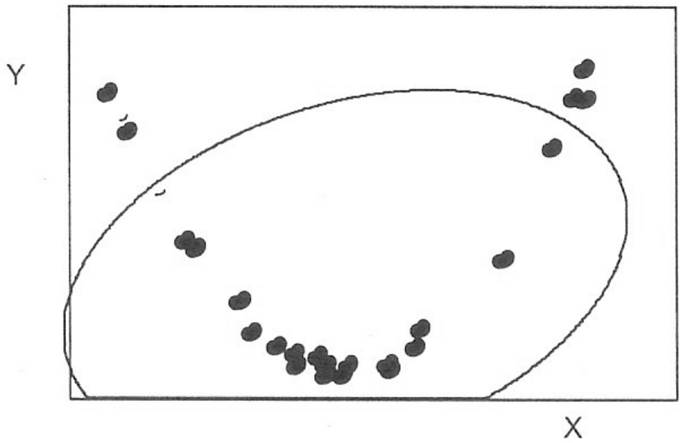
KORELACE

- Korelační analýza se zabývá mírou závislosti náhodných dat. Standardním výstupem korelační analýzy je koeficient popisující míru závislosti – nejčastěji korelační koeficient.
- Korelační koeficienty slouží jako míry vyjádření "těsnosti lineární vazby". Korelační analýza popisuje *lineární* vztahy mezi veličinami.

NEKORELOVANOST NEZNAMENÁ NEZÁVISLOST!



R = 0,193

PÁROVÝ (PEARSONŮV) KORELAČNÍ KOEFICIENT (R NEBO R_{XY})

$$R = \frac{\sum (x_i - \overline{x}) \cdot (y_i - \overline{y})}{(n-1)\, s_x\, s_y}$$

$$R = \frac{\sum (x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \cdot \sum (y_i - \overline{y})^2}}$$

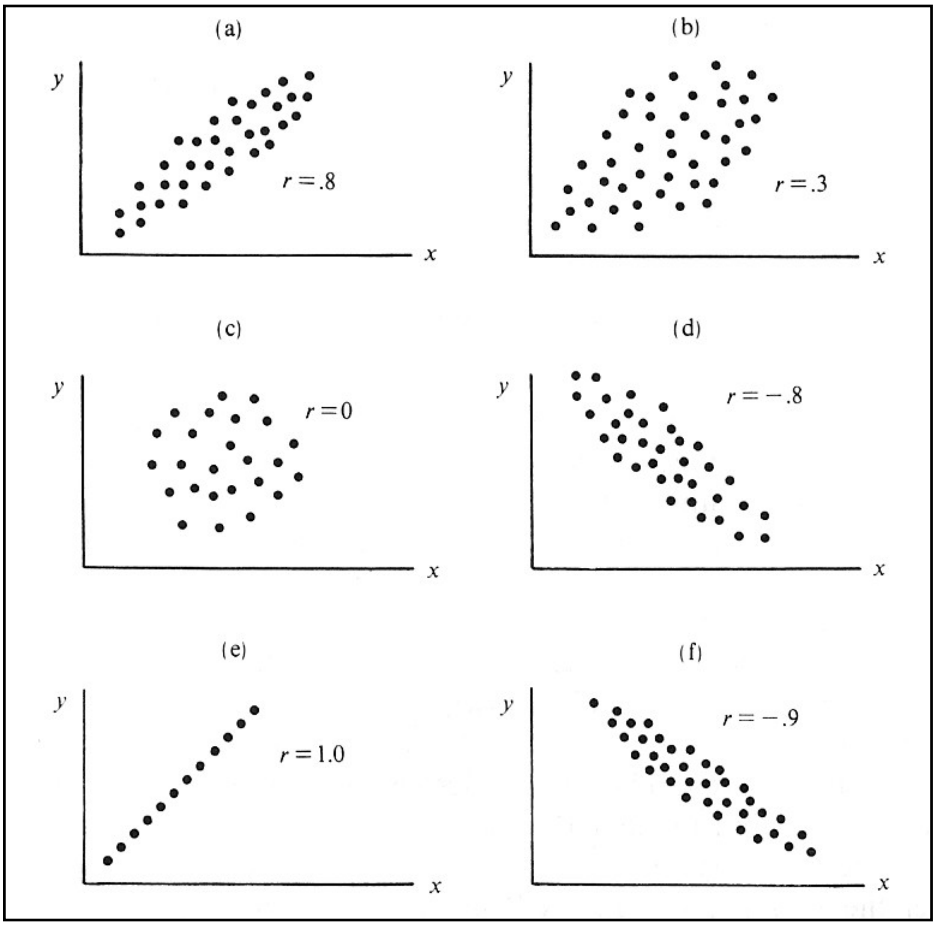
Korelační koeficient R může nabývat hodnot od –1 do +1.
Druhá mocnina korelačního koeficientu R^2 se nazývá koeficient determinace a nabývá hodnot od 0 do +1.

Kovariance (cov) – popisuje míru společné variability proměnných x a y. Pokud x a y nejsou v žádném vztahu, jejich kovariance je nulová.

$$cov(x, y) = \frac{\sum (x_i - \bar{x}).(y_i - \bar{y})}{n}$$

Párové korelační koeficienty se používají v jednoduché LR pro vyjádření korelace mezi x a y. Dále se používají ve vícenásobné LR a to pro vyjádření korelace mezi:

1. jednotlivými závisle proměnnými x_i a nezávisle proměnnou, tedy např.: x_2 a y, či x_3 a y,
2. vzájemně mezi závislými proměnnými, např.: x_1 a x_2, či x_1 a x_3.



Test významnosti korelačního koeficientu

H_0: R = 0; H_1: R ≠ 0

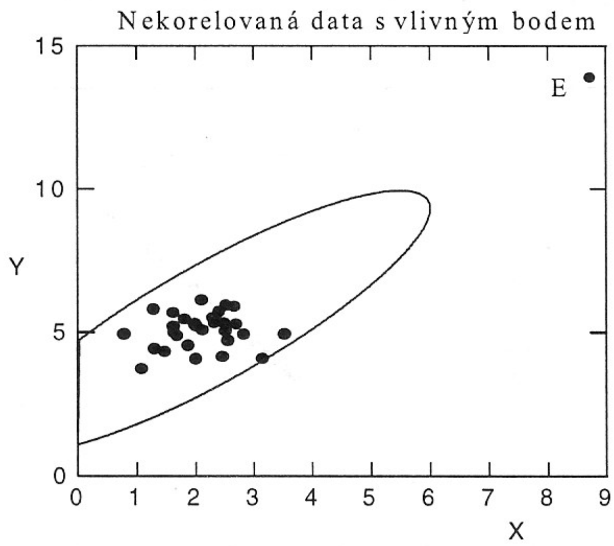
$$t = \frac{|R|\sqrt{n-2}}{\sqrt{1-R^2}}$$

testovací statistiku <u>t</u> porovnáváme s t_{krit}(0,975; n-2). Je-li <u>t</u> větší než kvantil Studentova rozdělení t_{krit}, zamítá se H_0.

Vliv extrémních hodnot na R

Extrémní bod ovlivňuje hodnotu R. Pokud je <u>E</u> správně změřený bod, který zvyšuje predikční rozsah (*golden point*), měli bychom se snažit získat takových bodů více. Je-li <u>E</u> chybný, zkresluje výsledek měření! <u>Při korelační analýze nestačí pouze vypočítat</u>

R a testovat jeho významnost. Měli bychom posoudit i graf závislosti mezi proměnnými x a y!



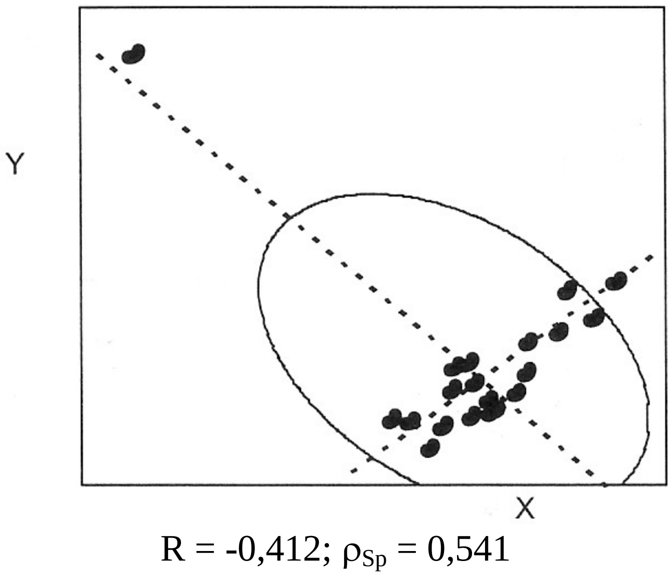
R = 0,863; ρ_{Sp} = 0,197

SPEARMANŮV (POŘADOVÝ) KORELAČNÍ KOEFICIENT ρ_{SP}

Je založen na srovnání pořadí hodnot. Tento koeficient je robustní. Jeho vztah ke klasickému korelačnímu koeficientu je analogií vztahu mediánu k aritmetickému průměru. Spearmanův koeficient není ovlivněn výskytem vlivných bodů.

$$\rho_{Sp} = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (P_{1i} - P_{2i})^2 \quad ,$$

kde P_{1i}, P_{2i} jsou čísla označující pořadí.



R = -0,412; ρ_{Sp} = 0,541

Korelační koeficienty v pokročilých regresních metodách

- **PARCIÁLNÍ KORELAČNÍ KOEFICIENT**

Popisují vztah mezi dvěma proměnnými x_i a x_j při zkonstantnění dalších proměnných.

Parciální korelační koeficient 1. řádu:

$$R_{1,2(3)} = \frac{R_{1,2} - R_{1,3} R_{2,3}}{\sqrt{(1 - R_{1,3}^2)(1 - R_{23}^2)}} \quad ,$$

kde $R_{1,2}$ je párový korelační koeficient mezi proměnnými 1 a 2,
$R_{1,3}$ je párový korelační koeficient mezi proměnnými 1 a 3,
$R_{2,3}$ je párový korelační koeficient mezi proměnnými 2 a 3,
$R_{1,2(3)}$ je korelační koeficient mezi proměnnými 1 a 2, je-li proměnná 3 konstantní.

Parciální korelační koeficient nultého řádu = párový korelační koeficient.

- **VÍCENÁSOBNÝ KORELAČNÍ KOEFICIENT** (používá se u vícenásobné LR)

$$R_{1(2,\ldots,m)} = \sqrt{1 - \frac{\det(\mathbf{R})}{\det(\mathbf{R}_{11})}} \quad ,$$

kde **det($\mathbf{R}_{ij}$)** označuje determinant matice vzniklé vypuštěním i-tého řádku a j-tého sloupce korelační matice **R** (matice párových korelačních koeficientů s 1 na hlavní diagonále).