

Vícerozměrná data

Sdružená (simultánní) absolutní četnost

$$n_{jk} = N(X = x_{[j]} \wedge Y = y_{[k]})$$

Sdružená (simultánní) relativní četnost

$$p_{jk} = \frac{n_{jk}}{n}$$

Marginální absolutní četnost varianty x_j

$$n_{j.} = N(X = x_{[j]}) = n_{j1} + \dots + n_{js}$$

Marginální relativní četnost varianty x_j

$$p_{j.} = \frac{n_{j.}}{n} = p_{j1} + \dots + p_{js}$$

Marginální absolutní četnost varianty y_j

$$n_{.k} = N(X = y_{[k]}) = n_{1k} + \dots + n_{rk}$$

Marginální relativní četnost varianty y_k

$$p_{.k} = \frac{n_{.k}}{n} = p_{1k} + \dots + p_{rk}$$

Kovariance

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

Výběrová kovariance

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

Pearsonův korelační koeficient

$$r_{xy} = \frac{1}{n} \sum_{i=1}^n \frac{x_i - \bar{x}}{s_{n,x}} \cdot \frac{y_i - \bar{y}}{s_{n,y}} = \frac{s_{n,xy}}{s_{n,x}s_{n,y}} = \frac{s_{xy}}{s_x s_y}$$

Spearmanův koeficient pořadové korelace

$$r_{xy}^s = 1 - \frac{6 \sum_{i=1}^n (p_i - q_i)^2}{n(n^2 - 1)}$$

Vícerozměrná náhodná veličina

Sdružená distribuční funkce vektoru $(X, Y)'$

$$F(x, y) = P(X \leq x, Y \leq y)$$

X a Y mají distribuční funkce

$$F_X(x) = F(x, \infty) \quad \text{a} \quad F_Y(y) = F(\infty, y).$$

Sdružená pravděpodobnostní funkce vektoru $(X, Y)'$

$$p(x, y) = P(X = x, Y = y)$$

Marginální pravděpodobnostní funkce X a Y jsou

$$p_X(x) = \sum_{y \in M_y} p(x, y), \quad x \in M_x,$$
$$p_Y(y) = \sum_{x \in M_x} p(x, y), \quad y \in M_y.$$

Sdružená hustota pravděpodobnosti vektoru $(X, Y)'$

$$P((X, Y) \in B) = \iint_B f(x, y) dx dy$$

Marginální hustoty X a Y jsou

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Náhodné veličiny X a Y jsou nezávislé právě tehdy, když

$$F(x, y) = F_X(x)F_Y(y).$$

Diskrétní náhodné veličiny jsou nezávislé, právě když

$$p(x, y) = p_X(x)p_Y(y),$$

spojité náhodné veličiny jsou nezávislé právě tehdy, když

$$f(x, y) = f_X(x)f_Y(y).$$

Střední hodnota vektoru $\mathbf{X} = (X, Y)'$

$$E(\mathbf{X}) = (E(X), E(Y))'$$

Kovariance

$$C(X, Y) = E[X - E(X)][Y - E(Y)] = E(XY) - E(X)E(Y)$$

Korelační koeficient

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{D(X)D(Y)}}$$

Test významnosti korelačního koeficientu

$H: \rho = 0 \rightarrow A: \rho \neq 0$

$$t = \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \sqrt{n - 2} \sim t(n - 2)$$

$W_\alpha: |t| \geq t_{1-\alpha/2}(n - 2)$

Test nezávislosti v kontingenční tabulce

$H: X$ a Y jsou nezávislé náhodné veličiny $\rightarrow A: X$ a Y jsou závislé náhodné veličiny

$$\chi^2 = \sum_{j=1}^r \sum_{k=1}^s \frac{(n_{jk} - o_{jk})^2}{o_{jk}}, \quad o_{jk} = \frac{n_{j.}n_{.k}}{n}$$

$W_\alpha: \chi^2 \geq \chi_{1-\alpha}^2(\nu), \nu = (r - 1)(s - 1)$