

Korreláció

Statisztika I., 3. alkalom

Két változó kapcsolata

A korábban használt eljárások két változót nagyságrendileg hasonlítanak össze.

Gyakran sokkal inkább az érdekel bennünket, hogy *két jelenség mennyire függ össze*.

Példák:

- Az anya aktivitásszintje mennyire függ össze a kisgyermek aktivitásszintjével.
- Mennyire erős a kapcsolat a munkahelyi légkör szubjektív megítélése és az egyén teljesítmény között.
- A felvételin elért pontszám mennyire jósolja be az első éves átlagot.

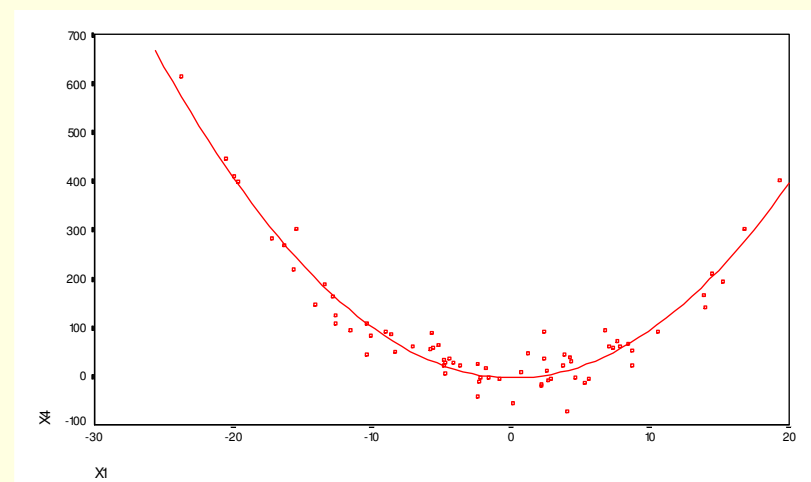
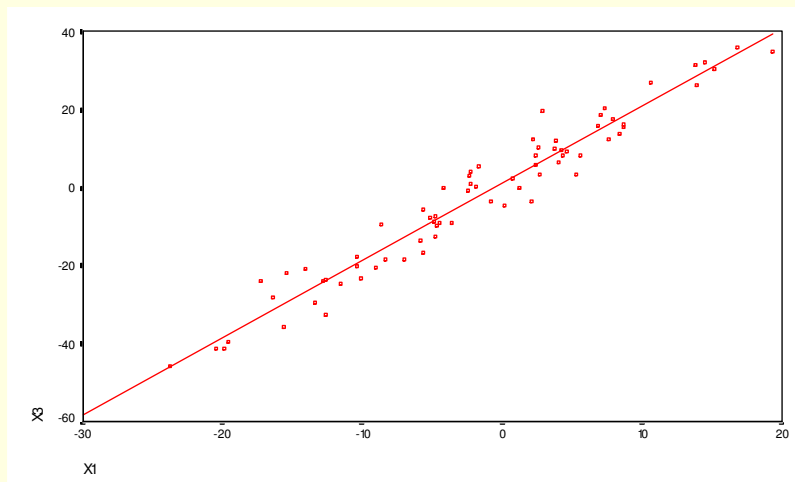
Ha nincs kapcsolat, a két jelenség független.

Ha van kapcsolat a két jelenség között, akkor fontos az összefüggés jellege.

Ha két változó normális eloszlású, akkor csak lineáris kapcsolat képzelhető el.

A pontdiagram (scatterplot) segít két változó kapcsolatának vizuális feltérképezésében.

Két változó kapcsolata



(Ábrák a Máth jegyzetből)

Korreláció

A lineáris kapcsolat erősségének mérésére a korreláció szolgál.

A korreláció értéke minimum -1 és maximum 1.

Ha a korreláció értéke 1, akkor a két változó kapcsolata tökéletes egyenes arányosság.

0, akkor nincs kapcsolat a két változó között, függetlenek.

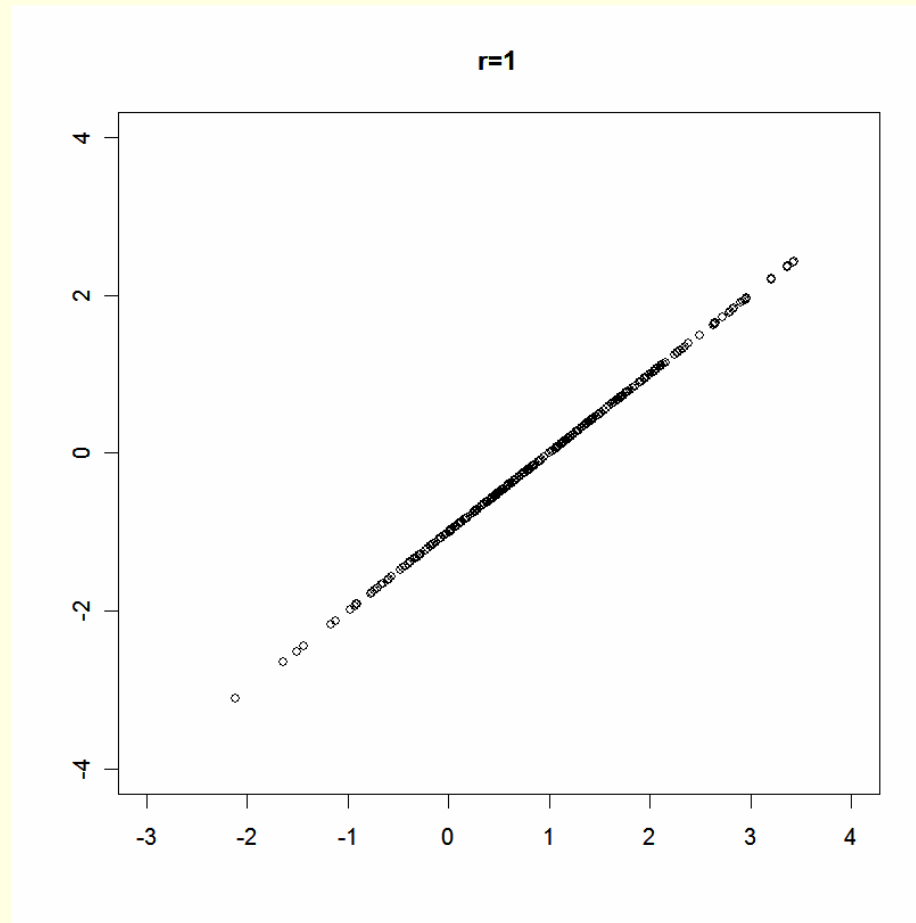
-1, akkor a két változó szintén tökéletes összhangban van, de kapcsolat jellege fordított arányosság.

a

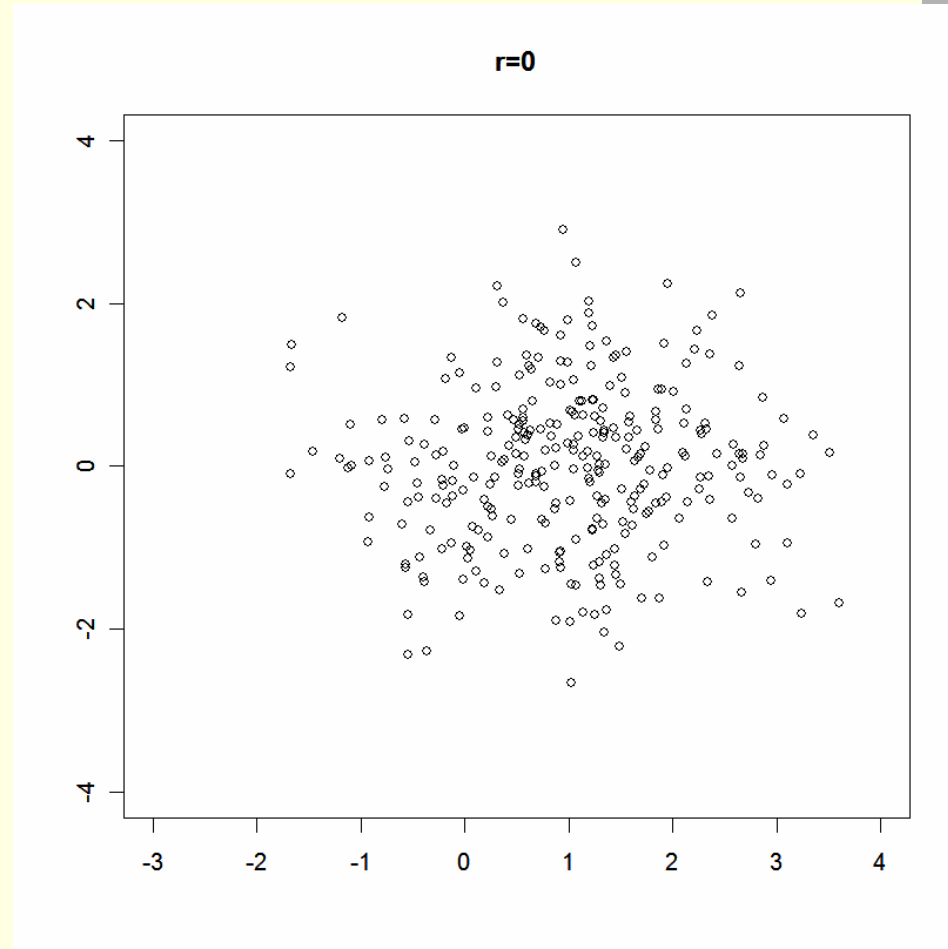
Ha két változó normális eloszlású, akkor csak lineáris kapcsolat képzelhető el, azaz, ha nincs közöttük lineáris kapcsolat, akkor függetlenek egymástól.

A korreláció kölcsönös kapcsolatot jelent.

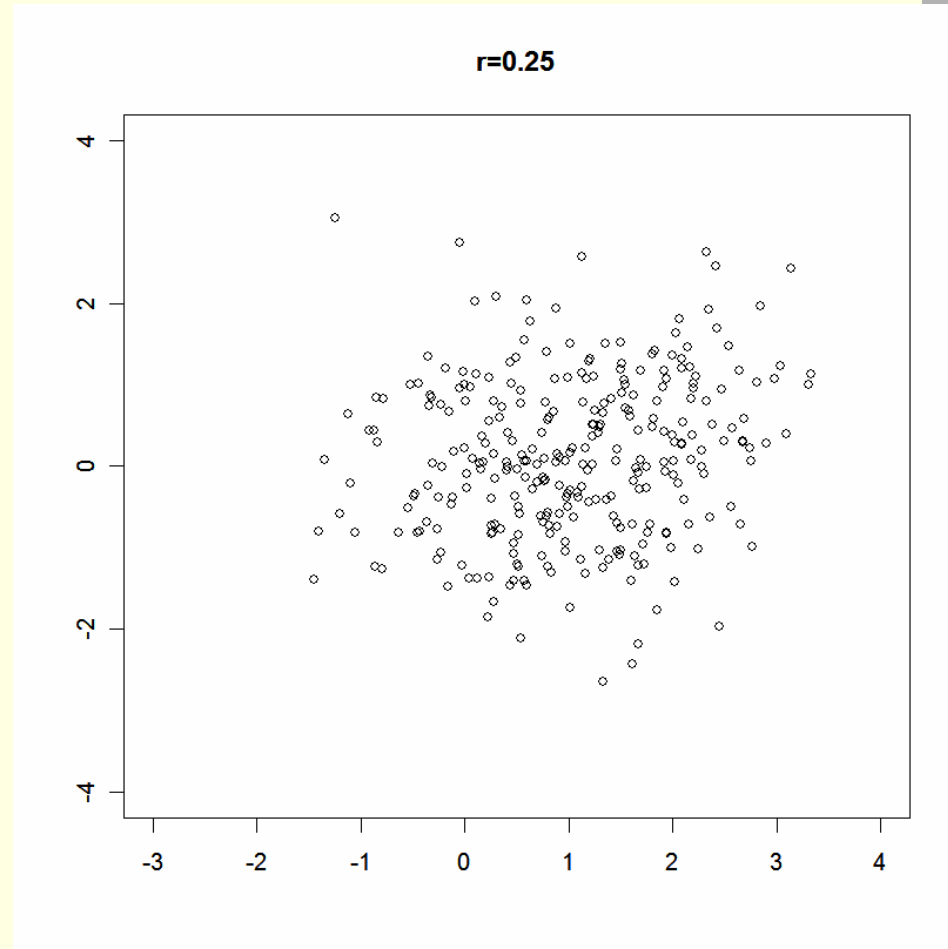
Korreláció



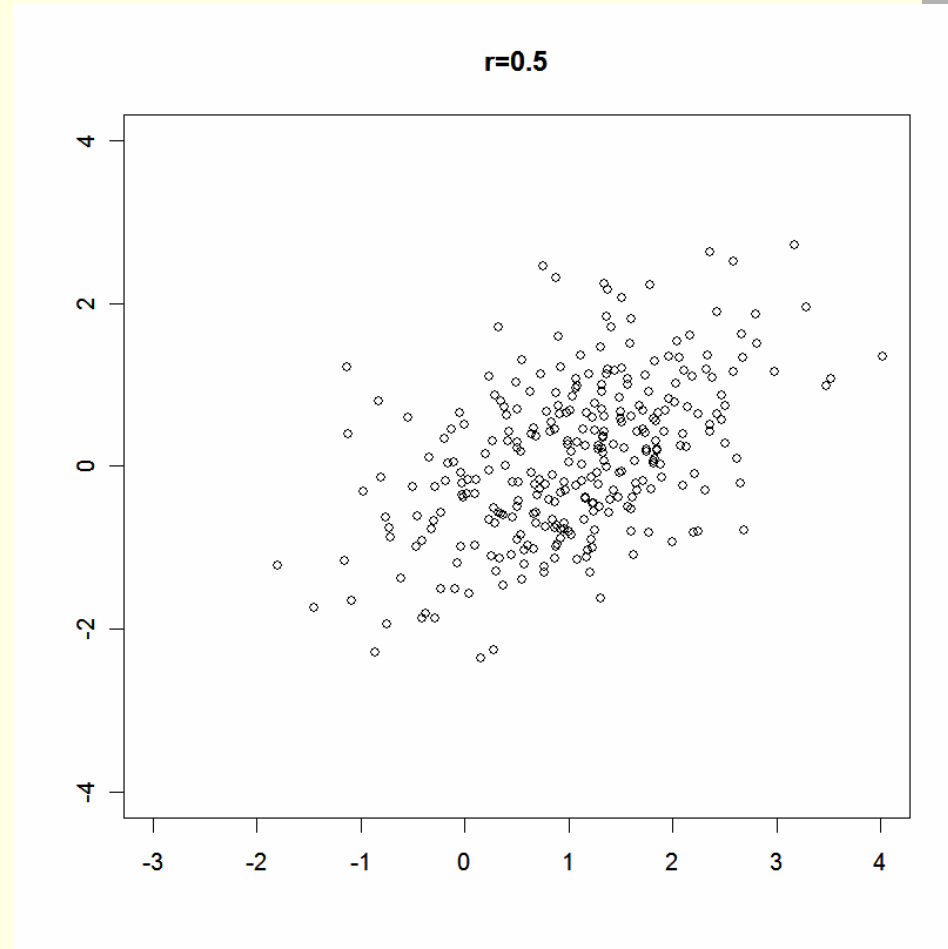
Korreláció



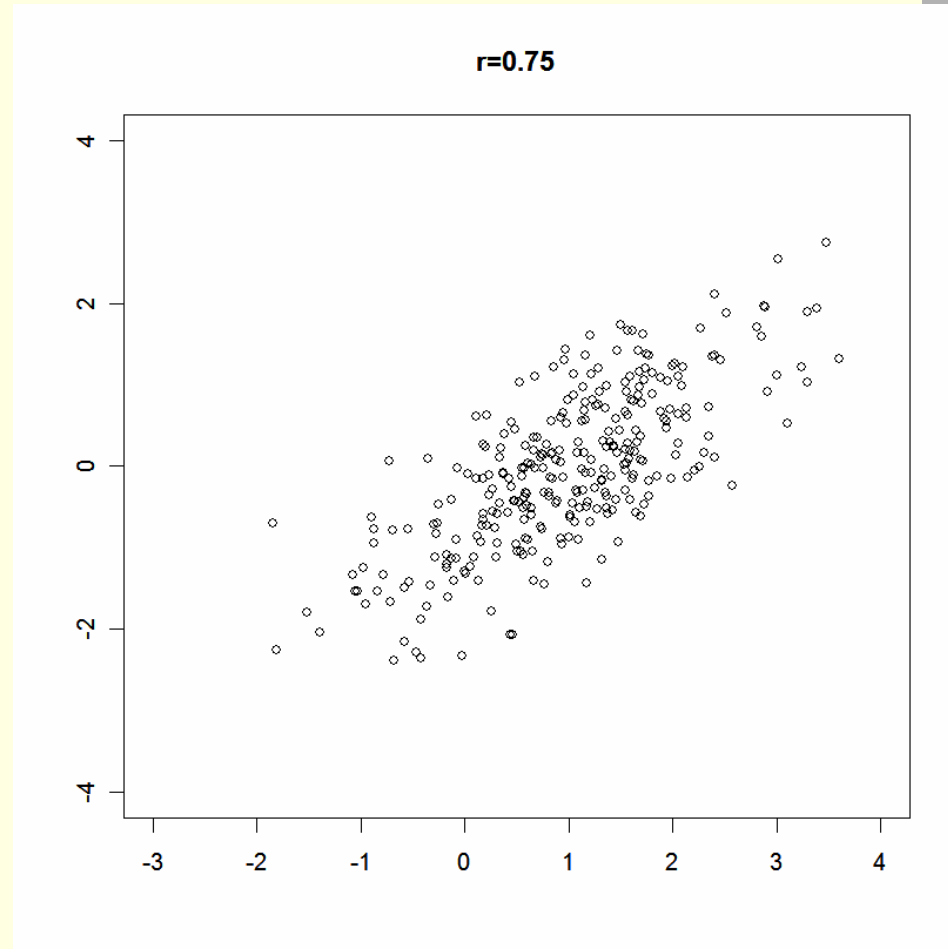
Korreláció



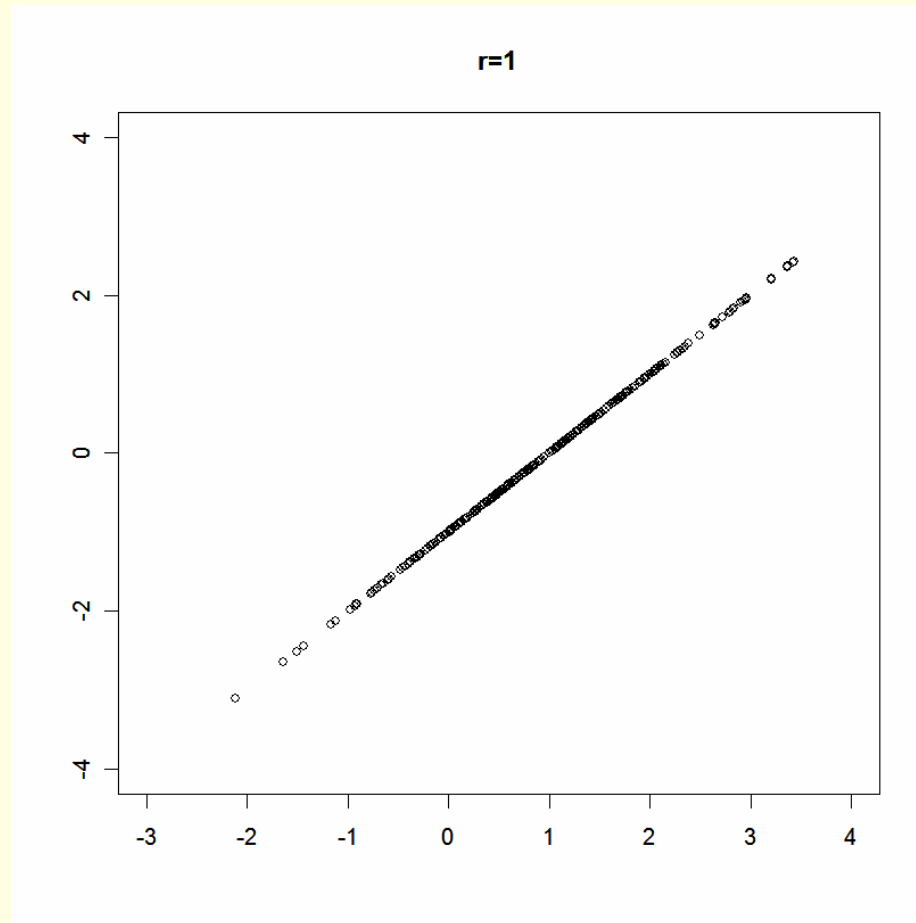
Korreláció



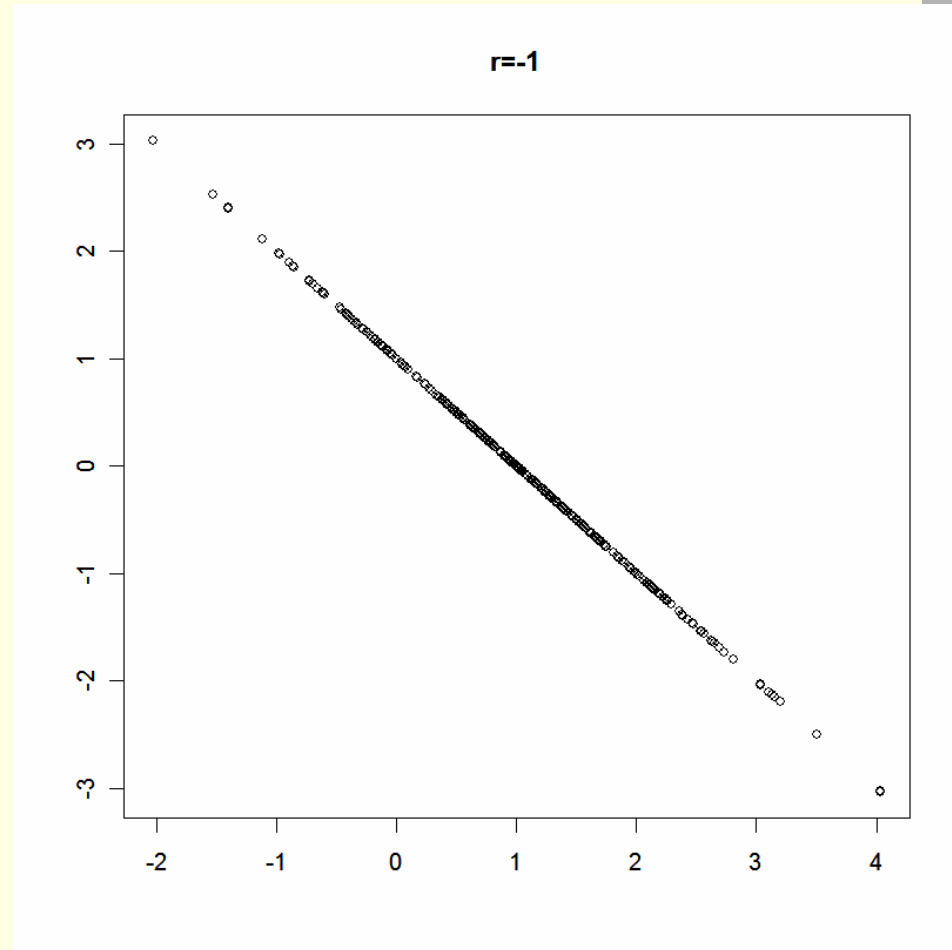
Korreláció



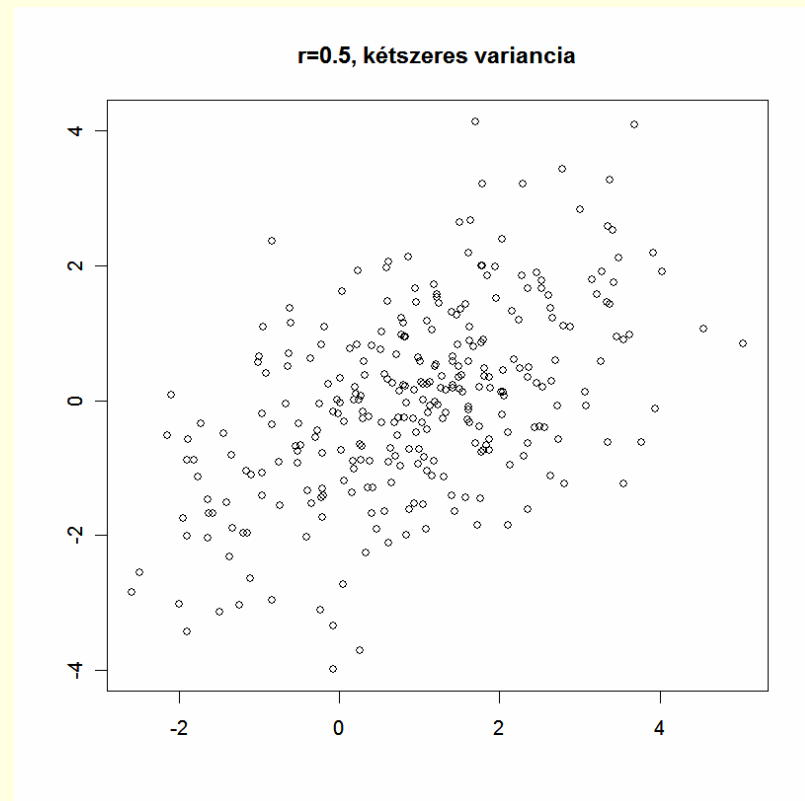
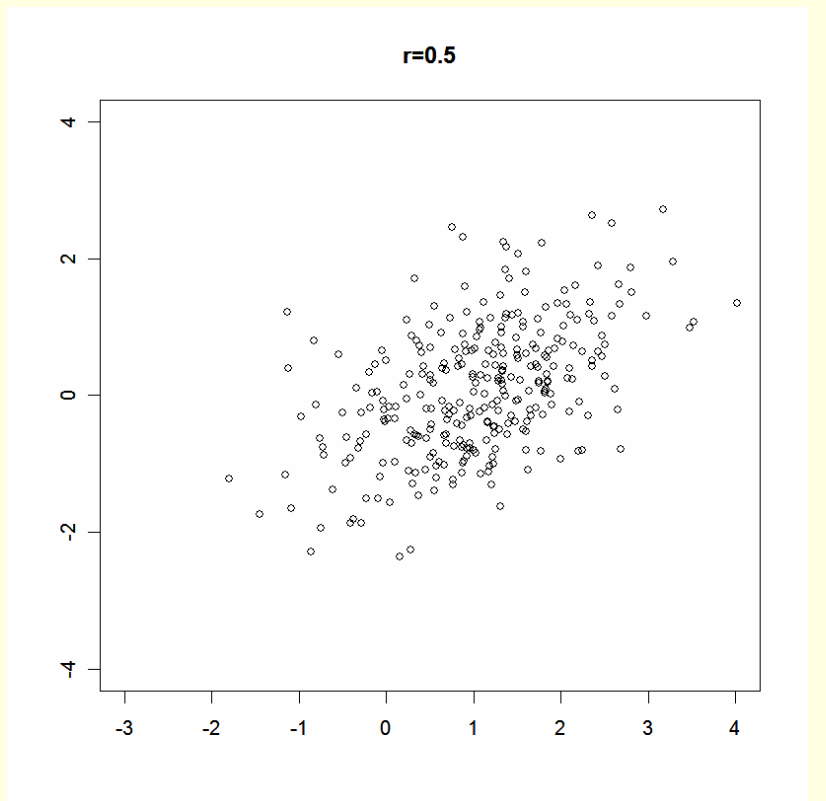
Korreláció



Korreláció



Korreláció



Korreláció

A korreláció populációra vonatkozó értelmezése csak akkor ésszerű ha mindkét változó random módon lett a populációból kiválasztva.

Több megközelítés létezik a korrelációértékből a kapcsolatra vonatkozó következtetésre.

(Guilford, 1950):

- 0 : nincs lineáris kapcsolat
- 0 - 0.2 (-0.2 - 0) : gyenge, majdnem hanyagolható kapcsolat
- 0.2 - 0.4 (-0.4 - -0.2) : biztos, de gyenge kapcsolat
- 0.4 - 0.7 (-0.7 - -0.4) : közepes korreláció, jelenős kapcsolat
- 0.7 - 0.9 (-0.7 - -0.9) : magas korreláció, markáns kapcsolat
- 0.9 - 1 (-1 - -0.9) : nagyon magas korreláció, erős függő kapcsolat

Korreláció

A korrelációs együttható kiszámítása a populációból:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

A korrelációs együttható kiszámítása a mintából:

$$r_{x,y} = \frac{s_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Karl Pearson nevéhez köthető, aki Francis Galton tanítványa volt, az ő emlékére Pearson-féle r-nek, Pearson-féle korrelációnak szokás nevezni.

A korrelációnál is lehet vizsgálni a nullától való eltérés valószínűségét a populációban. Az erre vonatkozó t-statisztika:

$$t = r \sqrt{\frac{N-1}{1-r^2}} \quad f=N-2$$

Korreláció

A korreláció különösen fontos a megbízhatóság és az érvényesség meghatározásában.

Az érvényesség (validitás) tulajdonképpen azt jelenti, hogy valóban azt méri-e a teszt, amit hivatott mérni. Konkurens validitás: Ennek ellenőrzésére, a mért értékek korrelációját egy referencia értékkel vizsgálják.

Pl. új intelligencia teszten elért eredményeket, már gyakorlatban használt IQ teszt eredményeivel korreláltatják. A korrelációs érték leggyakrabban 0 és .6 között van, de leginkább ennek alsó tartományában.

A megbízhatóság (reliabilitás), tulajdonképpen megismételhetőség. A tapasztalati reliabilitás a tesztérték saját magával vett korrelációja, ismételt mérés esetén. Ez általában sokkal magasabb, mint az érvényességi együttható értéke. Általában .9-es értéket követelnek meg. De számos fontos teszt .8-nál is jóval alacsonyabb értékeket produkál és ennek ellenére jól használható a gyakorlatban.