az egy csoportba tartozó itemek közt nagy legyen korreláció, míg a különböző csoportba tartozók között kicsi.

	Szeretem az embereket	Szeretek egyedül lenni	Sok barátom van	Szeretem a tömeget	Kiállok az ismerő-seimért	Pontos vagyok	Szétszórt-nak mondanak	Betartom az ígéreteim	Ritkán dühít fel valami
Szeretem az embereket	1,000								
Szeretek egyedül lenni	-,867	1,000							
Sok barátom van	,663	-,564	1,000						
Szeretem a tömeget	,414	-,332	,361	1,000					
Kiállok az ismerőseimért	,599	-,556	,458	,149	1,000				
Pontos vagyok	,059	-,036	,028	-,123	,528	1,000			
Szétszórtnak mondanak	-,098	,105	-,033	,066	-,535	-,876	1,000		
Betartom az ígéreteim	,102	-,128	,003	-,037	,408	,709	-,632	1,000	
Ritkán dühít fel valami	,035	-,044	,095	,172	-,104	-,175	,130	-,086	1,000

Tengelyek (faktorok, komponensek) és itemek töltései ezeken

- Az előbbi „csoportok” tulajdonképpen a **faktorok** faktorelemzésben illetve **komponensek** főkomponens elemzésben, vagy általános néven a **tengelyek**. (Az általános elméleti részben komponenst fogok írni).
  - Az elemzés eredménye elképzelhető annyi dimenziós térként, ahány komponens van
  - Hogy a komponensek hogyan jönnek létre, az a különböző elemzési módszerekben eltérő, és a koncepció megértéséhez most még nem is kell, így elég róla egyenlőre annyi, hogy ez

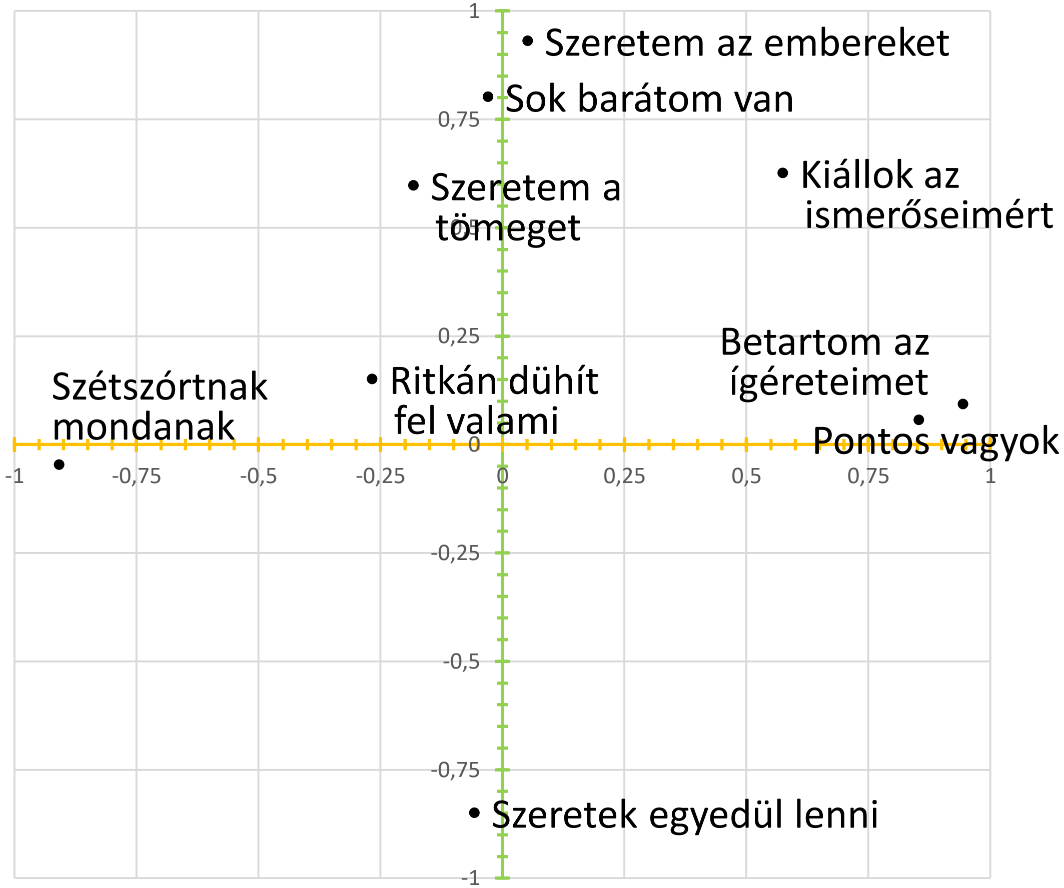
- **Komponenstöltés (ill. faktortöltés, loadings, jelölése: b_s)**
  - Az item és a komponens közötti Pearson korreláció (tehát értéke -1 és +1 között mozog)
  - A tengelyek mentén az itemek elhelyezhetők aszerint, hogy mennyire korrelálnak a komponenssel, így a töltések az itemek koordinátái lesznek
  - Minden item minden komponensre tölt valamilyen mértékben, de jó esetben egy-egy item egy komponensre tölt erősen, a többire pedig csak gyengén.
  - A példában a „Kiállok az ismerőseimért” nem igazán jó item, mert mindkét tengelyre viszonylag erősen tölt, és a „Ritkán dühít fel valami” sem jó item, mert egyik tengelyre sem tölt erősen. Hogy ezekkel mit érdemes kezdeni, később megbeszéljük.



Töltés mátrix

- **Komponenstöltés (ill. faktortöltés) mátrix**
  - A töltések mátrixba rendezésével kapjuk, jelölése b_s
  - Oszlopok jelölik a tengelyeket, sorok jelölik a változókat
- A tengelyek **jelentés**ének meghatározása a töltések alapján
  - A töltések mátrixos és/vagy grafikus megjelenítése segítségével az elemzést végző személy pszichológiai tudása alapján dönt a tengelyek jelentéséről.
  - Mi lehet a két tengely jelentése?
    - **Barátságosság** és **Lelkiismeretesség**

$$
\text{Töltés:}\quad b_s = \begin{array}{cl}
\begin{array}{cc} \text{Tengely1} & \text{Tengely2} \end{array} & \\
\begin{bmatrix}
.931 & .068 \\
-.878 & -.076 \\
.803 & -.021 \\
.574 & -.202 \\
.630 & .595 \\
.010 & .946 \\
-.063 & -.910 \\
.058 & .811 \\
.141 & -.274
\end{bmatrix} &
\begin{array}{l}
\text{Szeretem az embereket} \\
\text{Szeretek egyedül lenni} \\
\text{Sok barátom van} \\
\text{Szeretem a tömeget} \\
\text{Kiállok az ismerőseimért} \\
\text{Pontos vagyok} \\
\text{Szétszórtnak mondanak} \\
\text{Betartom az ígéreteimet} \\
\text{Ritkán dühít fel valami}
\end{array}
\end{array}
$$



Melyik itemek tartozzanak bele egy tengelyhez?

- Amikor értelmezzük a tengelyeket, el kell döntenünk, milyen erősségű töltéstől számítunk egy itemet az adott tengelyhez tartozónak. (Pl. a „Szeretem a tömeget” még értelmezzük-e a Tengely1-hez?)
  - Megbízhatóság szempontjából minta nagyságától függ. $\alpha = .01$ szig. szinten
    - 50 .722
    - 100 .512
    - 200 elemszámú mintán .364 -nél nagyobb töltésű változókat (abszolút értékben)
    - 300 .298
    - 600 .210
    - 1000 .162
    - De így nagy mintán sokszor értelmezhetetlenül sok változó tölt egy tengelyre
  - Magyarázóerő szempontjából a nagyméretű mintán a(z abszolút értékben vett) .400-nál nagyobb töltésű változókat érdemes értelmezni (ez 16% megmagyarázott varianciát jelent) (Stevens, 2012)
  - Tehát, ha egy 100 fős mintán néztük az előző elemzést, akkor
    - A Barátságosság: Szeretem az embereket, Szeretek egyedül lenni, Sok barátom van, Szeretem a tömeget, Kiállok az ismerőseimért
    - Lelkiismeretesség: Pontos vagyok, Szétszórtnak mondanak, Betartom az ígéreteimet, Kiállok az ismerőseimért

Tengely1	Tengely2	
.931	.068	Szeretem az embereket
-.878	-.076	Szeretek egyedül lenni
.803	-.021	Sok barátom van
.574	-.202	Szeretem a tömeget
.630	.595	Kiállok az ismerőseimért
.010	.946	Pontos vagyok
-.063	-.910	Szétszórtnak mondanak
.058	.811	Betartom az ígéreteimet
.141	-.274	Ritkán dühít fel valami

Számolható értékek

- Most már tudjuk, hogy a kérdőívnek van egy Barátságosság és egy Lelkiismeretesség tengelye. Jó lenne kiszámolni az emberek Barátságosság és Lelkiismeretesség pontszámát. Hogyan tehetjük ezt meg?
  - **1.** A tengelyhez tartozó **itemek pontjainak összeadásával** vagy átlagolásával (egyszerű, de kevésbé jó módszer)
    - Barátságosság: Szeretem az embereket, Szeretek egyedül lenni (fordított), Sok barátom van, Szeretem a tömeget, Kiállok az ismerőseimért
    - Lelkiismeretesség: Pontos vagyok, Szétszórtnak mondanak (fordított), Betartom az ígéreteimet, Kiállok az ismerőseimért

Emberek	Egyedül	Barát	Tömeg	Kiállok	Pontos	Szétszórt	Ígéret	Dühít	Barátságosság	Lelkiismeretesség
5	2-> 4	5	5	4	2	3 -> 3	3	1	5+4+5+5+4 = 23	2+3+3+4 = 12
2	3-> 3	2	2	2	2	3 -> 3	4	1	2+3+2+2+2 = 11	2+3+4+2 = 11
4	2-> 4	3	3	3	5	2 -> 4	3	2	4+4+3+3+3 = 17	5+4+3+3 = 15

- A módszer hátrányai:
  - A különböző itemszámú skálák értékei nem összehasonlíthatók egymással (nem mondhatom, hogy a harmadik személy Barátságossága magasabb, mint a Lelkiismeretessége, mert a Barátságosság több itemet tartalmaz)
  - A módszer csak azonos skálán mérődő itemek esetén használható (pl. nem lenne létrehozható az Egészségesség skála a „Hetente hányszor sportol?”, „Milyen gyakran iszik alkoholt?”, „Dohányzik?”, „Mennyire figyel a kiegyensúlyozott étkezésre?” itemekből, mert más a mértékegységük.
  - Nem súlyozza az itemeket, a „Szeretem az embereket” és a „Szeretem a tömeget" egyforma súllyal szerepel.

- **2. Komponens érték** (component score illetve faktor érték – factor score) számolásával
  - Az itemek összeadásánál **súlyozzuk az itemeket** a faktorelemzésből számolt **fontosságuk alapján**
  - Ehhez tudnunk kell az itemek fontosságát, amit a Komponens érték együttható ad majd meg
- **Komponens érték együttható** (component score coefficient) illetve faktorelemzésnél faktor érték együttható.
  - Jelölése B
  - Az itemek töltéséből indul ki, de figyelembe veszi az itemek közötti korrelációt is, így a komponens és az item közötti egyedi kapcsolatot méri. Ez lesz a súly, ami alapján megmondhatjuk, hogy egy-egy item mennyire fontos a Barátságosság és Lelkiismeretesség szempontjából
  - Például Barátságosság szempontjából a „Szeretem az embereket” lesz a legfontosabb, a „Szeretem a tömeget” kevésbé fontos, és a „Pontos vagyok” egyáltalán nem lesz fontos.
  - Egy item együtthatója mindig alacsonyabb, mint a töltése, hiszen nem tartalmazza azt a varianciát, melyet az adott item mellett más itemek is képesek megmagyarázni a tengelyből. (Hasonlóképp, ahogy a többszörös lineáris regresszióban egy prediktor magyarázó ereje alacsonyabb volt, mint amikor korrelációval néztük a kimeneti változóval való kapcsolatát.)

$$
B=\begin{array}{cl}
\begin{array}{cc}\text{Baráts} & \text{Lelkiism}\end{array} & \\
\begin{bmatrix}
.310 & -.025 \\
-.292 & -.019 \\
.272 & -.050 \\
.204 & -.102 \\
.181 & .180 \\
-.046 & .338 \\
.027 & -.322 \\
-.023 & .287 \\
.062 & -.105
\end{bmatrix} &
\begin{array}{l}
\text{Szeretem az embereket} \\
\text{Szeretek egyedül lenni} \\
\text{Sok barátom van} \\
\text{Szeretem a tömeget} \\
\text{Kiállok az ismerőseimért} \\
\text{Pontos vagyok} \\
\text{Szétszórtnak mondanak} \\
\text{Betartom az ígéreteimet} \\
\text{Ritkán dühít fel valami}
\end{array}
\end{array}
$$

- A komponensek pontszámának számolása előtt még egy lépés van vissza: **az itemeket standardizáljuk**, ennek köszönhetően akár különböző mértékegységű itemeket is be lehet vonni az elemzésbe.

Emberek	Egyedül	Barát	Tömeg	Kiállok	Pontos	Szétszórt	Ígéret	Dühít
5 -> 1,397	2 -> -0,928	5 -> 1,631	5 -> 1,681	4 -> 1,097	2 -> -0,819	3 -> 0,196	3 -> -0,368	1 -> -1,370
2 -> -0,809	3 -> 0,039	2 -> -1,213	2 -> -0,654	2 -> -1,188	2 -> -0,819	3 -> 0,196	4 -> 0,600	1 -> -1,370
4 -> 0,662	2 -> -0,928	3 -> -0,265	3 -> 0,125	3 -> -0,046	5 -> 1,281	2 -> -0,694	3 -> -0,368	2 -> -0,660

- Megjegyzések a standardizálással kapcsolatban:
  - Standardizálással minden itemet olyanná alakítunk, ahol az itemen a minta átlaga 0, szórása pedig 1. Például az Emberek itemen az első személy standardizált értéke pozitív, tehát ő az átlagosnál jobban szereti az embereket, a második személy standardizált értéke negatív, tehát ő az átlagosnál kevésbé szereti az embereket.
  - Főkomponens elemzést nem csak korrelációs táblákat használva, hanem kovariancia táblákat használva is lehet végezni. Kovarianciák elemzése során az itemeket nem standardizáljuk, csak centralizáljuk.
  - Itt nem az egész adatbázist látod, csak annak első három sorát.

- Most már ismerjük a súlyokat, a komponens szempontjából fontos itemek nagy súllyal, a nem fontos itemek kis súllyal fognak a pontszámba beszámítani, valamint egységesítettük az itemeket is, nincs más hátra mint kiszámolni a komponens értékeket!

Emberek	Egyedül	Barát	Tömeg	Kiállok	Pontos	Szétszórt	Ígéret	Dühít
1,397	-0,928	1,631	1,681	1,097	-0,819	0,196	-0,368	-1,370
-0,809	0,039	-1,213	-0,654	-1,188	-0,819	0,196	0,600	-1,370
0,662	-0,928	-0,265	0,125	-0,046	1,281	-0,694	-0,368	-0,660

B=

Baráts	Lelkiism	
.310	-.025	Szeretem az embereket
-.292	-.019	Szeretek egyedül lenni
.272	-.050	Sok barátom van
.204	-.102	Szeretem a tömeget
.181	.180	Kiállok az ismerőseimért
-.046	.338	Pontos vagyok
.027	-.322	Szétszórtnak mondanak
-.023	.287	Betartom az ígéreteimet
.062	-.105	Ritkán dühít fel valami

- Egy komponens értékét úgy kapjuk meg, hogy az itemek standardizált értékét az item adott komponensen lévő súlyával megszorozzuk, és ezeket összeadjuk.

A Barátságosság szempontjából fontos itemek nagy súllyal szerepelnek

Minél kevésbé fontos egy item, annál kisebb a súlya

Baráts. = **.310***Emberek **- .292***Egyedül + **.272***Barát + **.204***Tömeg + **.181***Kiállok **- .046***Pontos + **.027***Szétszórt **- .023***Ígéret + **.062***Dühít

A Barátságosság szempontjából negatívan megfogalmazott itemek negatív súllyal szerepelnek, így nem kell őket külön megforgatni

A Barátságosság szempontjából lényegtelen itemek olyan kis súllyal szerepelnek, hogy szinte mindegy is, hogy ott vannak

Lelkiis. = **- .025***Emberek **- .019***Egyedül **- .050***Barát **- .102***Tömeg + **.180***Kiállok + **.338***Pontos **- .322***Szétszórt + **.287***Ígéret **- .105***Dühít

A képletbe behelyettesítve az első személy barátságosság és lelkiismeretesség értéke:

Baráts. = **.310***1.397 **- .292***(-0.928) + **.272***1.631 + **.204***1.681 + **.181***1.097 **- .046***(-0,819) + **.027***0.196 **- .023***(-0.368) + **.062***(-1.370) = **1.656**

Lelkiis. = **- .025***1.397 **- .019***(-0.928) **- .050***1.631 **- .102***1.681 + **.180***1.097 + **.338***(-0.819) **- .322***0.196 + **.287***(-0.368) **- .105***(-1.370) = **-0.410**

- Megjegyzések a komponens értékhez
  - Kérdőívekben a komponenseket szokás skáláknak vagy alskáláknak is nevezni, tehát mondhatjuk, hogy létrehoztuk a Barátságosság és Lelkiismeretesség alskálákat, és kiszámoltuk az emberek pontszámát ezeken.
  - A komponens érték használatakor NEM kell az itemeket megforgatni, a negatív itemek (pl. „Szeretek egyedül lenni") egyszerűen negatív súllyal számolódnak
  - A létrejövő skálaérték standardizált formában van, tehát átlaga 0, szórása 1 a skáláknak
  - A pozitív skálaérték azt jelenti, hogy a személy az átlagnál magasabb értéket vesz fel a skálán (pl. az első személy az átlagnál barátságosabb, de kevésbé lelkiismeretes)
  - A személyek összehasonlíthatók egymással (pl. az első személy barátságosabb, mint a második)
  - Egy személy különböző skálapontszámai összehasonlítható egymással (pl. a harmadik személy barátságossága magasabb, mint lelkiismeretessége)
  - Az előzőekben bemutatott rengeteg számolást nem kell neked elvégezni, azt a skálapontszámok mentésével kérheted majd a főkomponens elemzés során.

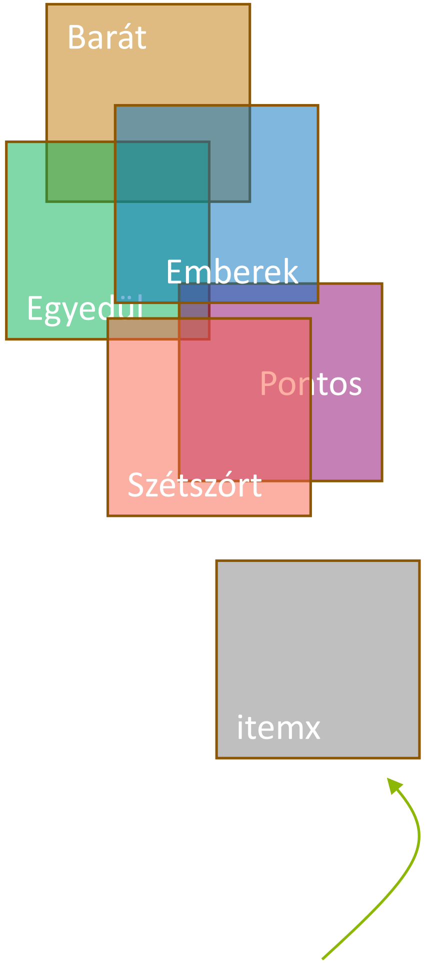
Emberek	Egyedül	Barát	Tömeg	Kiállok	Pontos	Szétszórt	Ígéret	Dühít	Barátságosság	Lelkiismeretesség
5	2	5	5	4	2	3	3	1	1.656	-0.410
2	3	2	2	2	2	3	4	1	-0.996	-0.088
4	2	3	3	3	5	2	3	2	0.311	0.578

Módszerek

- **Megerősítő (konfirmátoros) faktoranalízis (CFA – confirmatory factor analysis)**
  - látens változók struktúrájáról a kutató fejében már létezik egy modell, és ezt a meglévő modellt ellenőrzi
  - bonyolult összefüggésrendszerek modellezésére alkalmas
  - AMOS vagy Mplus
- **Feltáró (exploratív) faktoranalízis (EFA – exploratory factor analysis)**
  - adatok mögött rejtőző struktúra feltárása és a megtalált faktorok általánosítása a populációra
  - Másik mintán ellenőrizni kell
  - Többfajta matematikai megoldás (Főfaktor analízis, Maximum likelihood elemzés, stb.)
- **Főkomponens elemzés (PCA - principal component analysis)**
  - Legegyszerűbb módszer
  - Az adatok átrendezése úgy, hogy az itemek varianciáját (az általuk kifeszített teret) a lehető legkevesebb új dimenzióval írjuk le a lehető legkevesebb veszteséggel
  - „csak” merőleges transzformációk sorozata, mely az adatokat egy új koordináta rendszerbe transzformálja, de ha mögöttes összefüggésrendszer feltárására nem alkalmas
  - gyakorlatban a kérdőívek struktúrájának feltárásához általában megelégszünk vele, de egy új személyiségmodellt ne főkomponens elemzés alapján alkoss!

Kommunalitás

- A főkomponens elemzés és faktorelemzés összehasonlítása előtt tudnunk kell még:
- Több item összefüggésének vizsgálatakor egy item **teljes varianciája** két részből áll:
  - **Közös variancia** – más változókkal közös
  - **Egyedi variancia** – csak az adott változóhoz tartozik (az egyedi variancia része a random variancia – hiba is, de ezzel nem tudunk számolni)
- **Kommunalitás**:
  - megadja, hogy egy item varianciájának mekkora része közös
  - 1, ha az egész; 0-ha semennyi (a többi itemtől teljesen független az adott item)
- Faktoranalízis az itemek közös varianciájával dolgozik
  - Minden változóra többszörös lineáris regresszió– az R^2-ekből becsüljük - faktoranalízis
- Főkomponens elemzés az itemek teljes varianciájával dolgozik
  - Kezdetben a kommunalitást 100%-nak tekinti



Az ő kommunalitása 0 lenne, ez azt is jelenti, hogy egyik itemmel sem korrelál

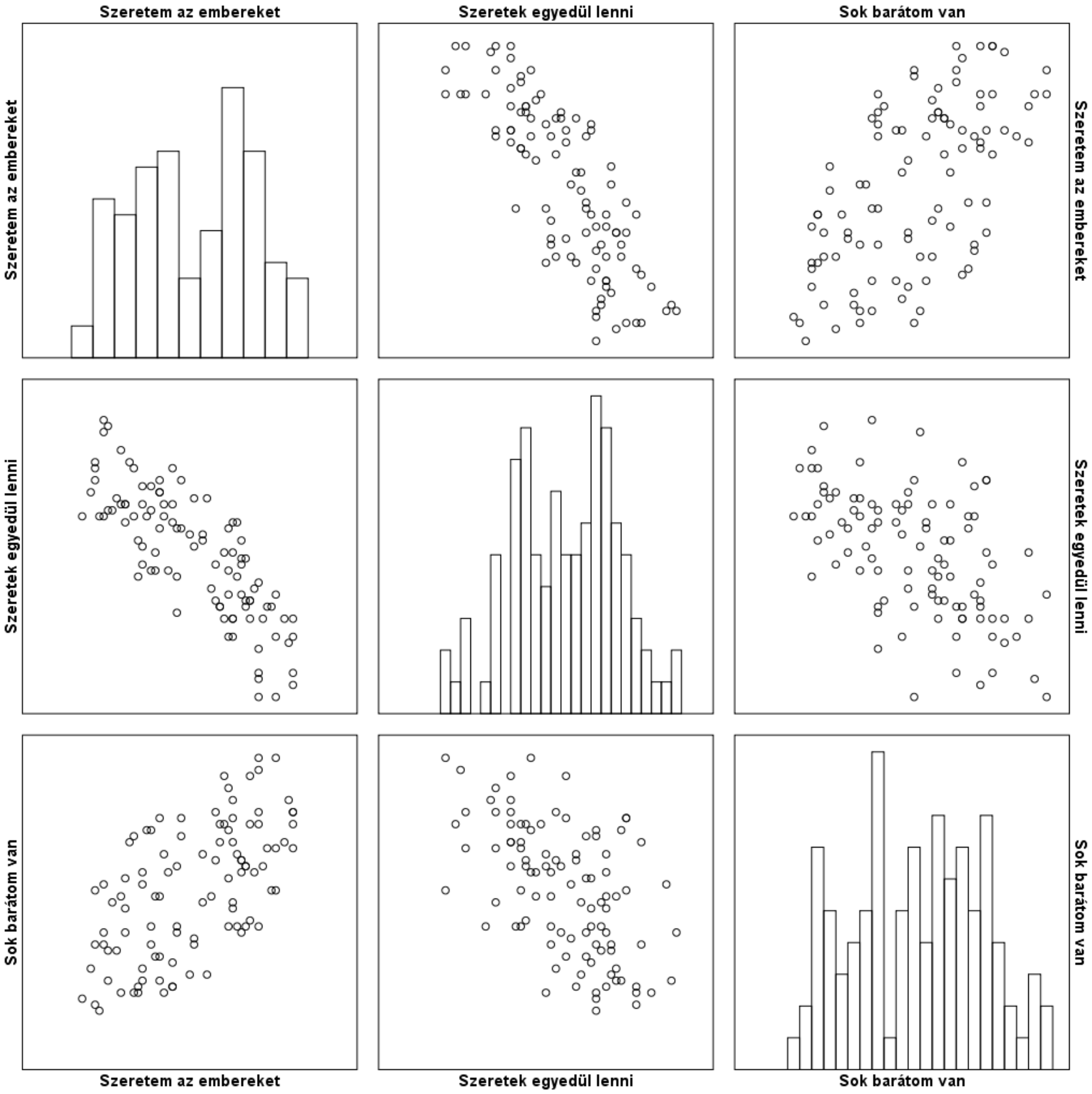
Faktor analízis / főkomponens analízis

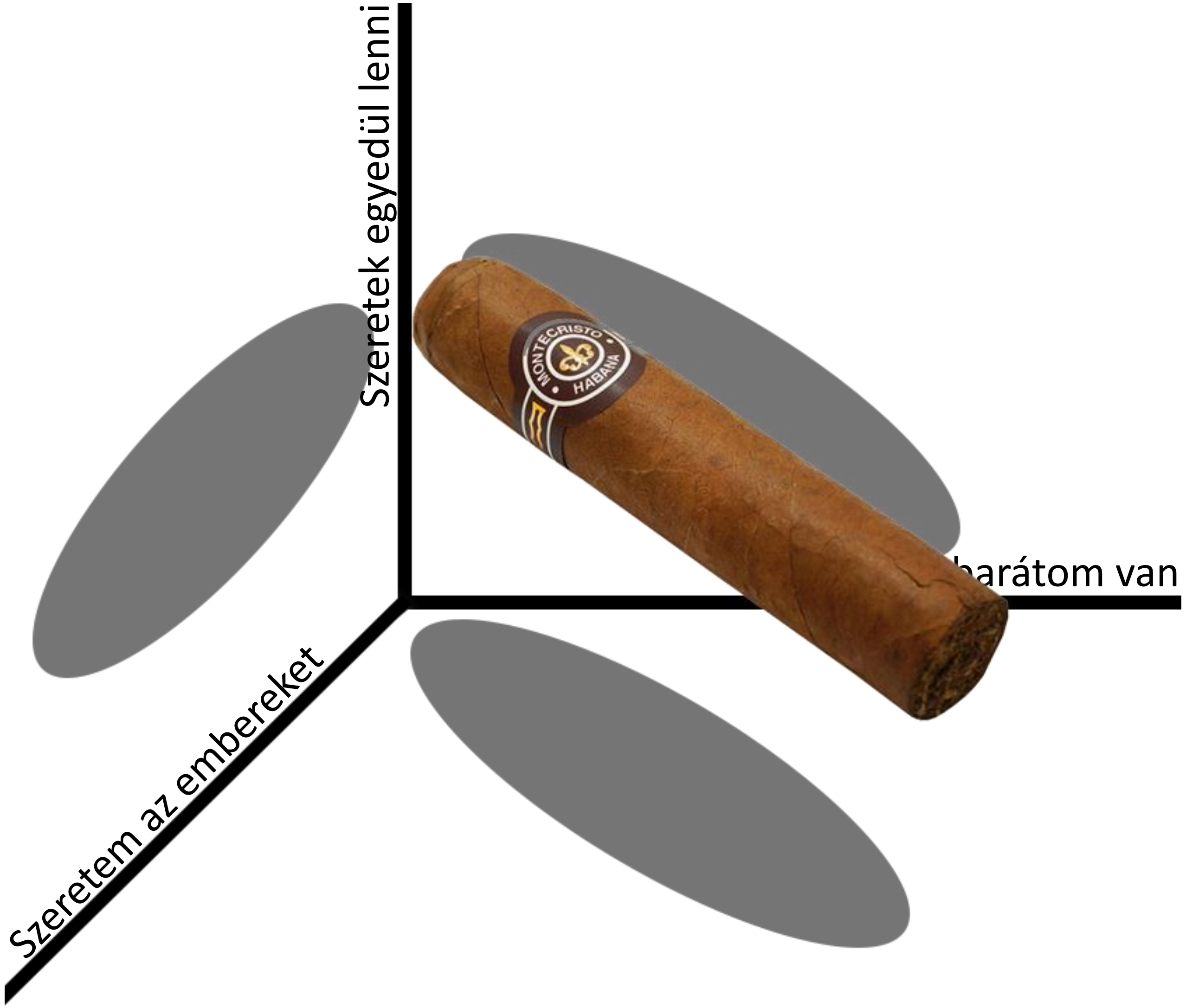
Faktor analízis	Főkomponens elemzés
Itemek közös varianciáját modellezi	Itemek teljes varianciáját elemzi
Kommunalitást lineáris regresszióval számolja	Kommunalitást 100%-nak tekinti
Faktorok matematikai modell alapján	Az eredeti adatot lineáris változók összességére bontja
Alkalmas látens változók vizsgálatára	Az adatokban meglévő lineáris komponenseket találja meg
Rengeteg feltétel (pl. hozza a többszörös regresszió minden feltételét)	egyszerű
30 változó felett magas kommunalitással nagyon hasonló eredményt hoznak	
Kevés változó, alacsony kommunalitás esetén nagy különbségek lehetnek, a főkomponens elemzés eredménye nem biztos, hogy értelmezhető	

A komponensek megtalálása

(azaz hogyan történik a varázslat)

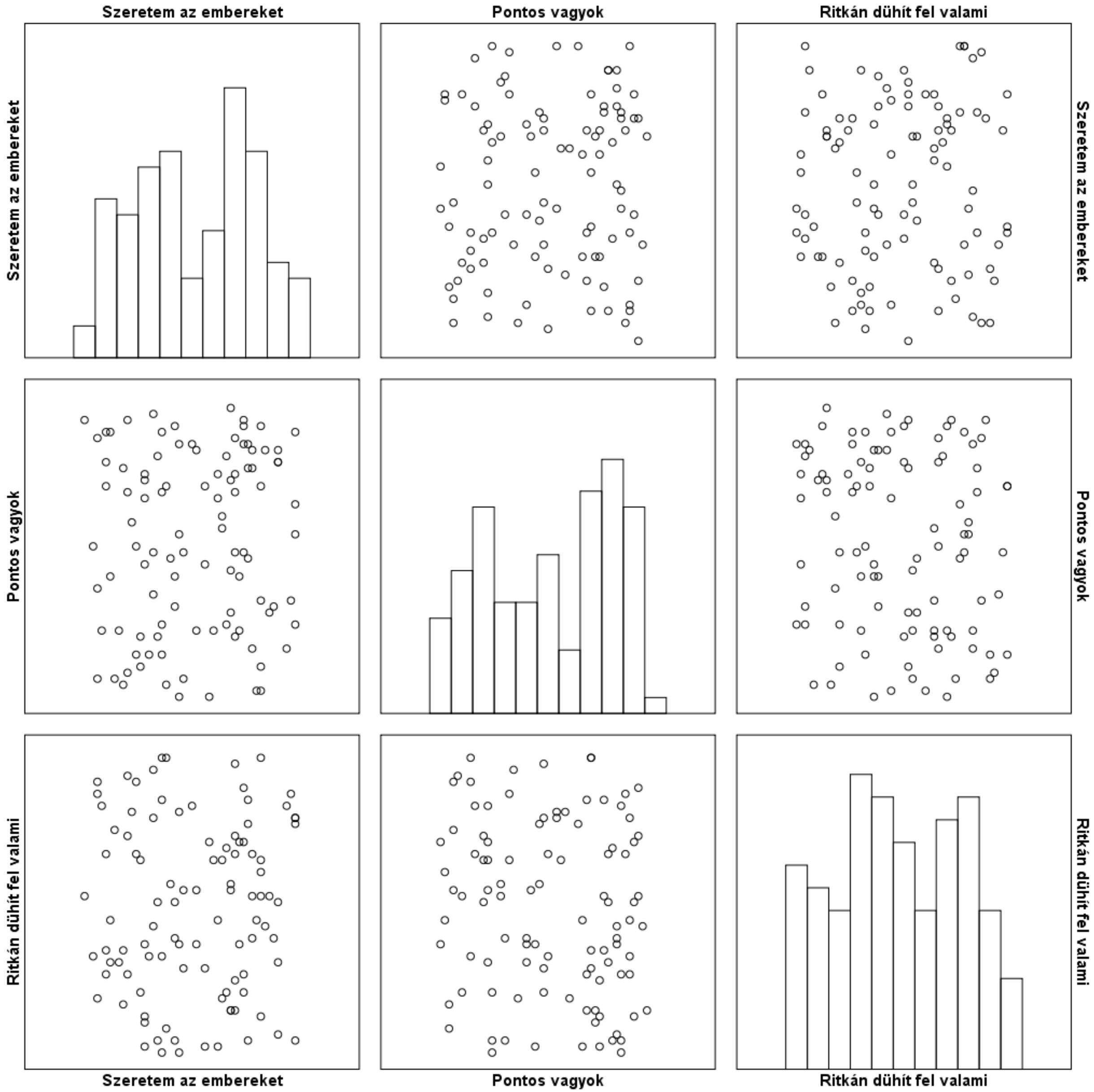
Hogy néz ki egy többdimenziós térben, ha az itemek korrelálnak?

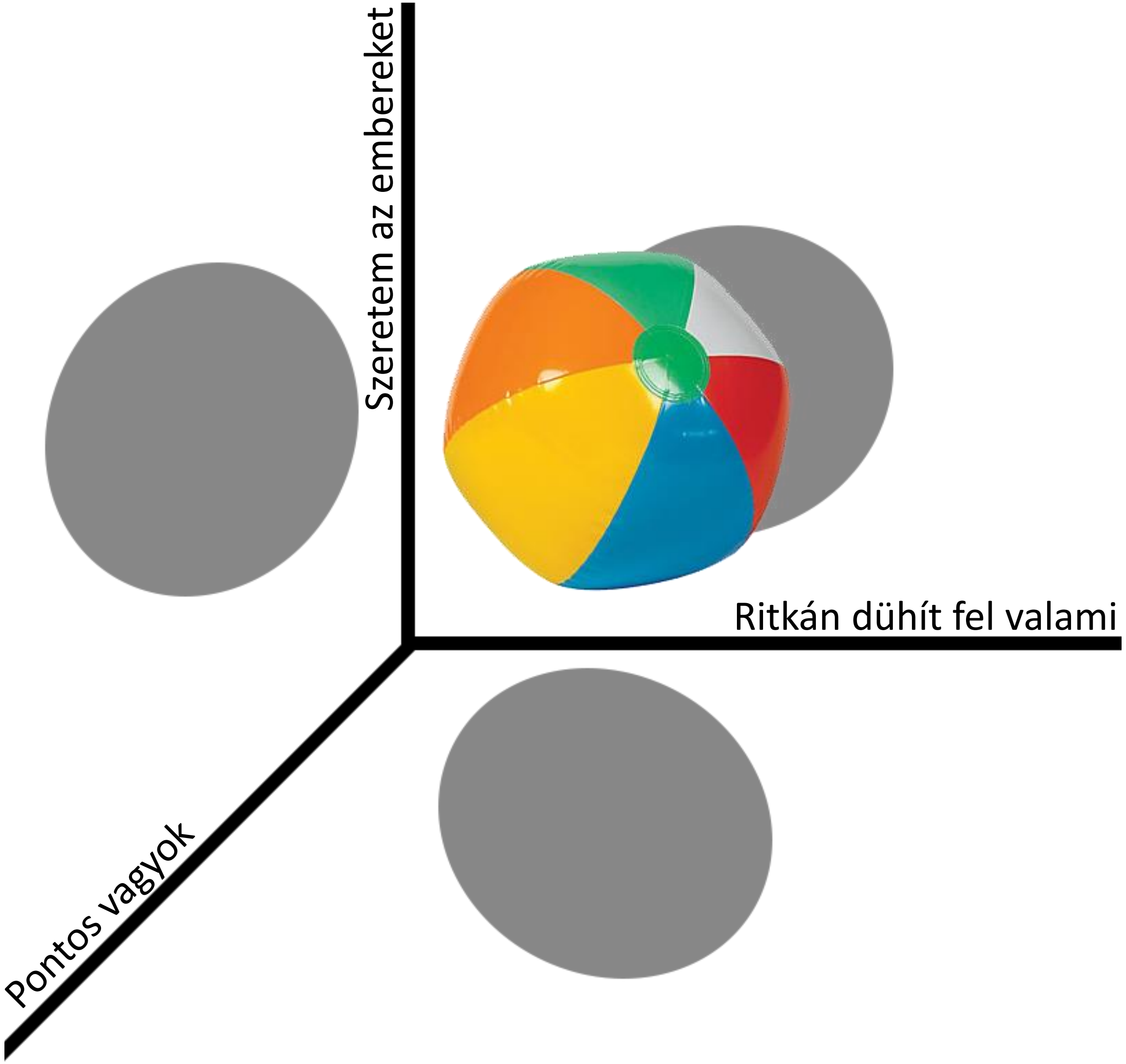




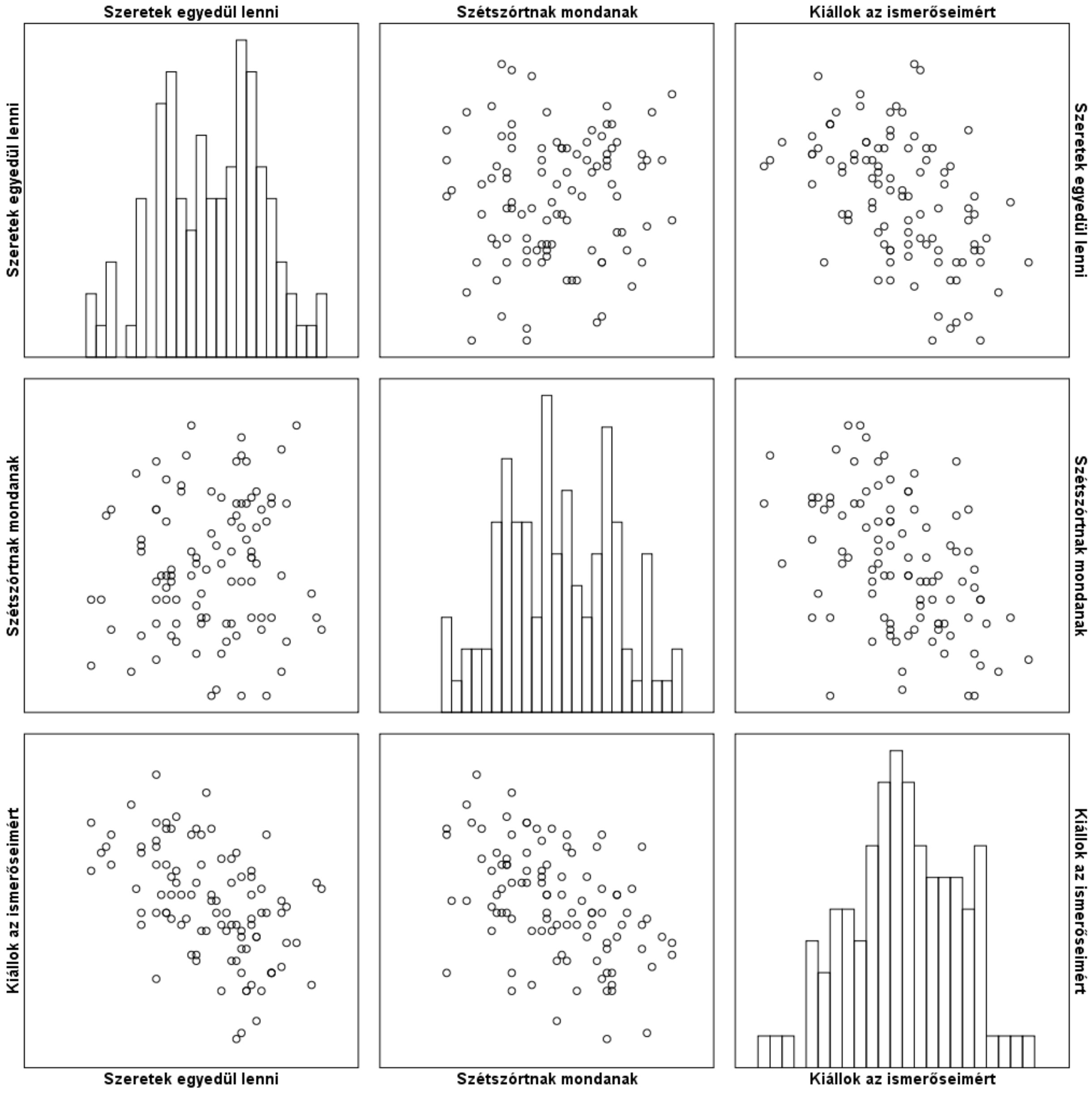
(a könnyebb átláthatóság kedvéért a Likert-skálás itemekhez zajt adtam)

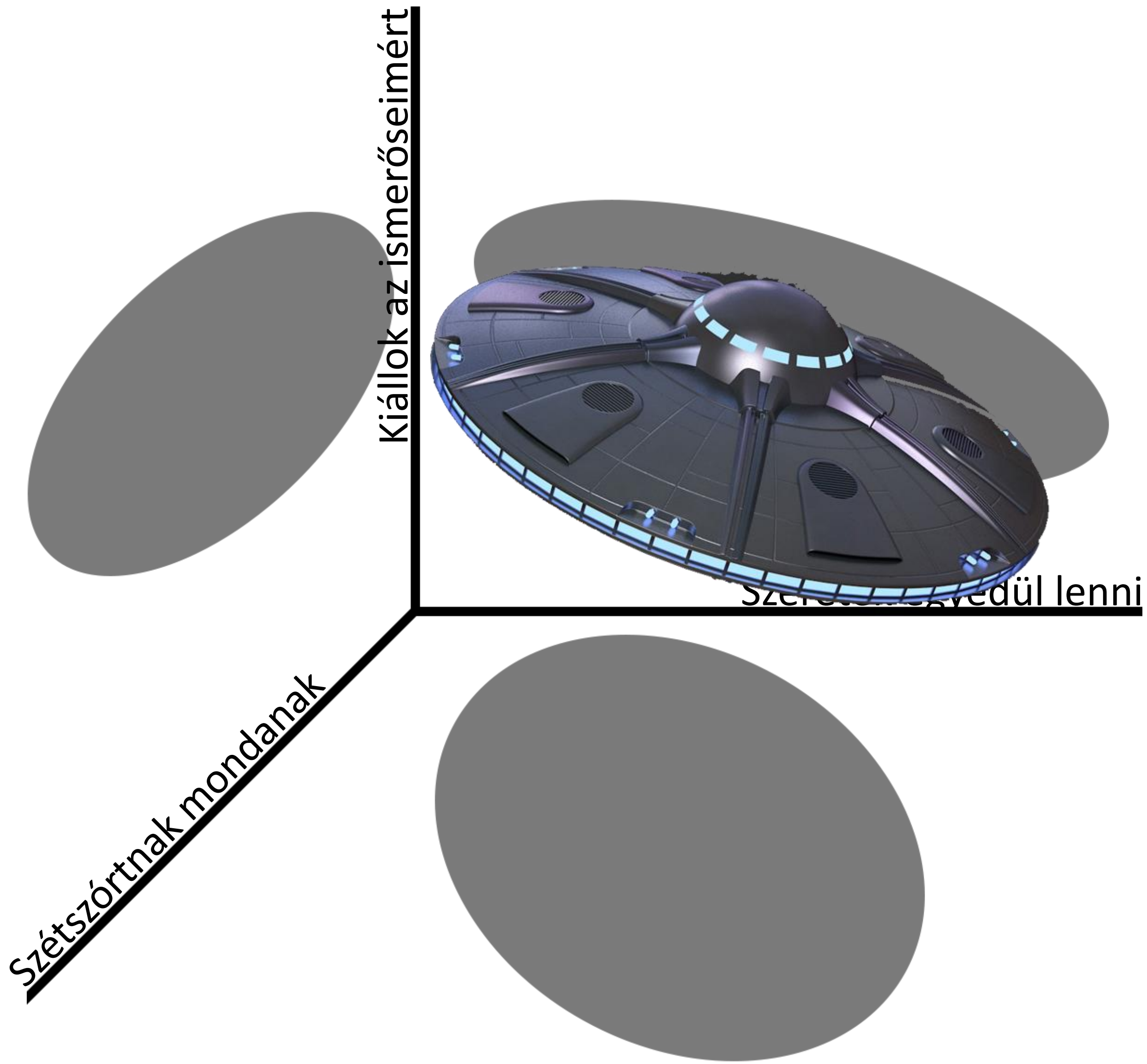
Hogy néznek ki, ha nem korrelálnak?





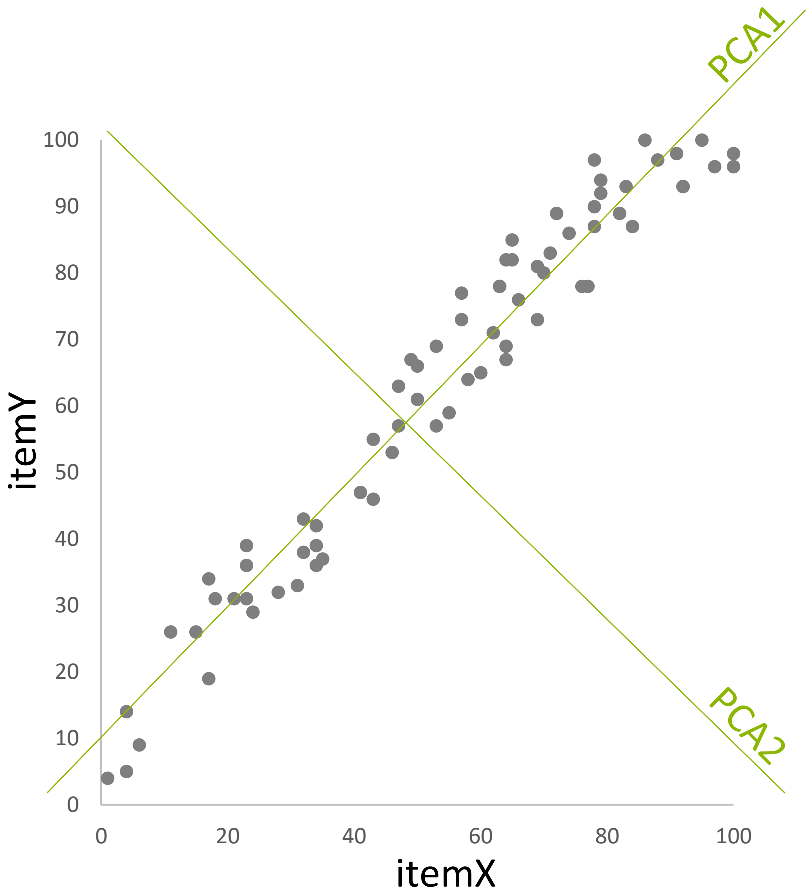
Hogy néznek ki különböző erősségű korrelációk?



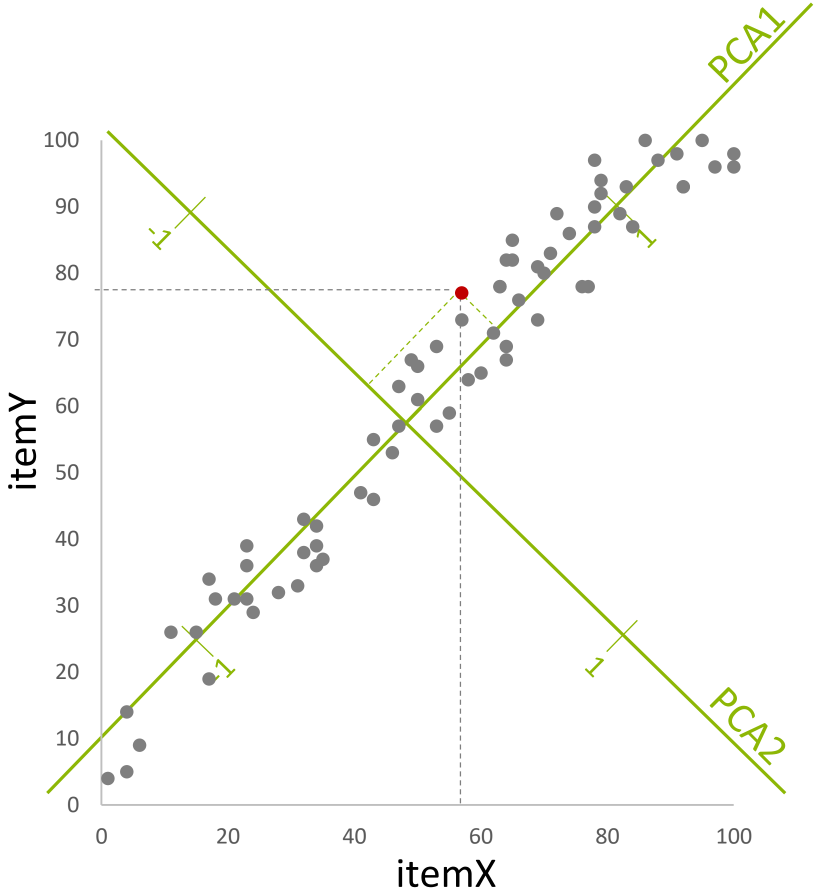


Főkomponens analízis

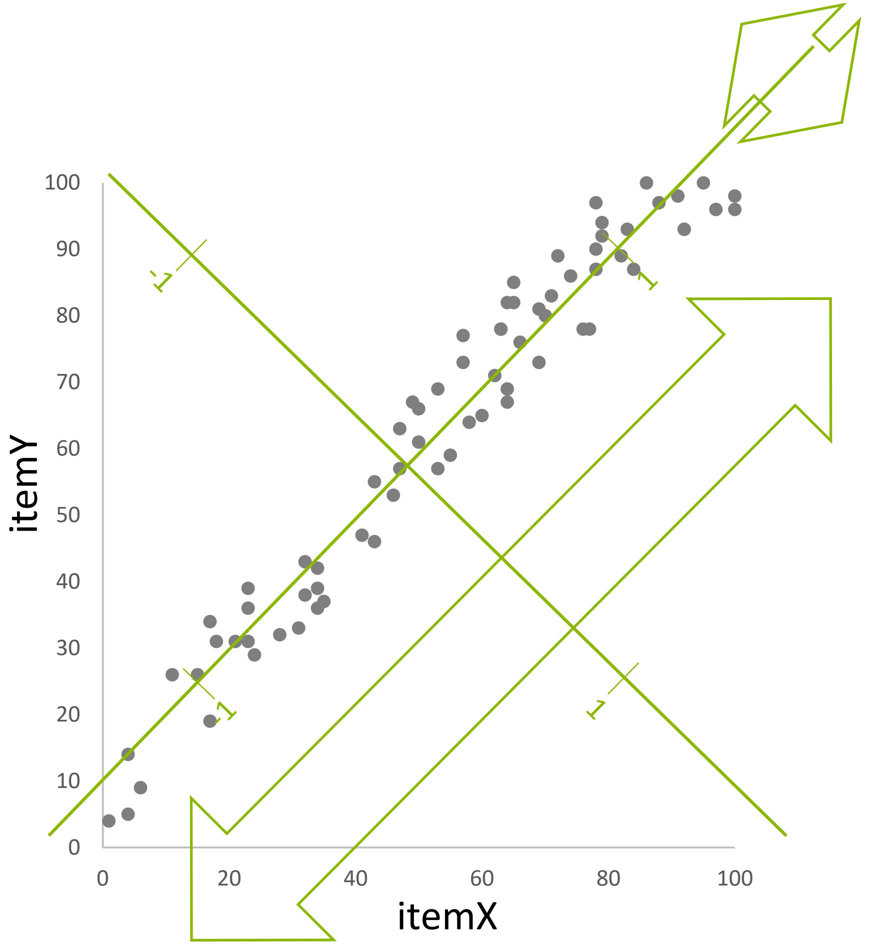
- Lépései
  - Az *n* darab item kifeszít egy *n* dimenziós teret, a pontfelhő ebben az *n* dimenziós térben helyezkedik el
  - Megkeressük azt az egyenest (első főkomponenst), mely az adatok varianciájának legnagyobb részét magyarázza. Azaz azt az egyenest, mely mentén legjobban szóródnak az adatok, ahol a pontfelhő a legvastagabb
  - Majd megkeressük az az egyenest (további főkomponensek), mely az előző egyenesekkel korrelálatlan / merőleges rá / független tőle, és a maradék variancának legnagyobb részét magyarázza
  - Ezt addig folytatjuk, amíg marad meg nem magyarázott variancia
  - Ennek eredményeként annyi komponenst (új dimenziót) kapunk, ahány dimenziós a tér eredetileg volt, tehát 2 item leírható egy két dimenziós térrel, és ebben két komponens található
  - A komponensek sorba vannak rendezve aszerint, hogy mennyit magyaráznak a pontfelhő varianciájából: az első a legtöbbet, az utolsó a legkevesebbet.

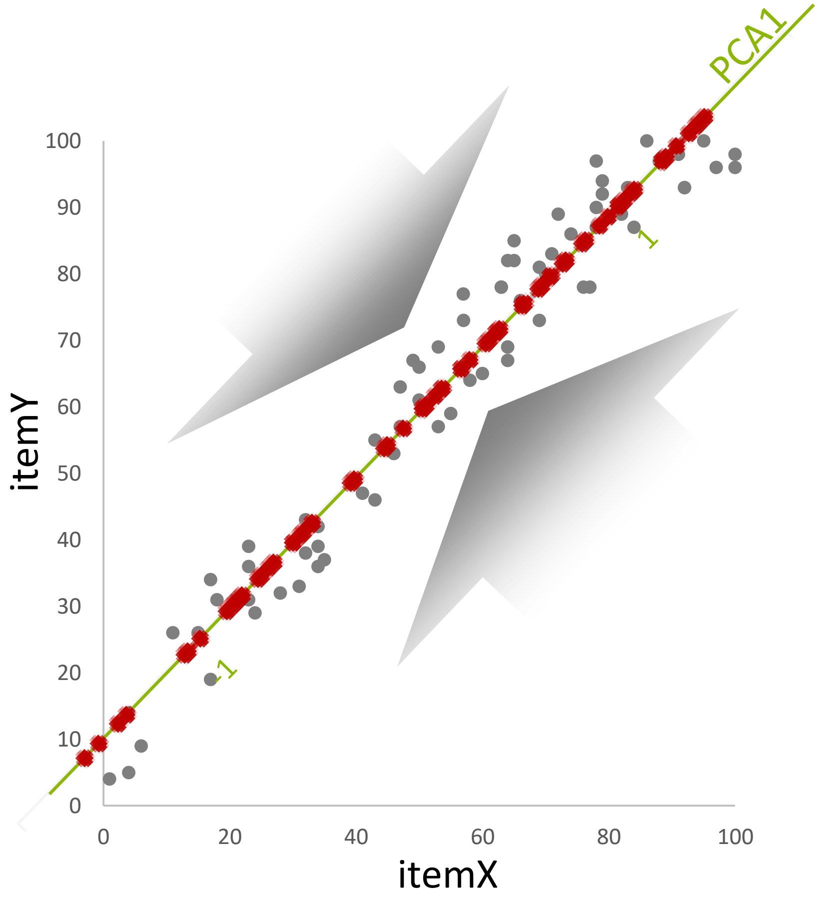


- Az új komponensek mentén ugyanúgy meg lehet határozni egy személy értékét, mint az eredeti itemek segítségével
- A pirossal jelölt személy az eredeti tengelyen X=57; Y=77 értékeket vette fel, az új tengelyeken pedig PCA1 = 0.4; PCA2 = - 0.1 (emlékeztető a főkomponens tengelyek z-értékekben skálázódnak)
- Ha minden tengelyt megtartunk, akkor az adatokban lévő varianciát 100%-ban magyarázni tudjuk vele, tehát mindenkinek pontosan meg tudjuk adni a helyét a többdimenziós térben a főkomponens elemzésből származó tengelyekkel.

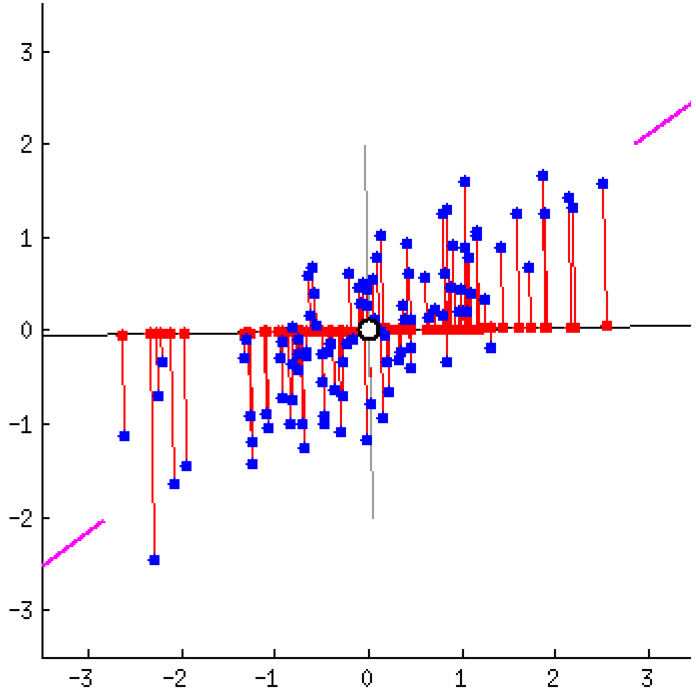


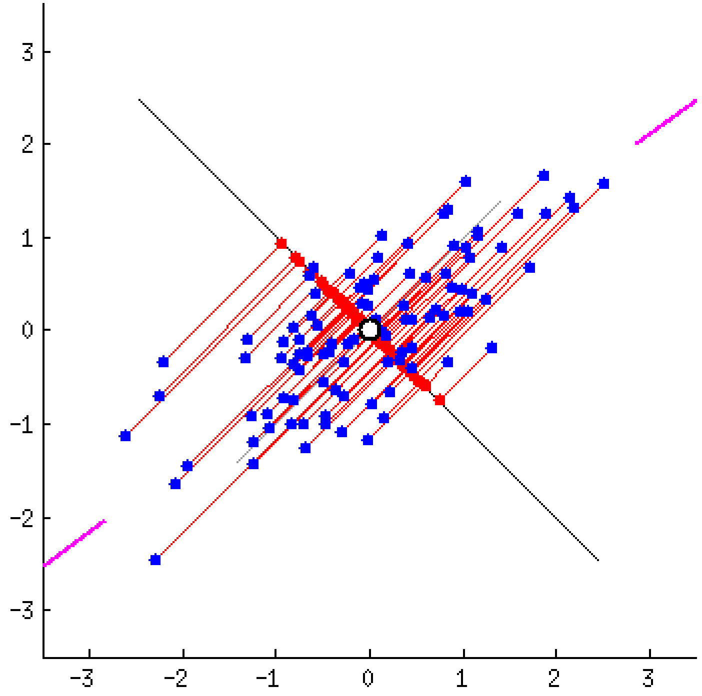
- Abban bízunk, hogy az adatok varianciájának nagy részét néhány főkomponenssel le lehet írni, azaz a többi főkomponens magyarázó ereje kicsi.
- A kis magyarázóerővel bíró komponensek viszonylag kis veszteséggel kidobhatók az elemzésből, ezáltal az adatok leírásához szükséges dimenziók (változók) száma csökkenthető.



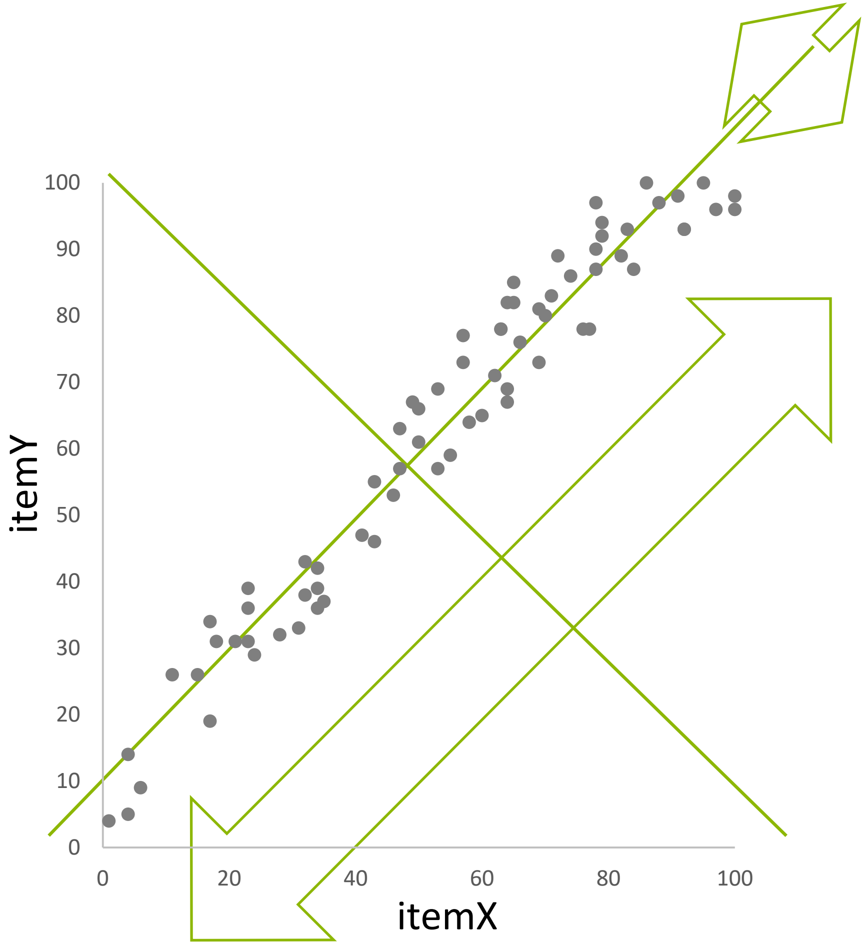


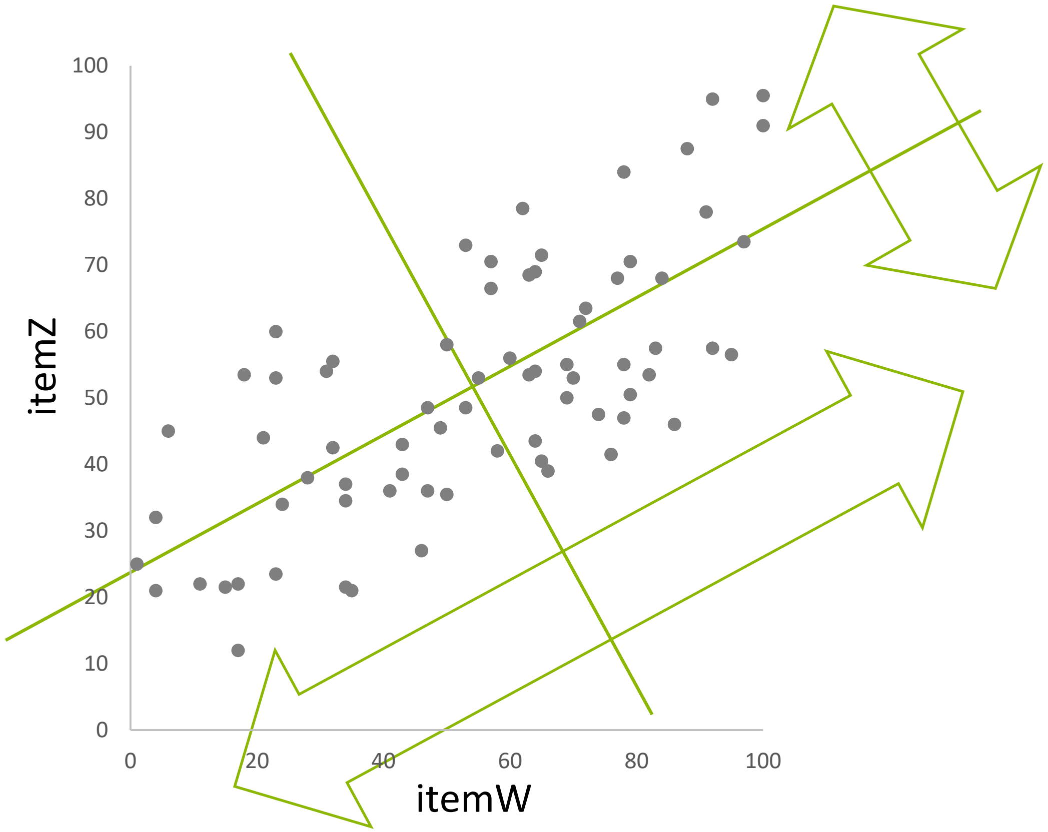
- Demostráció az előzőhöz. Az animációkat amoeba készítette Matlabbal, innen szedtem le őket: http://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues. Érdemes a linket megnézni, mert zseniális választ ad a „Hogyan magyarázd el a PCA-t a nagymamádnak” kérdésre!





- Ha nem mondhatom egy dimenzióra, hogy jelentéktelen, akkor megmarad. Például a jobb oldali pontfelhő leírásához szükség van mind a két komponensre, mert a második komponens kidobása túl nagy információveszteséget jelentene. Itt nem sikerült a dimenziók számát csökkenteni (kettő volt, kettő maradt), azaz nincsen adatredukció.



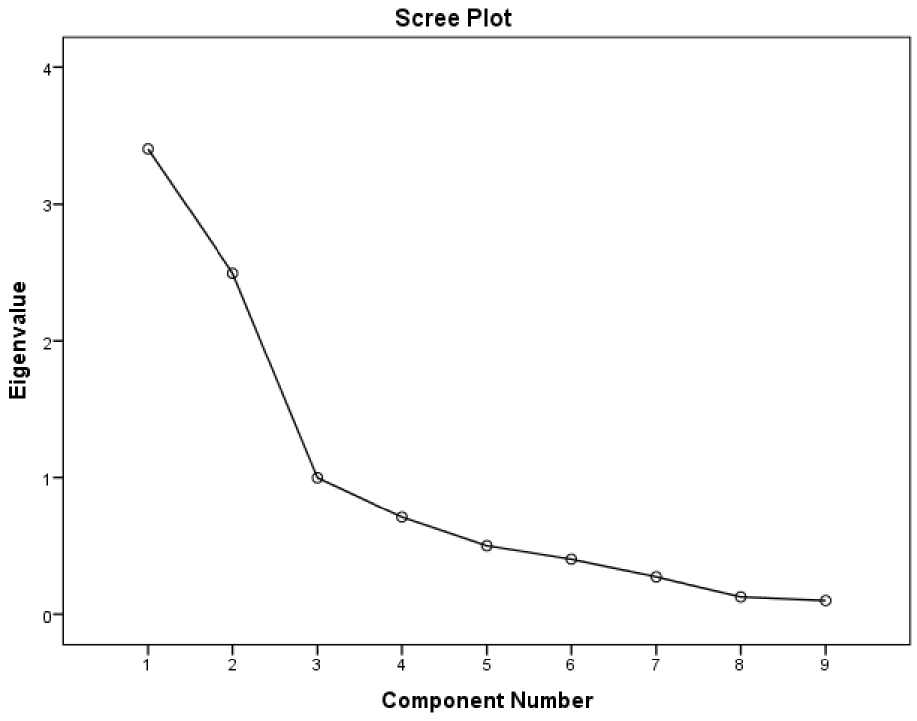


Sajátérték

- **Vissza az eredeti példánkhoz! Mivel 9 itemből áll a kérdőívünk, ezért létrejött 9 komponens. De hány komponenst tartsunk meg?**
- **Sajátérték (eigenvalue)**
  - Az adathalmaznak van valamennyi varianciája. Ebből a varianciából az első komponens 37,818%-ot magyaráz, a második 27,699%-ot, stb.
    Vedd észre, hogy, ha összeadod a megmagyarázott varianciákat, akkor megkapod a 100%-ot, azaz a 9 komponens megmagyarázza a 9 eredeti item teljes változatosságát.
  - A megmagyarázott varianciák átlaga 11,111%. Ezt megkapod úgy is, ha átlagolod őket, de úgy is, hogy 100%/9-cel (azaz a komponensek számával)
  - A **sajátérték** az komponens által megmagyarázott variancia és az átlagos variancia aránya. Például az első komponensnél ez 37,818/11,111=3.404. Jelentése, hogy az első komponens az átlagosnál háromszor több varianciát magyaráz meg.
    Látjuk, azt is, hogy a harmadik komponens majdnem pont az átlagbál van, a harmadiktól felfele, pedig már átlagosnál gyengébb komponenseket látunk.

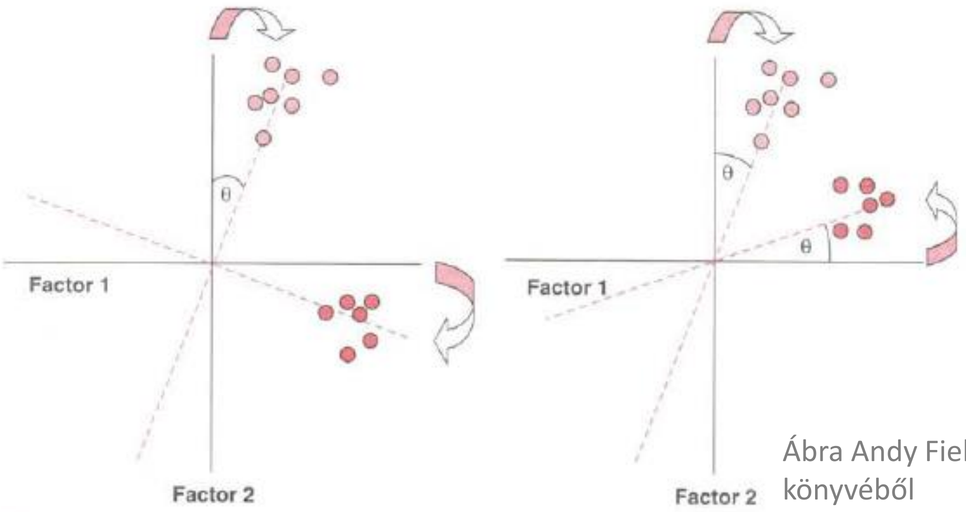
Komponens sorszáma	Hány százalékát magyarázza a komponens az adatok varianciájának?	Komponens sajátértéke
1	37,818	3,404
2	27,699	2,493
3	11,089	,998
4	7,876	,709
5	5,545	,499
6	4,454	,401
7	3,022	,272
8	1,397	,126
9	1,099	,099

- **Hány főkomponenst hagyjunk az elemzésben?**
  - Csak azokat, melyek a varianciából még elég sokat magyaráznak
  - Kaiser kritérium
    - Azokat, melyeknek a sajátértéke 1 fölött van (az átlagosnál többet magyaráznak a teljes varianciából)
    - Spss-ben ez történik alapból
    - Hátránya, hogy sok item esetén viszonylag sok komponenst „enged át”
  - Scree plot (könyök ábra)
    - Ha ábrázoljuk, hogy az egyes főkomponenseknek mennyi a sajátértéke, tipikus lefutást kapunk Néhány erős főkomponens után sok gyenge jön Ez a grafikonon a vonal töréseként (könyökeként) jelenik meg.
    - Azokat a főkomponenseket hagyjuk meg, melyek a töréspont előtt vannak
  - Paralel elemzés
    - Azokat tartsuk meg, melyek a véletlenül generált adatokból származó sajátértékek 95%-nál magasabbak.



Rotáció

- A létrejövő faktorok elforgatása, hogy minél inkább igaz legyen, a változók egy csoportja egy faktorra sokat tölt, a többire keveset
- Két rotációs módszer:
  - **Ortogonális:** megtartja a faktorok között a derékszöget / függetlenséget
  - **Ferde**: nem tartja meg a derékszöget / függetlenséget
    - Kutatói döntés: lehetséges, hogy a faktorok, melyeket találtam, tényleg nem függetlenek egymástól?
- **SPSS-ben**
  - **Ortogonális**:
    - **varimax:** a töltést egyenletesen szétosztja a faktorok között
    - **Quartimax**: kevés faktorra nagy töltést tesz
    - **Equamax**: kettő között
  - **Ferde:**
    - **Promax:** gyors, nagy adatállományokra tervezve
    - **Direct oblimin**: ha elméletileg várható a faktorok korrelációja



Ábra Andy Field könyvéből

Fogalom-összefoglaló

- **Itemek & korreláció**
  - Az elemzésbe bevett itemek egy n dimenziós teret feszítenek ki (n = itemek száma). Ha az itemek korrelálatlanok lennének, akkor e többdimenziós térben a személyek válaszai egyenletes pontfelhőt alkotnának. Helyette az itemek korrelálnak egymással, a pontfelhőben elnyúlások/sűrűsödések vannak
- **Komponens**
  - Az itemek közötti korrelációt kihasználva az n dimenziós pontfelhőnket írjuk le kevesebb dimenzióval (tengellyel), ezeket nevezzük komponensnek. Mivel egy n dimenziós teret kevesebb dimenzióval próbálunk leírni, természetesen valamennyi információt vesztünk a redukcióval.
- **Saját érték**
  - Megadja, hogy egy komponens mennyit magyaráz a többdimenziós pontfelhő varianciájából. A komponens „erősségét” írja le, ez alapján döntjük el, melyik komponenst tartjuk meg, melyiket dobjuk ki
- **Komponens töltés**
  - Megadja, egy item mennyire korrelál egy tengellyel (komponenssel), az item komponenssel való kapcsolatát írja le. Ez alapján döntjük el, melyik itemeket érdemes egy komponens értelmezésekor figyelembe venni
- **Forgatás**
  - Ha megvannak a tengelyeink (komponenseink), úgy forgatjuk őket, hogy minél inkább igaz legyen, hogy a tengelyre néhány item nagyon tölt, más itemek meg egyáltalán nem. Így könnyebb a őket értelmezni
- **Komponens érték**
  - A személy pontszáma az adott komponensen, skálán.
- **Komponens érték együttható**
  - Korrigált komponens töltés, a komponens érték kiszámolásakor ezzel súlyozzuk az itemeket, hogy a tengely szempontjából „fontosabb” tételek nagyobb hangsúllyal kerüljenek be a skála értékébe