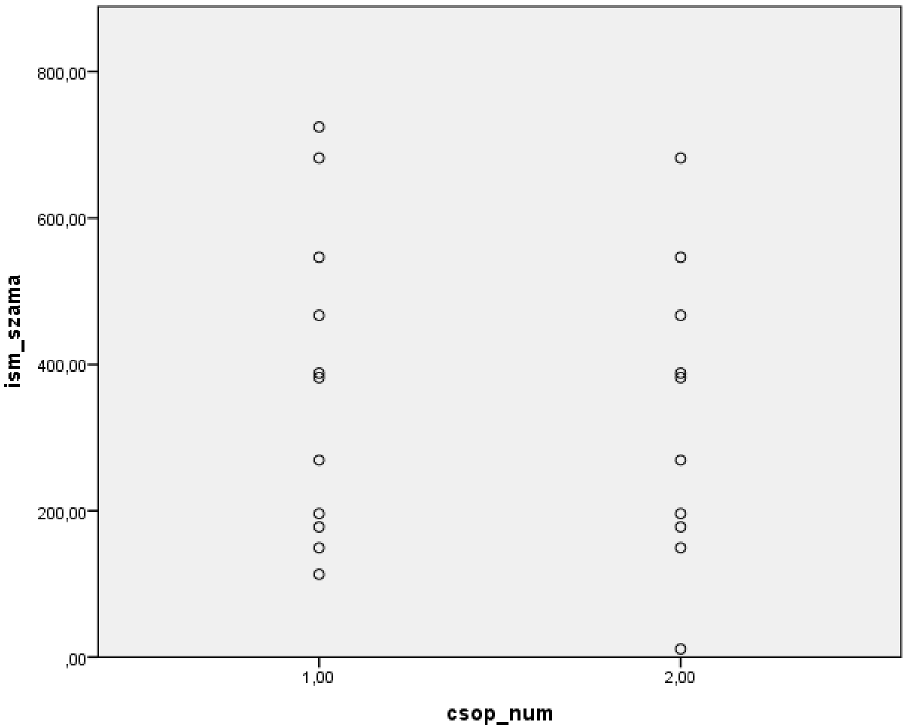
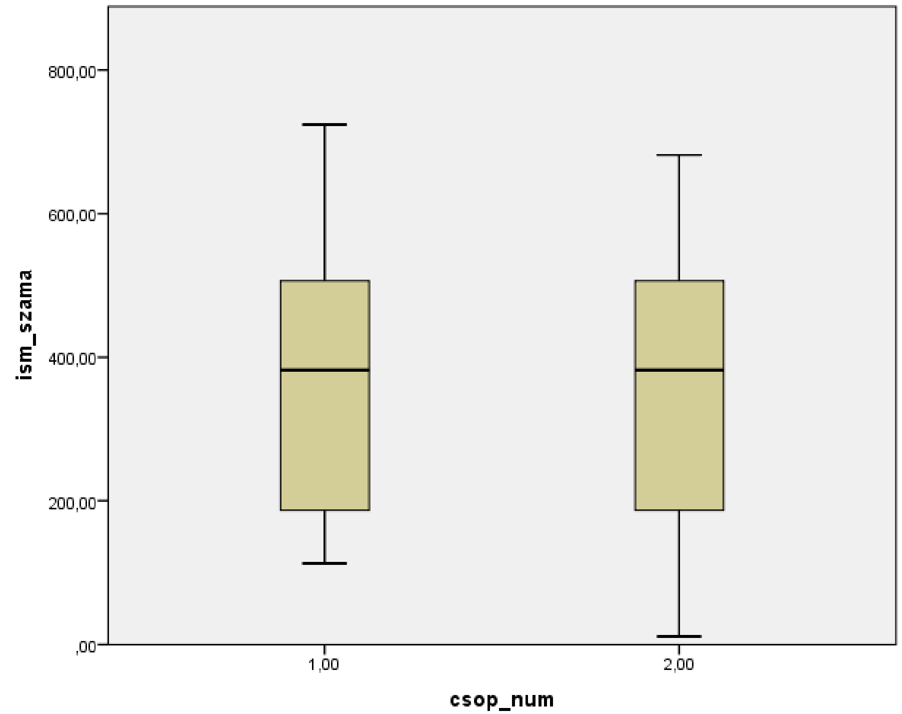
Feladatok:
pontdiagram és dobozdiagram





Hogyan csináltuk?

Alakmutatók:
ferdeség, csúcsosság

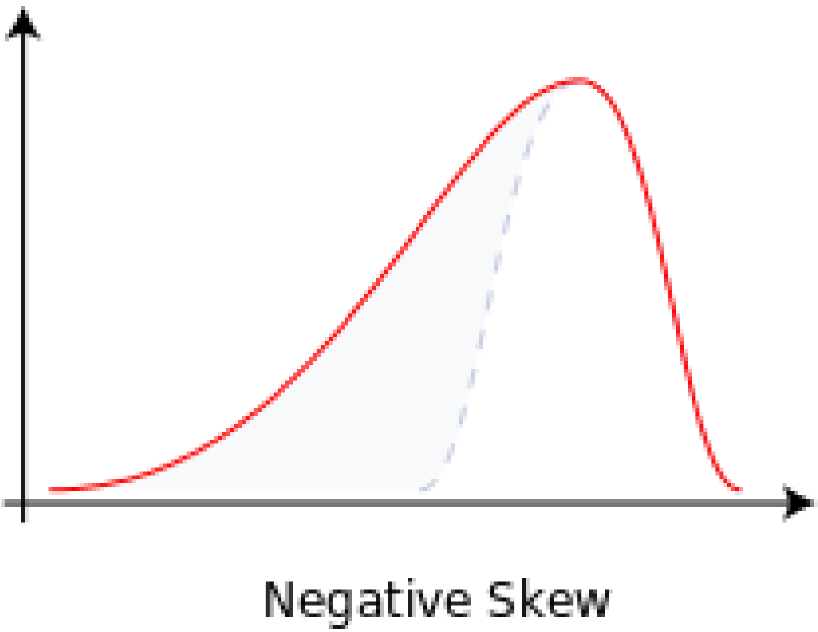
Alakmutatók a ferdeség és csúcsosság mérésére

- Ez eloszlás centrumát (középérték) és az adatok centrum körüli terpeszkedését (szóródási mutatók) már láttuk, de ez nem minden.

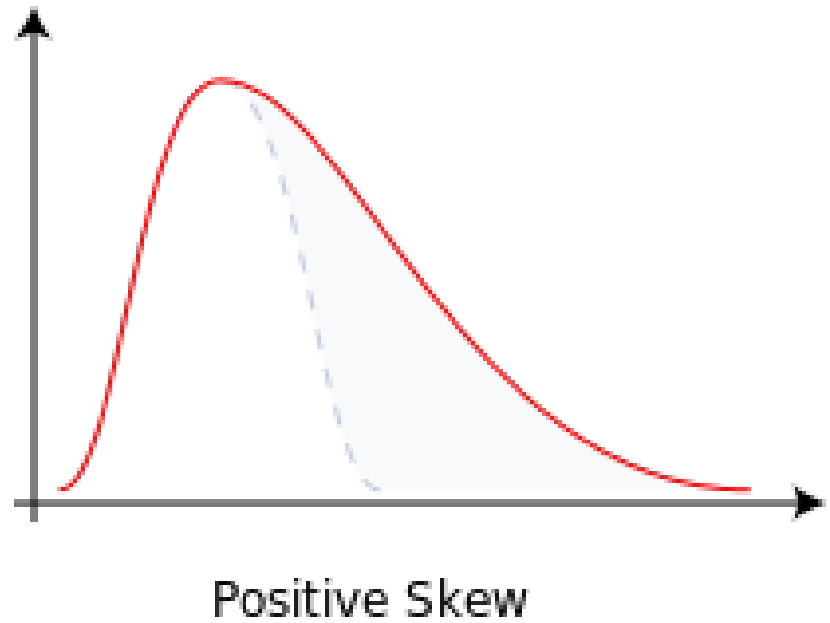
- Az eloszlás további fontos jellemzői:
- **Ferdeség (*skewness*):** szimmetria mértéke
- **Lapultság/csúcsosság (*kurtosis*) :** mennyire húzódik szét az eloszlás két széle, mennyire hegyes a centrumnál, mennyire lejtős

Ferdeség *(skewness)*

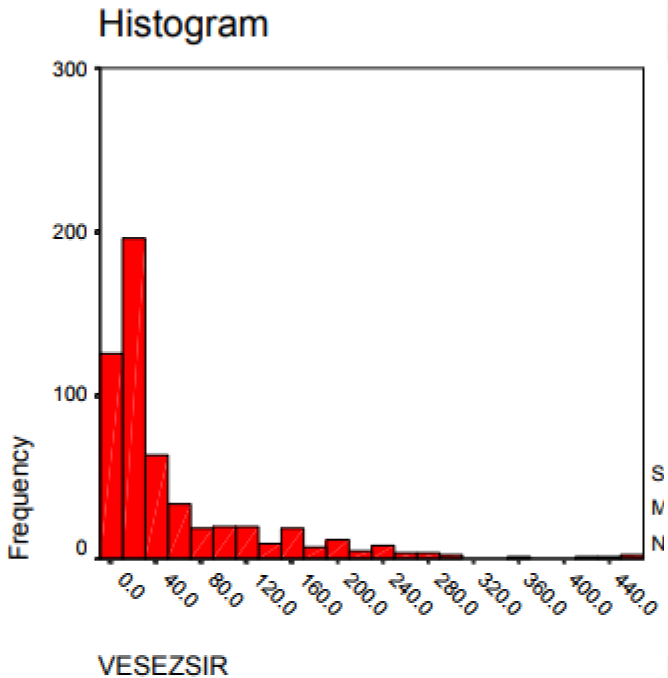
Balra ferde/negatív eloszlás



Jobbra ferde/pozitív eloszlás

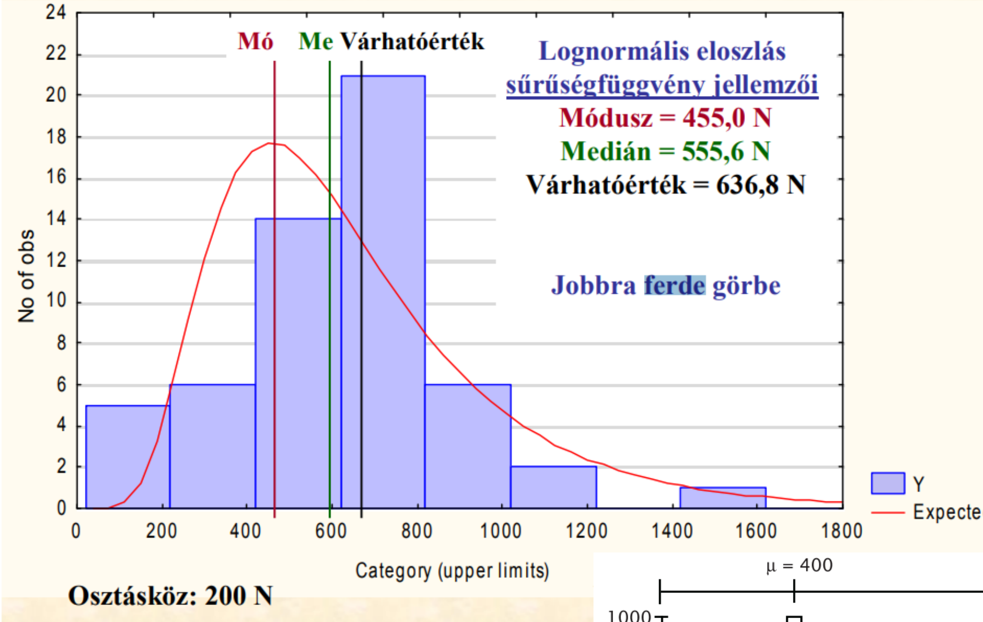


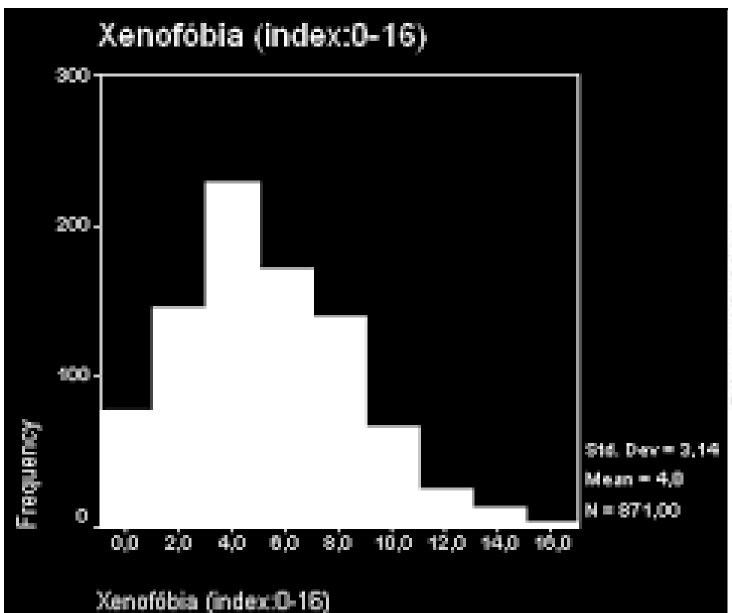
Szimmetrikus (pl. standard normális eloszlás):
ferdeség = 0

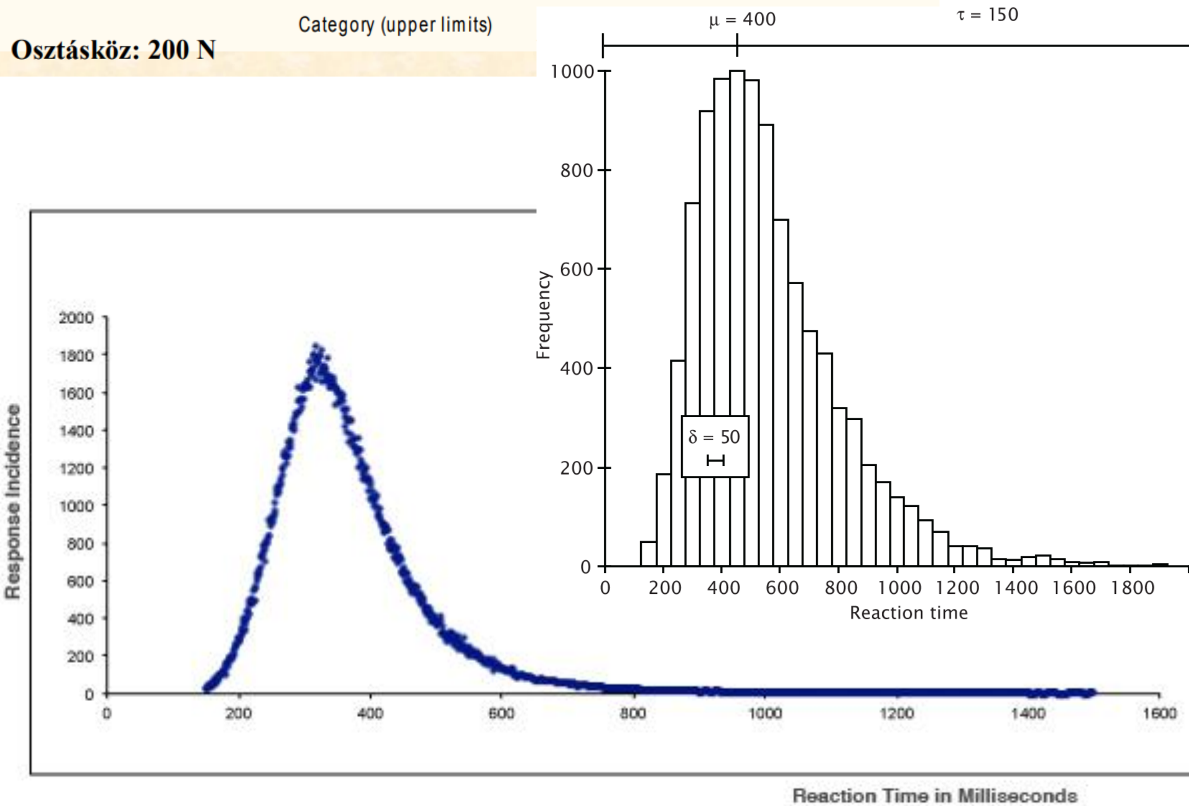


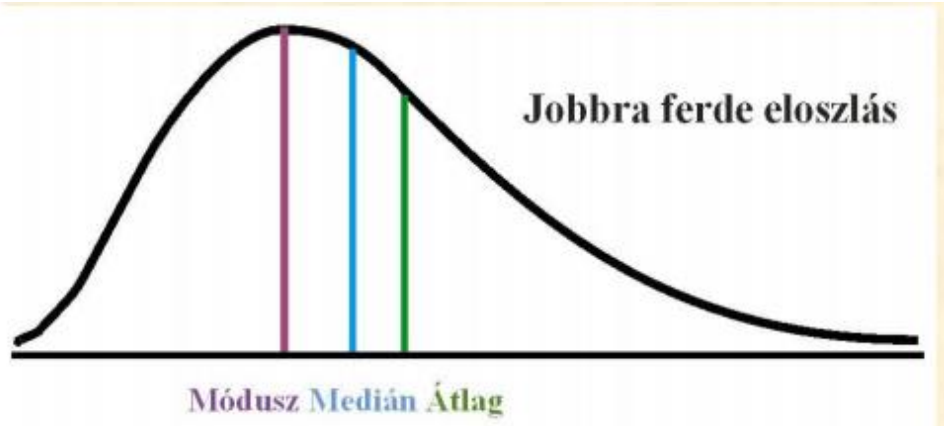
Őzek vese körüli zsír mennyisége:

Az óriás harcsa súlya valószínűségének gyakorisági görbéje (jobbra ferde) *lognormális eloszlás* feltételezésével

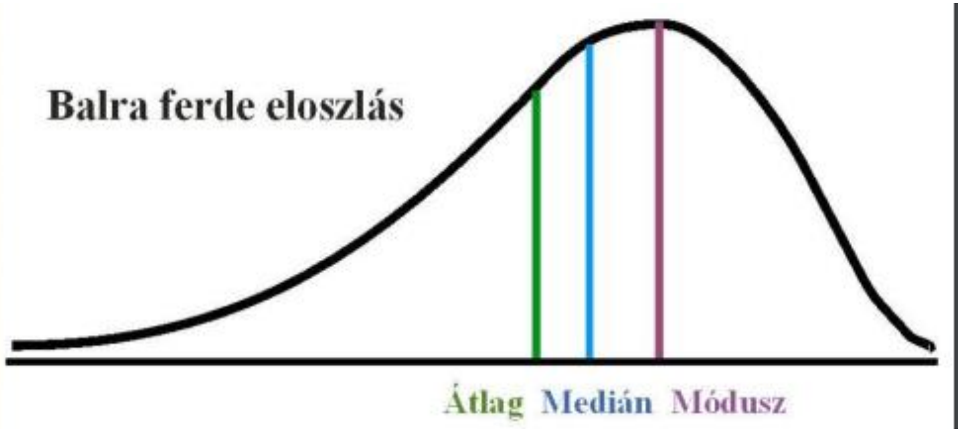




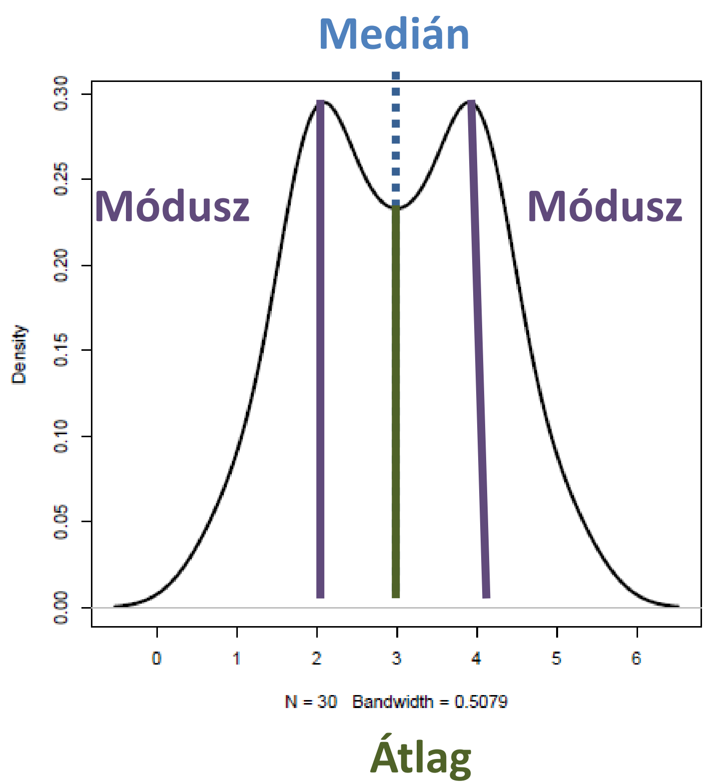




Jobbra ferde eloszlás
Módusz Medián Átlag



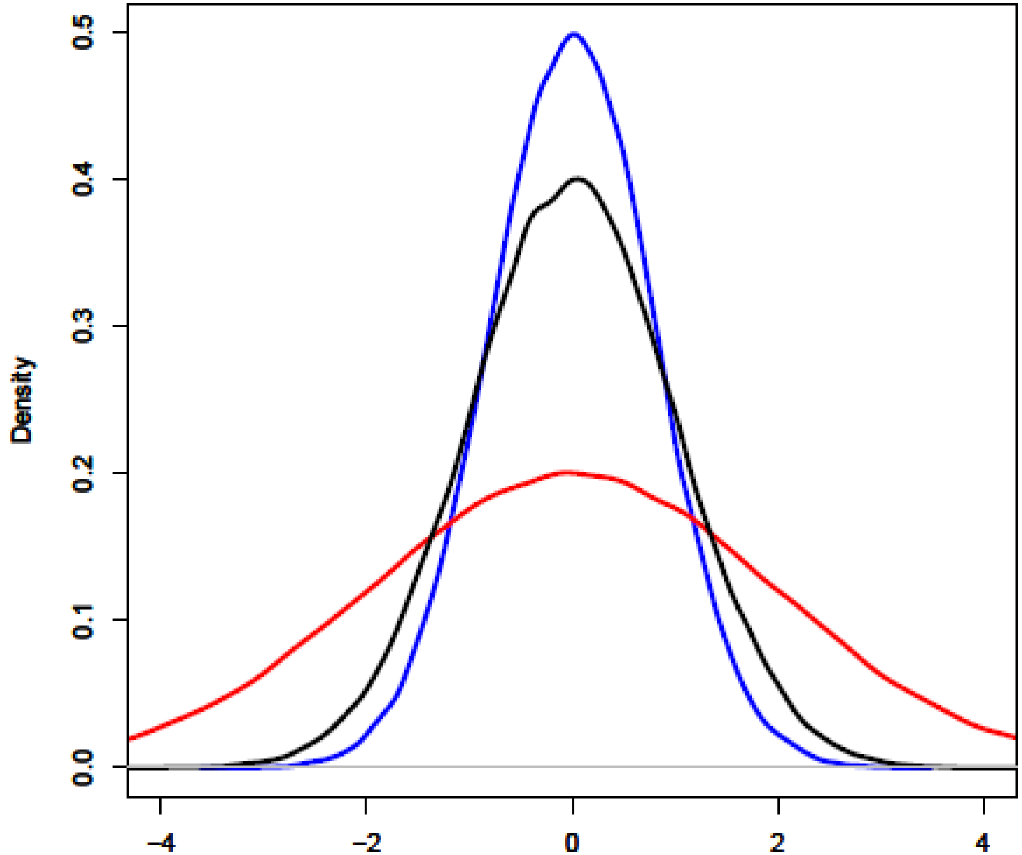
Balra ferde eloszlás
Átlag Medián Módusz



Medián
Módusz
Módusz
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0 1 2 3 4 5 6
N = 30 Bandwidth = 0.5079
Átlag

Lapultság/csúcsosság *(kurtosis)*

normális eloszlás:
csúcsosság = 0



Csúcsosabb eloszlás: a csúcsosság **pozitív**

Lapultabb eloszlás: a csúcsosság **negatív**

A lapultság valójában arra utal, hogy az 1 szórásnyira lévő adatok valóban az adatok 68%-át tartalmazzák-e! (ld. később)

Standardizálás (normalizálás)

- A ferdeség és csúcsosság paraméterek **függetlenek a konkrét értékektől (hol van az eloszlás közepe, milyenek a szóródási viszonyok)**.
- Viszont az eloszlások átlaga, szórása sokszor eltérhet → a „nyers adatok” nem összevethetők!
- Tehát a ferdeségi és csúcsossági együtthatók számításához valahogy **„közös nevezőre” kell hoznunk a különböző változókat**, hogy összehasonlíthatók legyenek.
- Ehhez egy **lineáris transzformáció**val megszüntetjük a konkrét szakmai dimenziót (mértékegységet, pl. kg, cm).

Standardizálás (normalizálás)

- Ha egy újszülött 56 cm és 4 kg, akkor melyik mérték szerint kiemelkedőbb (extrémebb)?
- Közvetlenül nem összevethető.
- Közös mérce: adjuk meg az adat nagyságát azzal, hogy a **populáció átlagától hány szórásnyira van**.

Magyar csecsemők átlagai és		az adatok szórása:
Testhossz:	49,8 cm	2,5 cm
Tömeg:	3,24 kg	0,5 kg

Az adott baba ehhez képest:

Testhossz: (56−49,8)/2,5 = 2,48

Tömeg: (4−3,24)/0,5 = 1,5

Standardizálás (normalizálás)

- Ha egy újszülött 56 cm és 4 kg, akkor melyik mérték szerint kiemelkedőbb (extrémebb)?
- Közvetlenül nem összevethető.
- Közös mérce: adjuk meg az adat nagyságát azzal, hogy a **populáció átlagától hány szórásnyira van**.

Magyar csecsemők átlagai és		az adatok szórása:
Testhossz:	49,8 cm	2,5 cm
Tömeg:	3,24 kg	0,5 kg

Az adott baba ehhez képest:

Testhossz: (56−49,8)/2,5 = 2,48 szélsőségesebb

Tömeg: (4−3,24)/0,5 = 1,5

Standardizálás (normalizálás) formalizálása: **Z-transzformáció**

- *X* kvantitatív változó
- Átlaga a populációban: μ
- Szórása a populációban: σ
- *X* változó standardizáltja: *Z*

$$Z = \frac{X - \mu}{\sigma}$$

- Ha valaki átlagos: *Z* = 0
- **Előjel!**
- Pozitív érték: a populációátlagnál *nagyobb* érték;
- negatív: a populációátlagnál *kisebb* érték.

Standardizálás (normalizálás): **Z-transzformáció**

- *Z* elméleti átlaga mindig 0
- *Z* elméleti szórása mindig 1
- Az így transzformált értékek átlaga és szórása tehát **független**
  - az eredeti értékek **mértékegységtől** (skálaléptékétől) és
  - **nagyságszintjétől** (középértékétől),
- és csak az eloszlás azon jellemzőit tükrözi, melyek ezektől függetlenek (lapultság, ferdeség), ezért a változók összehasonlíthatók!

Standardizálás a mintában

- Egy adott érték: x
- Mintaátlag: $\bar{x}$
- Minta szórása: s_s

$$z = \frac{x - \bar{x}}{s_s}$$

- Jelentése: **az adott érték hány szórásnyival tér el a mintaátlagtól (előjelesen!).**
- Vajon mikor használjuk?

Standardizálás: 1. példa

Likert skálás adatok: egy adott kérdésre válaszadás 1-től 5-ig terjedő skálán.

- egyesek (*a* típus) extrém értékeket adnak (1, 5), mások (*b* típus) csak a skála közepét használják (2, 3, 4)
- a két ksz átlaga hasonló (3 körüli), de a szórás különbözik!

adatközlő	típus	válaszok	átlag	szórás	Z-score
da	a	2	**3**	**1**	-1
da	a	3			0
da	a	4			1
bm	b	1	**3,33333**	**2,08167**	-1,1209
bm	b	5			0,800641
bm	b	4			0,320256

$$z = \frac{x - \bar{x}}{s_S}$$

Standardizálás: 1. példa

- Ha arra vagyunk kíváncsiak, hogy a kísérletben releváns kérdések esetében az adatközlők magukhoz képest melyik irányba és mennyire térnek el a legtöbbet adott értéktől, azaz nem a érdekel, hogy pontosan milyen értékeléseket adnak, hanem hogy az kirívó-e, vagy semleges.
- Ehhez az adatközlők saját egyéni átlagértékeire normalizálunk.

adatközlő	„típus”	válaszok	átlag	szórás	Z-score
da	a	2	3	1	-1
da	a	3			0
da	**a**	**4**			**1**
bm	b	1	3,33333	2,08167	-1,1209
bm	b	5			0,800641
bm	**b**	**4**			**0,320256**

Ugyanaz a 4-es értékelés a szűkebb tartományt használó esetében messzebb van, azaz „többet ér”, nagyobb a jelentősége.

Standardizálás: példa 2

- Ha az egész mintára standardizálunk: megállapíthatjuk, hogy mik a kilógó adatok, mert normális eloszlásban 3 szórásnál messzebb alig vannak adatok!
- Lásd később a normális eloszlás tulajdonságainál
- Mi a gond a kilógó adatokkal?
- Elrontják a normális eloszlást → ekkor nem használhatók parametrikus próbák.

Standardizálás: példa 3

- Korreláció!
- Két változó közti (nem feltétlenül ok-okozati) összefüggés vizsgálata (ha az egyik változik, változik-e a másik és ha igen, hogyan).
- Vajon az ember testtömege (kg) nő az életkor előrehaladtával (év)?
- A kettő közti összefüggés vizsgálatához el kell tüntetni a (szakmai) **dimenziót** (= mértékegység!), és valahogy közös nagyságrenden kezelni az értékeket!

Feladat 1-3

Januárban született 12 csecsemő testsúlyadatai:

csecsemok.xlsx

3 3,4 3,5 3,83 3,6 2,9 2,52 3,9 4 3,8 3,75 4,4

Testhosszadatai:

49,8 47,5 48 46,3 47 44 49 48,1 49,5 49 45 49

Nemek:

Nő ffi nő ffi ffi nő ffi nő nő nő nő nő nő

Számoljuk ki: összesített és csoportokra bontott átlagot, mediánt, szórást. → excel.

Ábrázoljuk grafikonon a nemek eloszlását a mintában! darabteli()

Feladat 4-5-6

Januárban született 12 csecsemő testsúlyadatai:

3 3,4 3,5 3,83 3,6 2,9 2,52 3,9 4 3,8 3,75 4,4

Testhosszadatai:

49,8 47,5 48 46,3 47 48 49 48,1 49,5 49 45 49

Nemek:

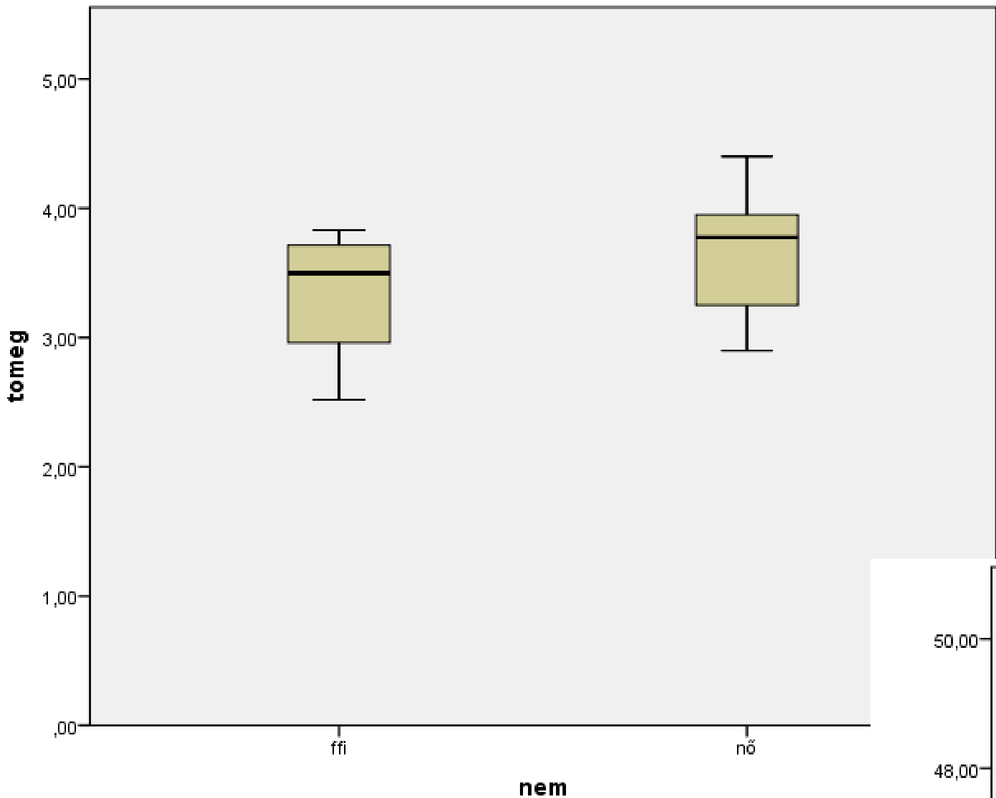
Nő ffi nő ffi ffi nő ffi nő nő nő nő nő nő

A fiúk vagy a lányok születési súlya nagyobb?
A fiúk vagy a lányok születési hossza nagyobb? →

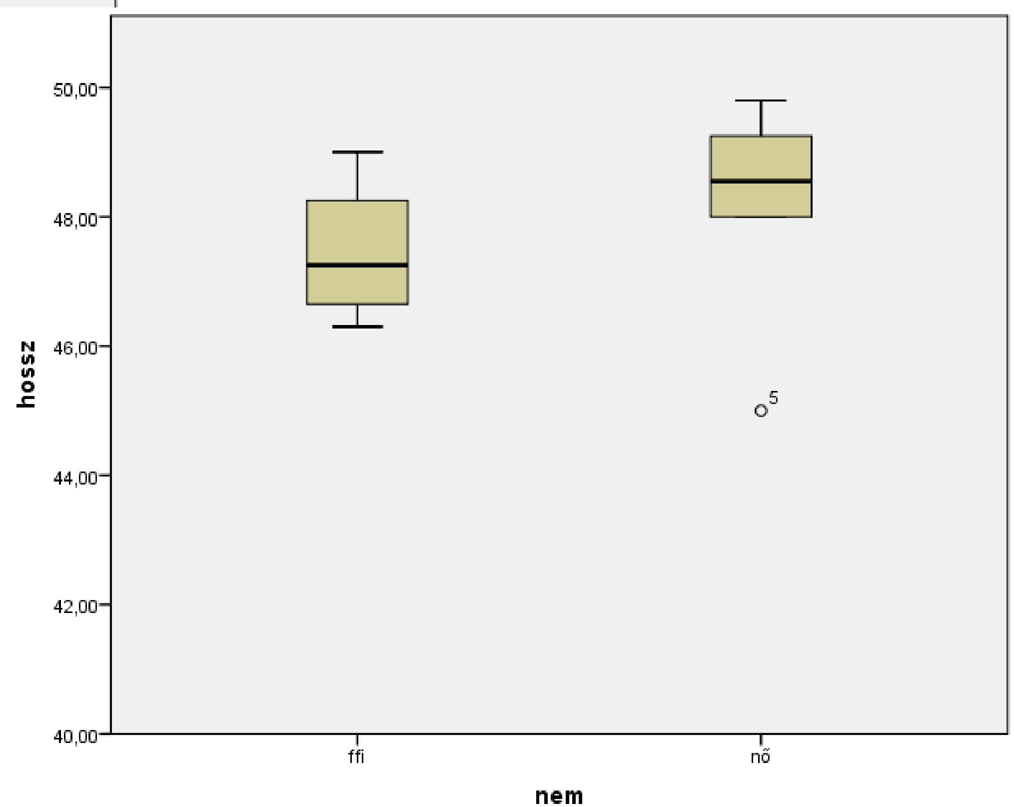
Számoljuk ki: átlag + szórás (excel)

Ábrázoljuk (SPSS):

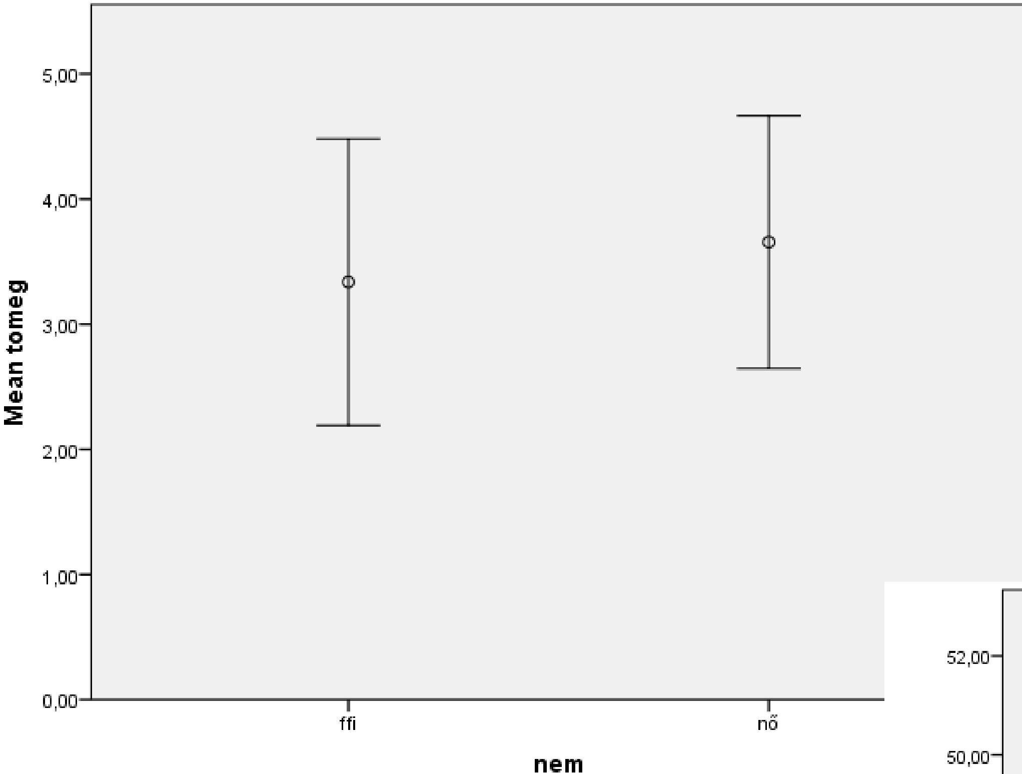
- medián és interkvartilis tartomány (SPSS)
- pont diagram hibasávokkal (SPSS)
- oszlopdiagram hibasávokkal (excel)



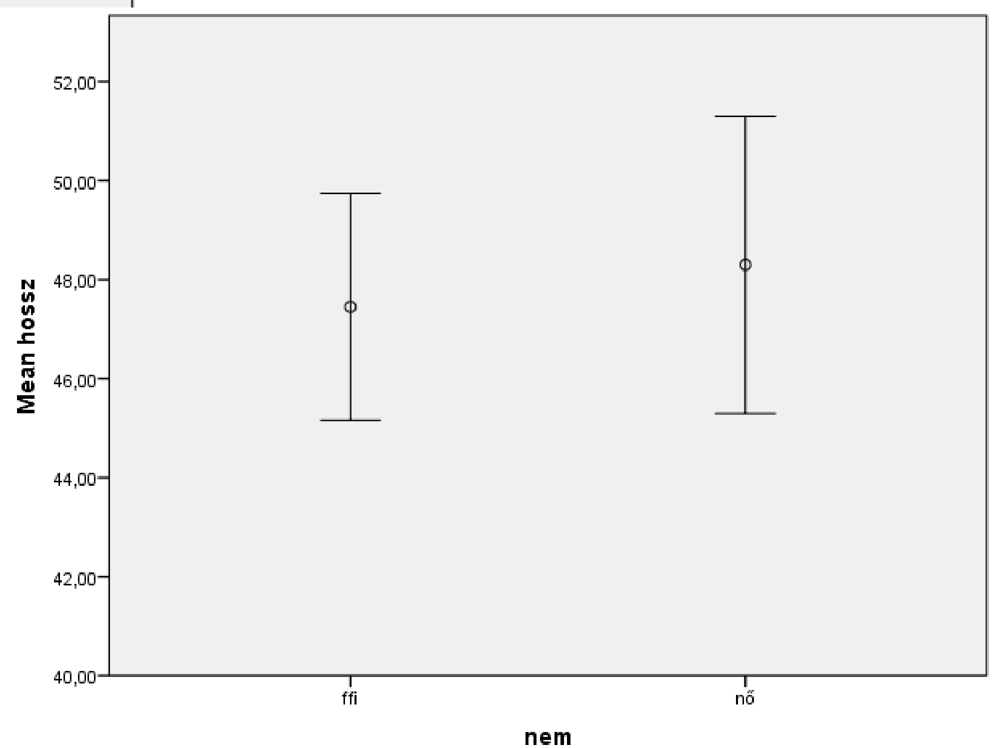
5,00
4,00
3,00
2,00
1,00
,00
tomeg
ffi
nő
nem



50,00
48,00
46,00
44,00
42,00
40,00
hossz
5
ffi
nő
nem



Error Bars: +/- 2 SD



Error Bars: +/- 2 SD

Feladat II

Súlyban vagy hosszban kiemelkedőbbek a lányok?

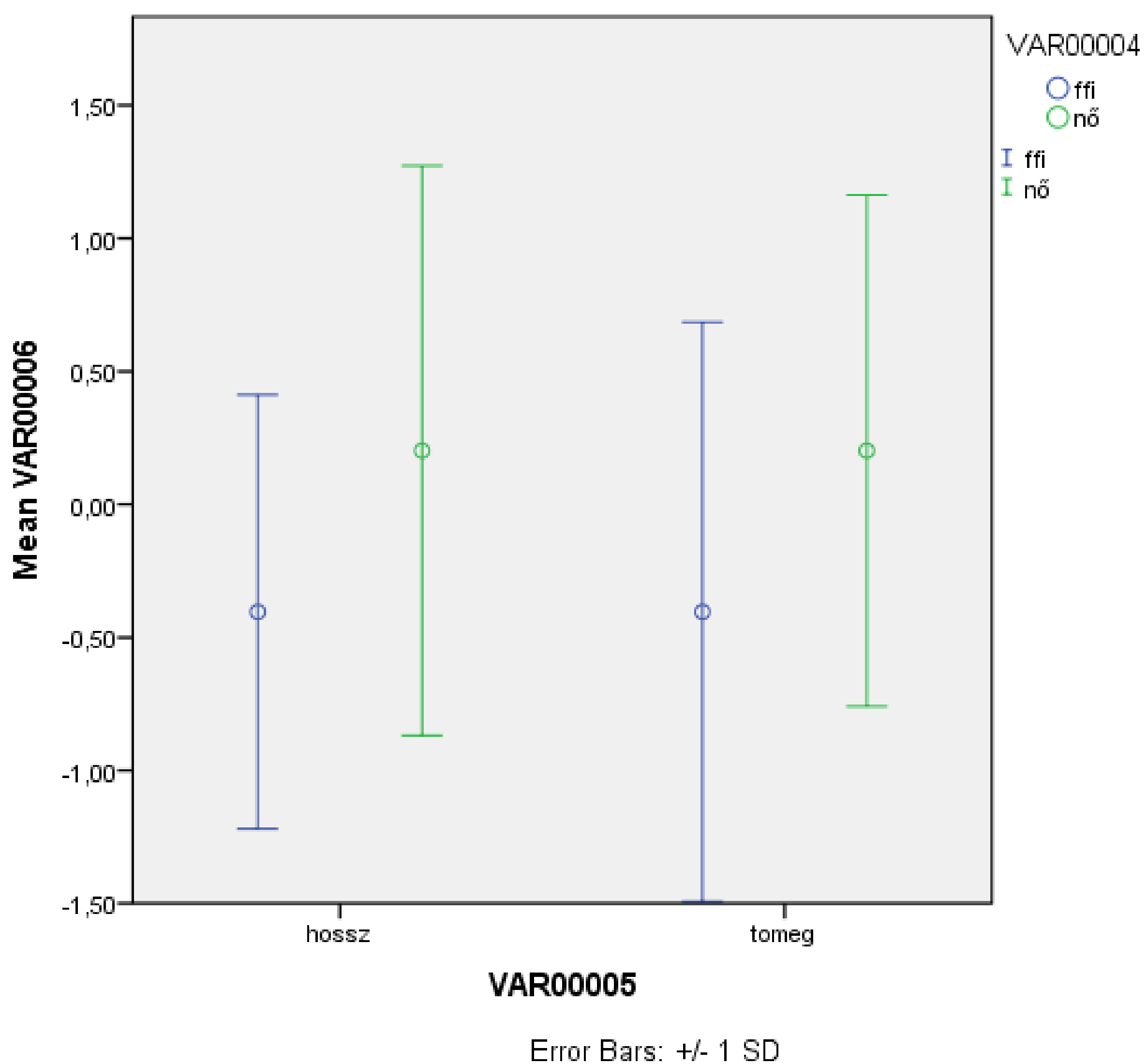
Mit kell tennünk, hogy összehasonlíthatók és együttesen ábrázolhatók legyenek a tömeg és hossz adatok?

Normalizáljuk az adatokat. → excel.

$$z = \frac{x-\bar{x}}{s_s}$$

= (x-átlag())/szórása()

A standardizált adatok összehasonlítása grafikusan: SPSS → csoportosított pont diagram hibasávval ($\pm$ 1 SD)


VAR00004
ffi
nő
ffi
nő
1,50
1,00
0,50
0,00
-0,50
-1,00
-1,50
Mean VAR00006
hossz
tomeg
VAR00005
Error Bars: +/- 1 SD

Házi feladat

- Formázzuk meg az SPSS adattábláinkat!
- Formázzuk meg az ábráinkat!

Excel

- Rögzített sorok/oszlopok
- Sorba rendezés
- Aritmetikai függvények *(szum, szumha, szórása, átlag, medián, módusz)*
- Logikai függvények *(bal, jobb, darab, darab2, darabteli)*
- Szövegből oszlopok
- Grafikonok típusai (köv óra)