


Fórum:Testy2/Pearsonův korelační koeficient

Pearsonův korelační koeficient měří sílu lineární závislosti mezi dvěma veličinami. Pomůže nám například vyčíslit, jak silná je vazba mezi výsledky ve dvou různých testech, nebo mezi výsledkem v testu a průměrnou známkou na vysvědčení.

 **Tip: Korelace neznamená kauzalitu**

Předpokládáme, že máme u n jedinců dvojice hodnot $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. Pearsonův korelační koeficient je pak dán vztahem

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

kde $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ je aritmetický průměr prvních měření a \bar{Y} je aritmetický průměr druhých měření.

Ekvivalentně lze Pearsonův korelační koeficient vyjádřit pomocí součinů z-skórů

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s_X} \right) \left(\frac{Y_i - \bar{Y}}{s_Y} \right)$$

kde $s_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ je směrodatná odchylka (standard deviation, SD) prvních měření, s_Y směrodatná

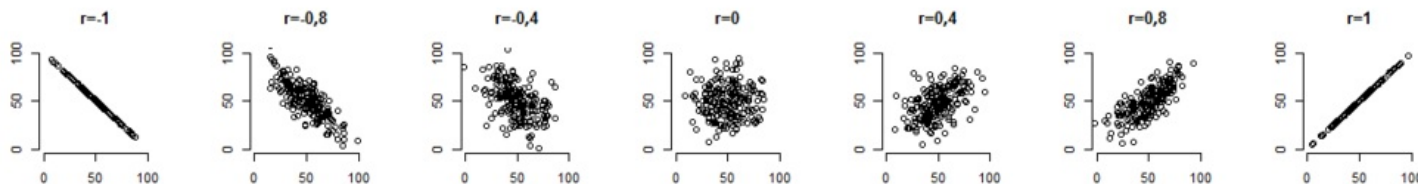
odchylka druhých měření. z-skór $\left(\frac{X_i - \bar{X}}{s_X} \right)$ říká, jak daleko je hodnota X_i od průměru \bar{X} , přičemž za jednotku vzdálenosti bereme směrodatnou odchylku s_X .

V případě, kdy jeden ze znaků je binární (např. sledujeme-li závislost binární položky Y a celkového počtu bodů X, tzv. index RIT), se korelační koeficient redukuje na tzv. *bodově-biseriální korelační koeficient*

$$r_{bis} = \frac{\bar{X}_1 - \bar{X}_0}{s} \frac{n_0 n_1}{n(n-1)},$$

kde \bar{X}_1 je průměr celkového počtu bodů těch studentů, kteří mají hodnotu sledované položky rovnou 1, \bar{X}_0 je průměr celkového počtu bodů těch studentů, kteří mají hodnotu sledované položky rovnou 0, s^2 je směrodatná odchylka celkových bodových zisků, n_1 je počet jedniček a n_0 je počet nul.

Korelační koeficient nabývá pouze hodnot z intervalu od -1 do 1. Svých extrémních hodnot (tedy 1 a -1) nabývá pouze pokud všechny body (X_i, Y_i) leží na jedné přímce. Korelační koeficient je roven 1, pokud je mezi veličinami vztah přímé úměry (tedy čím větší je hodnota jedné veličiny, tím větší je hodnota i druhé veličiny). Pokud je mezi veličinami vztah nepřímé úměry, je korelační koeficient roven -1.



Obr. 10.4 Korelační koeficient nabývající hodnot z intervalu od -1 do 1

Jsou-li veličiny nezávislé, je korelace mezi nimi nulová. Nicméně Pearsonův korelační koeficient je pouze odhad populačního korelačního koeficientu, a při každém výběru n -tice studentů vyjde hodnota odhadu nepatrně jiná. Proto i pro nezávislé náhodné veličiny zpravidla vyjde Pearsonův korelační koeficient nenulový. Nulovost pak můžeme testovat pomocí statistického testu (viz např. [2]).

1. SIES, Helmut. A new parameter for sex education. *Nature*. 1988, roč. -, vol. -, no. 332, s. 495, ISSN (Print) 0028-0836, (Online) 1476-4687. DOI: 10.1038/332495a0 (<http://dx.doi.org/10.1038%2F332495a0>).
2. ZVÁRA, Karel. *Základy statistiky v prostředí R*. 1. vydání. Praha : Karolinum, 2013. 249 s.

