

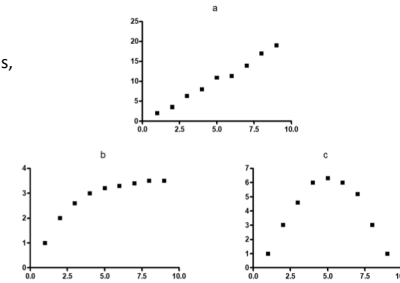
Orvosi statisztika, informatika és telemedicina

7. előadás:
Korreláció, egyszerű lineáris regresszió
2021. október 20.
Agócs Gergely

Változók közötti reláció (kapcsolat)

A kapcsolat típusa lehet:

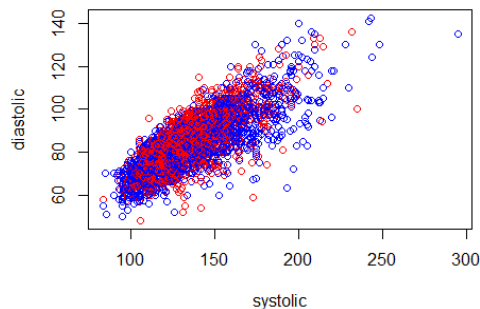
- monoton
 - pozitív (ha x nő, akkor y is nő) vagy negatív (ha x nő, akkor y csökken)
 - lineáris vagy nem lineáris (exponenciális, logos, hatványos)
- nem monoton
 - parabolikus
 - szinuszos
- nincs kapcsolat



2

Korreláció

- **monoton**,
 - **szimmetrikus** (nem megmondható, hogy melyik függne melyiktől)
- kapcsolat **két véletlen** (véletlen hibájú, „nem beállított”) változó között.



3

Korreláció

Korreláció **erősségének** kifejezésére:

Korrelációs együtthatók (r):

ha **lineáris korrelációt** feltételezünk: **Pearson r**

ha **monoton** (nem feltétlenül lineáris): **Spearman rang r**,

A korrelációs együtthatók **értéke**:

-1 től +1

negatív: negatív korreláció

pozitív: pozitív korreláció

minél közelebb esik |1|-hez annál erősebb a korreláció



$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{s_{xy}}{s_x s_y}$$

„Középtől vett távolság” –
mint y és mind x irányban

4

Student t-próba a korrelációs együttható (vagy a meredekség) szignifikanciájára

Mire vagyok kíváncsi

2 változó (lineárisan) korrelált-e (r értéke különbözik 0-tól)

Változó típusa

2 számszerű változó (X és Y)

Feltételek

Egymástól független megfigyelések (x és y párok)

Szimmetrikus, lineáris kapcsolatot feltételezünk

x és y véletlen változók

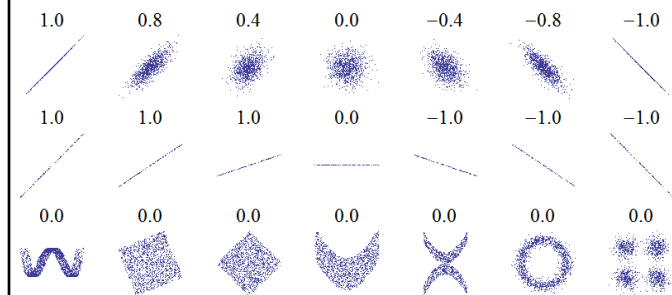
Megjegyzés:

A próbában a H_0 : **korrelációs együttható = 0** -t tesztljük.

5

Megjegyzések

- Mindig ábrázoljunk!
- Korreláció nem jelent okozati összefüggést
eg: <http://www.fastcodesign.com/3030529/infographic-of-the-day/hilarious-graphs-prove-that-correlation-isnt-causation>



Regresszió

Függvény kapcsolat (NEM szimmetrikus) egy függő (cél, eredmény, Y) változó és egy független (magyarázó, prediktáló, X) változó(k) között. [Y véletlen változó, X nem feltétlenül]

Y függ X-től – a függőségi viszony iránya **klínikailag** feltételezett, statisztikailag nem vizsgálható.

Kapcsolódó kérdések:

- Van (adott típusú) függvényszerű kapcsolat? (statisztikai kapcsolat, nem ok-okozati)
- Mekkora Y értéke, ha X:....? (becslés)
- Mekkora X értéke, ha Y:....? (becslés)
- Milyen függvény írja le legjobban Y X-től való függését?

7

Lineáris regresszió

Lineáris függvénykapcsolat feltételezett.

2 változó esetében a regresszió – korreláció kérdései, számolásai legtöbbször egymással „ekvivalensé” tehetőek.

8

Lineáris regresszió

Az egyenes becslésére: OLS (Ordinary Least Square method – azaz legkisebb négyzetek módszere)

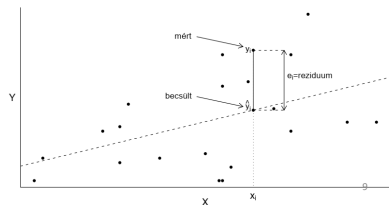
tengelymetszet

meredekség

Lineáris függvény: $Y = \beta_0 + \beta_1 * X + \epsilon$

Hibatag (reziduum, maradvány): pont-egyenes függőleges távolsága (a becsült és mért értékek különbsége)

Az OLS szerinti legjobb egyenes az, ahol a **legkisebb** a pont-egyenes függőleges távolságok **négyzetösszege**.



Student t-próba a korrelációs együttható (vagy a meredekség) szignifikanciájára

Mire vagyok kíváncsi

Y változó X-től lineárisan függ-e.

Változó típusa

2 számszerű változó (X és Y)

Feltételek

Egymástól független megfigyelések (x és y párok)

Lineáris kapcsolatot feltételezünk

x értékei „hiba nélkül” mérhetőek

a reziduumok eloszlása minden x-re normál

és variációjuk minden x-re megegyezik

Megjegyzés:

A próbában a H_0 : **meredekség = 0** -t teszteljük.

10

Meredekség és R^2

Meredekség

Az Y-ban bekövetkező átlagos változás X egységnyi változtatására

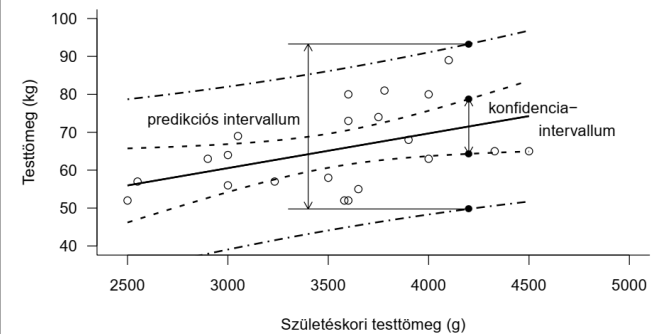
R^2 – meghatározottsági (determinációs) együttható

az r négyzete

Az Y változó variációjának (változatosságának) hány százaléka magyarázható az X variációjával (változásával)

11

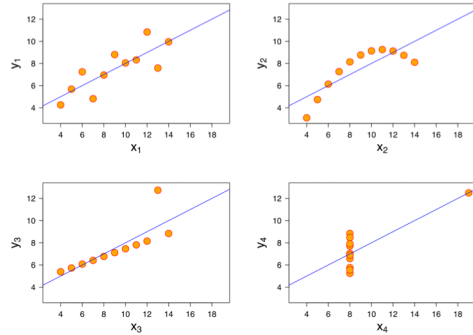
Konfidencia és predikciós intervallumok



12

Anscombe-féle négyes (1973)

- nagyon eltérő grafikus megjelenés
- azonos paraméterek (átlag, szórás, korrelációs együttható, egyenes egyenlete)



13