

Korreláció

statisztika1

Felhasznált irodalmak

- Balázs Katalin órai diasorai
- Field, A., Miles, J., & Field, Z. (2012).
Discovering Statistics Using R. SAGE
Publications.

Bevezetés

- Matektanulás
 - Többet gyakorolok -> jobb jegyet kapok (pozitív kapcsolat)
 - Többet gyakorolok-> jegyem változatlan (függetlenség)
 - Többet gyakorolok-> rosszabb jegyet kapok (negatív kapcsolat)

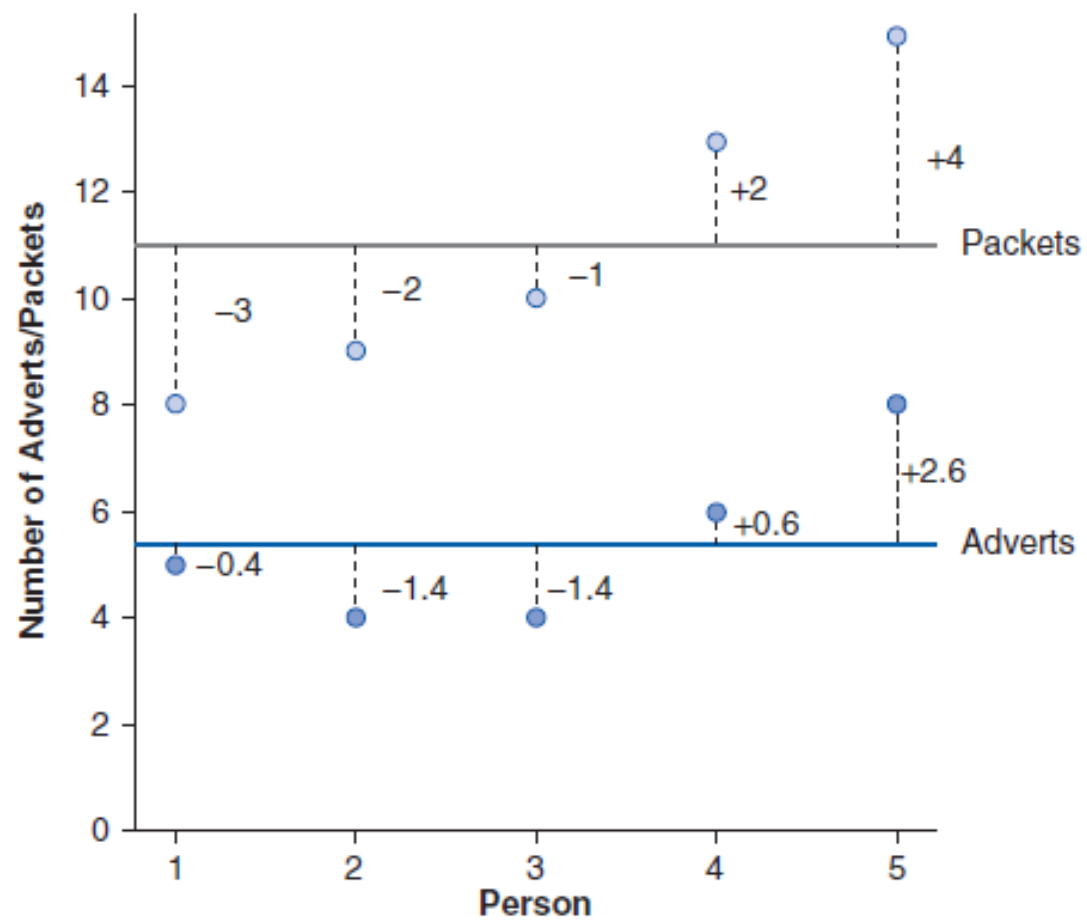
Hogyan mérhető 2 változó kapcsolata?

- variancia

$$\frac{\sum (x_i - \bar{x})^2}{N - 1} = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{N - 1}$$

- Egyik változóban változás milyen hatással van a másikra?

személy	1	2	3	4	5	átlag	szórás
reklám	5	4	4	6	8	5,4	1,67
csoki	8	9	10	13	15	11	2,92



$$\text{cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

- Sztenderdizálás
- Korreláció

Korrelációs együttható

$$r = \frac{\text{COV}_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(N - 1)s_x s_y}$$

t-statisztika a szign. szinthez

$$t_r = \frac{r\sqrt{N - 2}}{\sqrt{1 - r^2}}$$

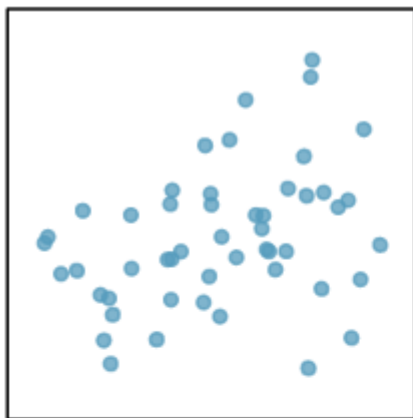
Korrelációs együttható

- A kovariancia a változók közötti kapcsolatot „durva” mérőszáma
- Ha sztenderdizáljuk az értéket, Pearson-féle korrelációs együtthatót kapunk
- -1 és $+1$ közé esik
- $+1$: tökéletes pozitív kapcsolat
- 0 : nincs lineáris kapcsolat
- -1 : tökéletes negatív kapcsolat

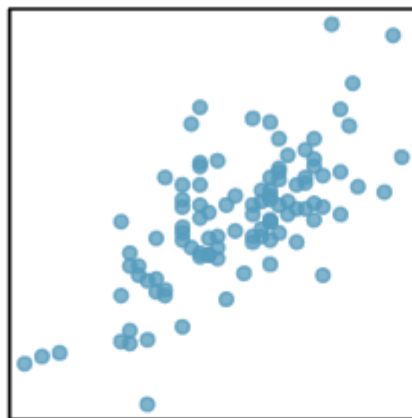
- Karl Pearson



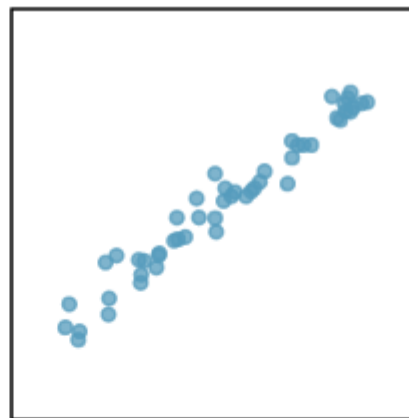
- Kapcsolat erőssége:
 - 0 : nincs lineáris kapcsolat
 - 0 - 0.2 (-0.2 - 0) : gyenge, majdnem hanyagolható kapcsolat
 - 0.2 - 0.4 (-0.4 - -0.2) : biztos, de gyenge kapcsolat
 - 0.4 - 0.7 (-0.7 - -0.4) : közepes korreláció, jelentős kapcsolat
 - 0.7 - 0.9 (-0.7 - -0.9) : magas korreláció, markáns kapcsolat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



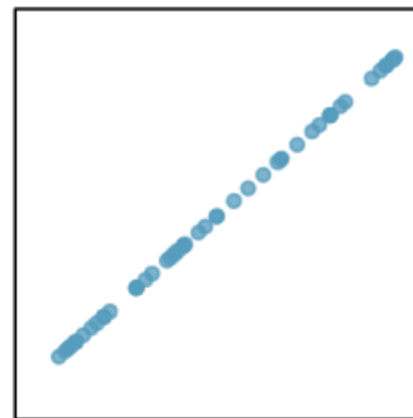
$R = 0.33$



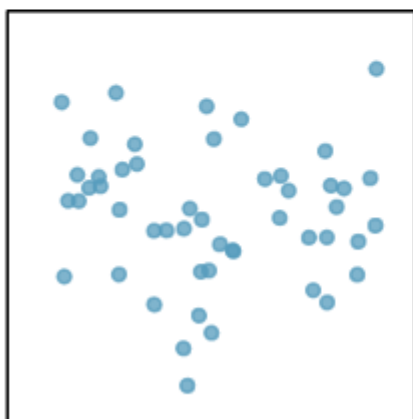
$R = 0.69$



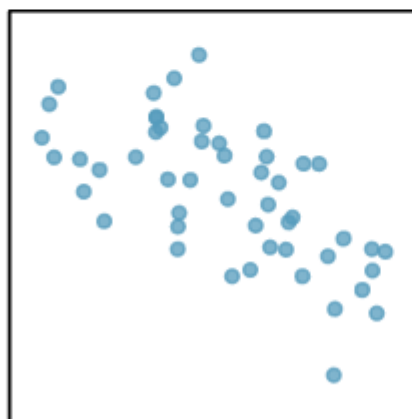
$R = 0.98$



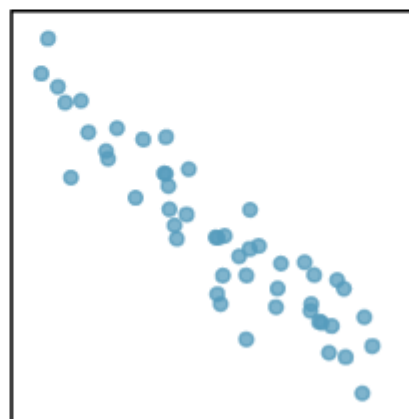
$R = 1.00$



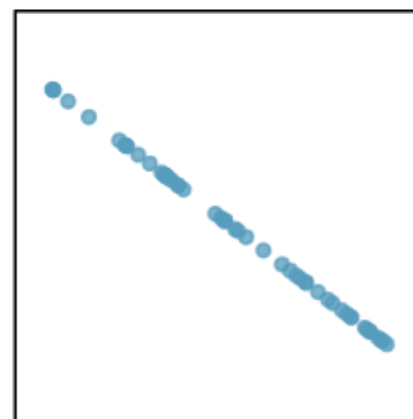
$R = -0.08$



$R = -0.64$



$R = -0.92$



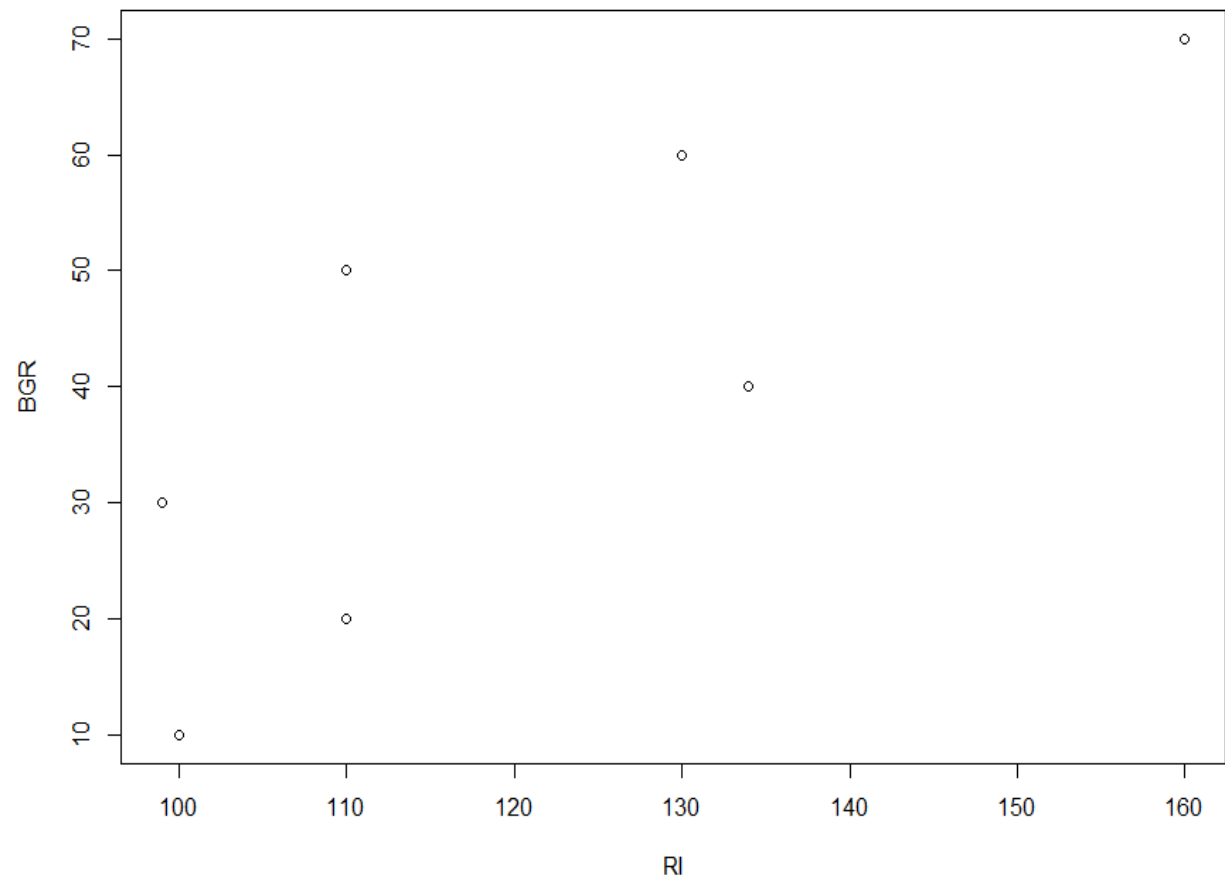
$R = -1.00$

- Harmadik változó?!
- Kölcsönös kapcsolat. Oksági kapcsolat nem vizsgálható vele.

feltételek

- Normál eloszlású változók
- Lineáris kapcsolat

- `plot(BGR~RI)`



- `cor.test(BGR,RI)`

Pearson's product-moment correlation

data: BGR and RI


`t = 3.0416, df = 5, p-value = 0.0287` 

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

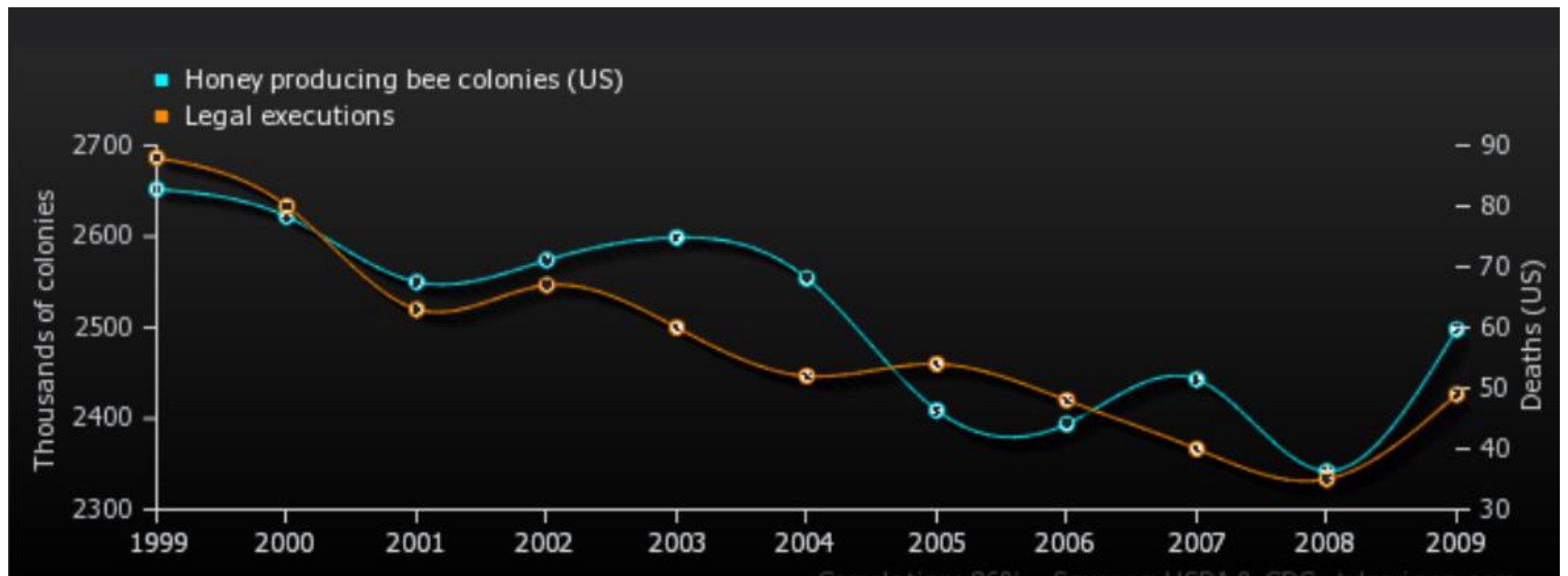
0.1338740 0.9701386

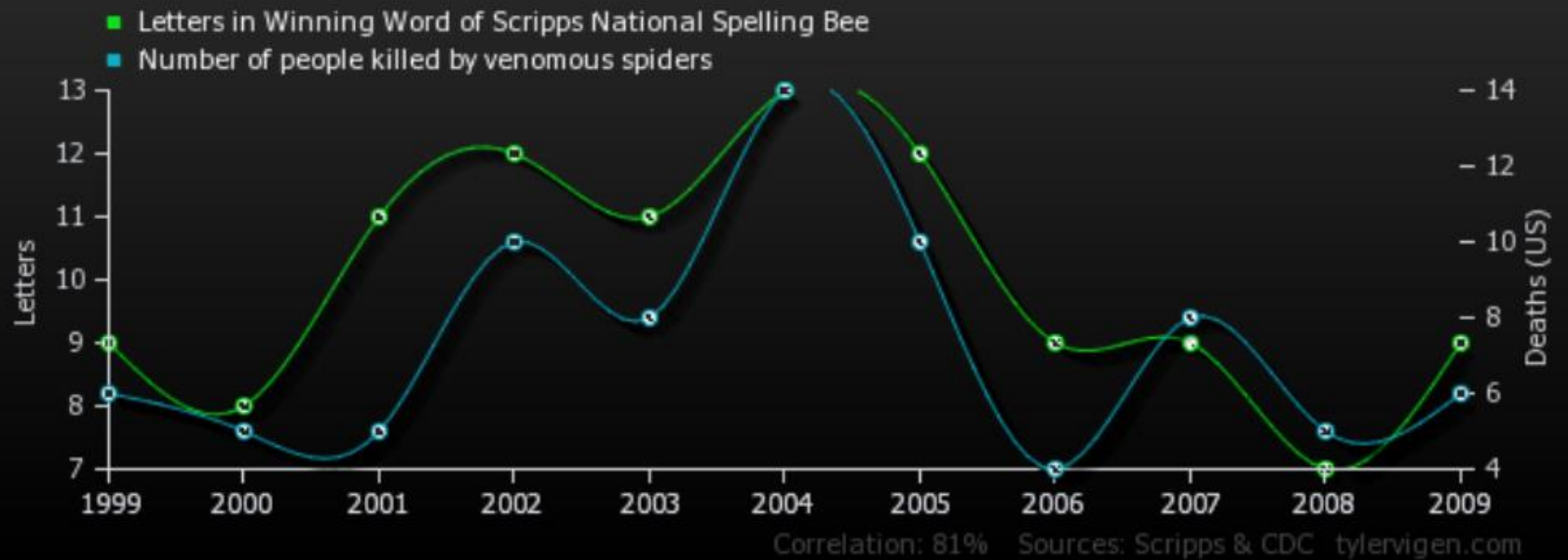
sample estimates:

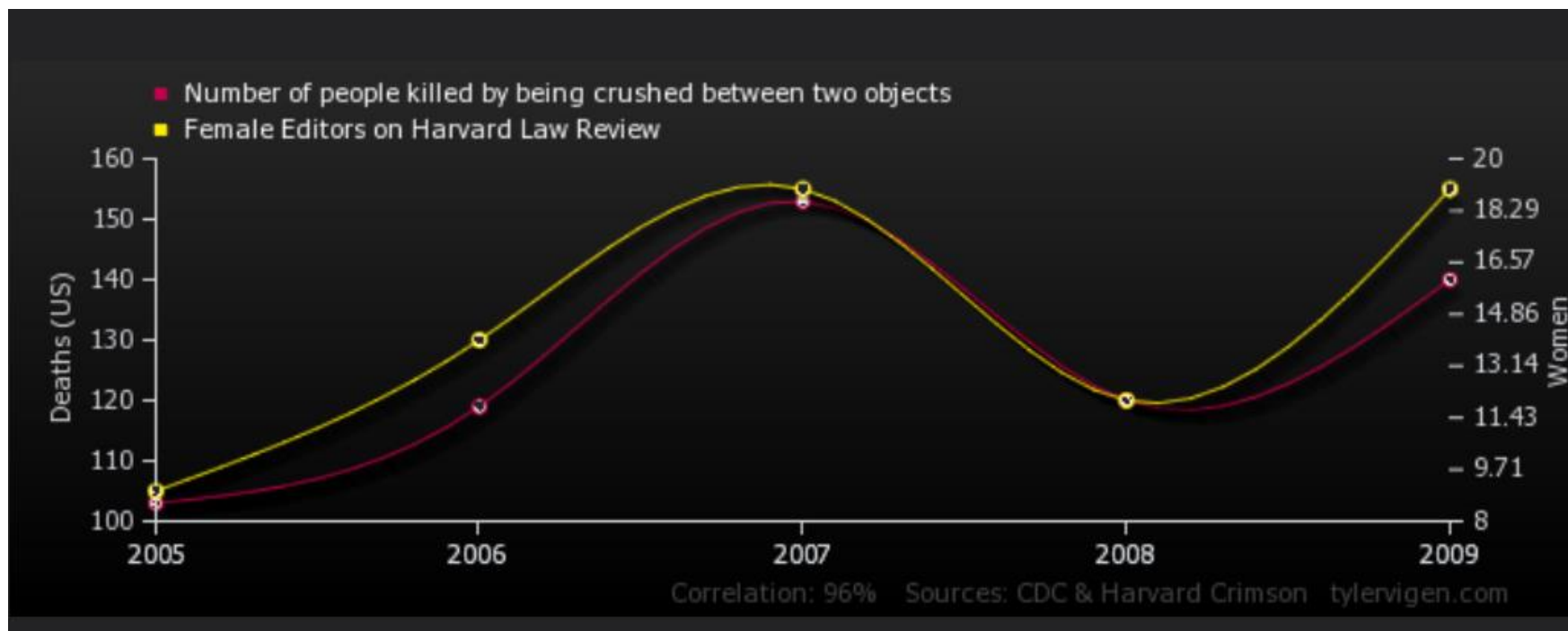
cor
0.805705 

- 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 együtttható 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.







The graph displays two data series over a ten-year period. The red line represents the number of deaths by falling into swimming pools, and the yellow line represents the number of films Nicolas Cage appeared in. Both series show a strong positive correlation, with a correlation coefficient of 0.77. The data points are as follows:

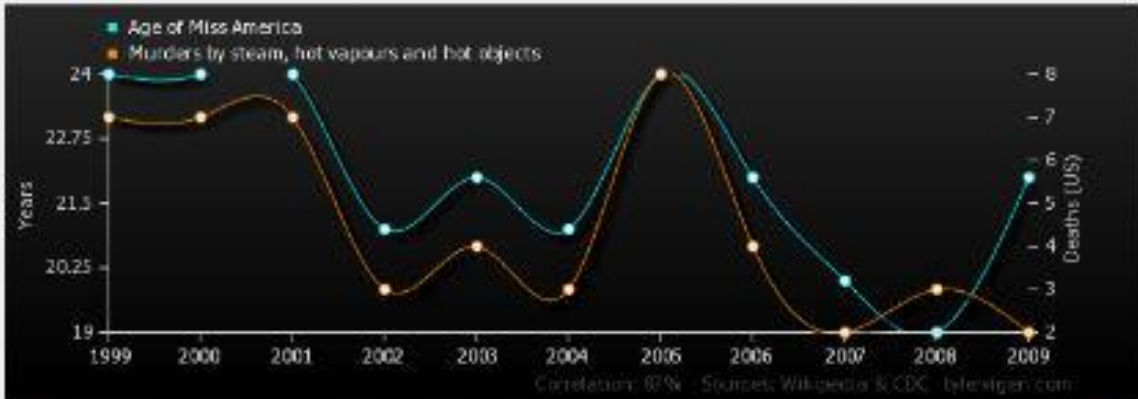
Year	Deaths (US)	Films
1999	106	93
2000	104	93
2001	104	91
2002	103	114
2003	89	80
2004	93	80
2005	94	94
2006	93	114
2007	118	130
2008	92	80
2009	104	130

Correlation: 0.77 Sources: CDC & IMDb tylerjvigen.com

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Number people who drowned by falling into a swimming-pool Deaths (US) (CDC)	109	102	102	98	85	95	96	98	123	94	102
Number of films Nicolas Cage appeared in Films (IMDB)	2	2	2	3	1	1	2	3	4	1	4

Correlation: 0.666004

Murders by steam, hot vapours and hot objects



	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Age of Miss America Years (Wikipedia)	24	24	24	21	22	21	24	22	20	19	22
Murders by steam, hot vapours and hot objects Deaths (US) (CDC)	7	7	7	3	4	3	8	4	2	3	2

Correlation: 0.870127