I. Analýza závislosti kvantitativních znaků

Jednoduchá korelace a regrese

funkce lineární (přímka):

$$y' = a + b \cdot x_i$$

Soustava normálních rovnic

$$na_{yx} + b_{yx} \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$a_{yx} \sum_{i=1}^{n} x_i + b_{yx} \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

Výpočtové vzorce

$b_{yx} = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$	$b_{xy} = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum y_i^2 - \left(\sum y_i\right)^2}$
$b_{yx} = \dfrac{\operatorname{cov} xy}{s_x^2}$	$b_{xy} = \dfrac{\operatorname{cov} xy}{s_y^2}$
$b_{yx} = r \cdot \dfrac{s_y}{s_x}$	$b_{xy} = r \cdot \dfrac{s_x}{s_y}$
$a_{yx} = \overline{y} - b_{yx} \cdot \overline{x}$	$a_{xy} = \overline{x} - b_{xy} \cdot \overline{y}$
$\operatorname{cov} xy = \overline{xy} - \overline{x} \cdot \overline{y}$	

Síla závislosti – korelační koeficient

$$r_{xy} = r_{yx} = \frac{\operatorname{cov} xy}{s_x \cdot s_y} \qquad r_{yx} = r_{xy} = \pm\sqrt{b_{yx} \cdot b_{xy}}$$

$$r_{yx} = r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right] \cdot \left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}}$$

Síla závislosti – Spearmanův koeficient pořadové korelace

$$r_S = 1 - \frac{6 \cdot \sum d_i^2}{n \cdot (n^2 - 1)},$$

kde d_i = rozdíly pořadí R_x a R_y

funkce logaritmická:

$$y' = a + b \cdot \log x_i$$

$$na + b\sum \log x_i = \sum y_i$$
$$a\sum \log x_i + b\sum (\log x_i)^2 = \sum (\log x_i) \cdot y_i$$

funkce kvadratická (parabola):

$$y' = a + b \cdot x_i + c \cdot x_i^2$$

$$na + b\sum x_i + c\sum x_i^2 = \sum y_i$$
$$a\sum x_i + b\sum x_i^2 + c\sum x_i^3 = \sum y_i x_i$$
$$a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 = \sum y_i x_i^2$$

funkce kubická:

$$y' = a + b \cdot x_i + c \cdot x_i^2 + d \cdot x_i^3$$

$$na + b\sum x_i + c\sum x_i^2 + d\sum x_i^3 = \sum y_i$$
$$a\sum x_i + b\sum x_i^2 + c\sum x_i^3 + d\sum x_i^4 = \sum x_i y_i$$
$$a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 + d\sum x_i^5 = \sum x_i^2 y_i$$
$$a\sum x_i^2 + b\sum x_i^4 + c\sum x_i^5 + d\sum x_i^6 = \sum x_i^3 y_i$$

funkce hyperbolická (funkce lomená):

$$y' = a + b \cdot \frac{1}{x_i}$$

$$na + b\sum \frac{1}{x_i} = \sum y_i$$
$$a\sum \frac{1}{x_i} + b\sum \frac{1}{x_i^2} = \sum \frac{y_i}{x_i}$$

funkce mocninná:

$$y' = a \cdot x_i^b$$

$$n \log a + b\sum \log x_i = \sum \log y_i$$
$$\log a \sum \log x_i + b\sum (\log x_i)^2 = \sum (\log y_i)(\log x_i)$$

funkce exponenciální:

$$y' = a \cdot b^{x_i}$$

$$n \log a + \log b \sum x_i = \sum \log y_i$$
$$\log a \sum x_i + \log b \sum x_i^2 = \sum x_i \log y_i$$

Index korelace

$$I_{yx} = \sqrt{\frac{\sum y_i'^2 - \frac{1}{n}(\sum y_i)^2}{\sum y_i^2 - \frac{1}{n}(\sum y_i)^2}} \quad , \qquad I_{yx} = \sqrt{1 - \frac{\sum (y_i - y_i')^2}{\sum (y_i - \bar{y})^2}}$$

$$\sum y_i'^2 = (\sum y_i) \cdot y_i'$$

$$I_{yx} = \sqrt{1 - \frac{s_\varepsilon^2}{s_y^2}} \qquad s_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2$$

Vícenásobná lineární regrese a korelace

Rovina: $y' = a + b_{yx_1 \cdot x_2} \cdot x_1 + b_{yx_2 \cdot x_1} \cdot x_2$

$$b_{yx_1 \cdot x_2} = \frac{b_{yx_1} - b_{yx_2} \cdot b_{x_2x_1}}{1 - b_{x_2x_1} \cdot b_{x_1x_2}} \qquad b_{yx_1 \cdot x_2} = \frac{s_y}{s_{x_1}} \cdot \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{1 - r_{x_1x_2}^2}$$

$$b_{yx_2 \cdot x_1} = \frac{b_{yx_2} - b_{yx_1} \cdot b_{x_1x_2}}{1 - b_{x_2x_1} \cdot b_{x_1x_2}} \qquad b_{yx_2x_1} = \frac{s_y}{s_{x_2}} \cdot \frac{r_{yx_2} - r_{x_1x_2} \cdot r_{yx_1}}{1 - r_{x_1x_2}^2}$$

$$a = \bar{y} - b_{yx_1 \cdot x_2} \cdot \bar{x}_1 - b_{yx_2 \cdot x_1} \cdot \bar{x}_2$$

Síla závislosti

a) úplný korelační koeficient

$$R_{y.x_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 \cdot r_{x_1x_2} \cdot r_{yx_1} \cdot r_{yx_2}}{1 - r_{x_1x_2}^2}}$$

b) parciální korelační koeficienty

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}} \qquad r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)}}$$

$\beta -$ koeficienty

$$\beta_{yx_1x_2} = b_{yx_1 \cdot x_2} \cdot \frac{s_{x_1}}{s_y} \qquad \beta_{yx_2x_1} = b_{yx_2 \cdot x_1} \cdot \frac{s_{x_2}}{s_y}$$

Maticový počet

Regresní funkce

$$\vec{y} = X\vec{b} + \vec{\varepsilon}$$

Soustavu normálních rovnic lze zapsat

$$X^T \cdot \vec{y} = X^T \cdot X \cdot \vec{b}$$

Index korelace

$$I_{yx} = \sqrt{\frac{\vec{b}^T \cdot X^T \cdot \vec{y} - \frac{1}{n}(\Sigma y_i)^2}{\vec{y}^T \cdot \vec{y} - \frac{1}{n}(\Sigma y_i)^2}}$$

Testování výběrových charakteristik korelace a regrese

Test korelačního koeficientu

$$t_r = \frac{|r_{yx}|}{\sqrt{\frac{1 - r_{yx}^2}{n - 2}}}$$

t_r má Studentovo t-rozdělení pro (n – 2) stupně volnosti

Test dílčího korelačního koeficientu

$$t = \frac{r_{yx_1 \cdot x_2 \ldots\ldots x_k}}{\sqrt{\frac{1 - r^2_{yx_1 \cdot x_2 \ldots\ldots x_k}}{n - k - 1}}}$$

k ..... počet nezávisle proměnných

Testovací kritérium má Studentovo t-rozdělení pro (n – k – 1) stupňů volnosti.

Test totálního korelačního koeficientu

$$F = \frac{\frac{R^2_{y \cdot x_1 x_2 \ldots\ldots x_k}}{k}}{\frac{1 - R^2_{y \cdot x_1 x_2 \ldots\ldots x_k}}{n - k - 1}}$$

Testovací kritérium má F-rozdělení pro počet stupňů volnosti (k; n – k – 1), kde k je počet nezávisle proměnných.

Test regresního koeficientu

$$t_b = \frac{|b_{yx}|}{s_{b_{yx}}} \qquad s_{b_{yx}} = \frac{s_y}{s_x} \cdot \sqrt{\frac{1-r_{yx}^2}{n-2}}$$

$$t_b = \frac{|b_{xy}|}{s_{b_{xy}}} \qquad s_{b_{xy}} = \frac{s_x}{s_y} \cdot \sqrt{\frac{1-r_{yx}^2}{n-2}}$$

t_b má Studentovo t-rozdělení pro (n – 2) stupně volnosti

Test o dílčím regresním koeficientu

$$t = \frac{b_{yx_1 \cdot x_2 \ldots\ldots x_k}}{s_{b_{yx_1 \cdot x_2 \ldots x_k}}} \qquad s_{b_{yx_1 \cdot x_2 \ldots\ldots x_k}} = \frac{s_y}{s_x} \cdot \frac{1-R^2_{y \cdot x_1 x_2 \ldots x_k}}{(1-R^2_{y \cdot x_1 x_2 \ldots\ldots x_k})(n-k-1)}$$

Testovací kritérium má Studentovo t-rozdělení pro počet stupňů volnosti (n – k – 1).

Testování regresní funkce pomocí analýzy rozptylu

Variabilita	Součet čtverců	Stupně volnosti	Rozptyl	Testovací statistika
regrese	$S_1 = \sum_{i=1}^{n} (y_i' - \bar{y})^2$	p – 1	$s_1^2 = \frac{S_1}{p-1}$	$F = \frac{s_1^2}{s_r^2}$
kolem regrese	$S_r = \sum_{i=1}^{n} (y_i - y_i')^2$	n – p	$s_r^2 = \frac{S_r}{n-p}$	

p = počet parametrů regresní funkce

Statistika F má Snedecorovo F-rozdělení o $[(p-1),(n-p)]$ stupních volnosti.

Bodový odhad korelačního koeficientu $\hat{\rho}_{yx}$

$$\hat{\rho}_{yx} = \sqrt{1-(1-r)^2 \cdot \frac{n-1}{n-2}}$$

Intervalový odhad korelačního koeficientu ρ_{yx}

je-li n < 100, potom Fisherova z-transformace

$$z \in (z_r \pm u_\alpha \cdot s_z)$$

$$s_z = \frac{1}{\sqrt{n-3}}$$

je-li n > 100, potom $r_{yx} \pm u_\alpha \cdot s_r$

$$s_r = \frac{1-r_{yx}^2}{\sqrt{n-2}}$$

Interval spolehlivosti pro dílčí korelační koeficient

$$r_{yx_1 \cdot x_2 \dots x_k} \pm u_\alpha \cdot s_z$$

$$s_z = \frac{1}{\sqrt{n-k-2}}$$

Intervalový odhad regresního koeficientu β_{yx}:

$$b_{yx} \pm t_{\alpha(n-2)} \cdot s_{b_{yx}}, \qquad \text{kde } s_{byx} = \frac{s_y}{s_x} \cdot \sqrt{\frac{1-r_{yx}^2}{n-2}}$$

$$b_{xy} \pm t_{\alpha(n-2)} \cdot s_{b_{xy}}, \qquad \text{kde } s_{bxy} = \frac{s_x}{s_y} \cdot \sqrt{\frac{1-r_{yx}^2}{n-2}}$$

Interval spolehlivosti pro dílčí regresní koeficient $\beta_{yx_1.x_2 \dots x_k}$

$$b_{yx_1 \cdot x_2 \dots x_k} \pm t_{\alpha(n-k-1)} \cdot s_b, \qquad \text{kde } s_{b_{yx_1.x_2 \dots x_k}} = \frac{s_y}{s_x} \cdot \frac{1-R^2_{y \cdot x_1 x_2 \dots x_k}}{(1-R^2_{y \cdot x_1 x_2 \dots x_k})(n-k-1)}$$

II. Analýza závislosti kvalitativních znaků

1. Asociace (2 x 2)

Průběh závislosti (asociační přímka)

$$\frac{(a+c)'}{n} = A + B\frac{(a+b)}{n}$$

$$B = \frac{n(a)-(a+b)\cdot(a+c)}{(a+b)\cdot(c+d)}$$

$$A = \frac{(a+c)}{n} - B \cdot \frac{(a+b)}{n}$$

Síla závislosti

Koeficient asociace $$V = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Yuleův koeficient asociace $$Q = \frac{ad-bc}{ad+bc}$$

Koeficient koligace $$Y = \frac{1-\sqrt{\frac{bc}{ad}}}{1+\sqrt{\frac{bc}{ad}}}$$

Testování v asociační tabulce

χ^2- test

$$\chi^2 = \frac{n\left(ad-bc\right)^2}{\left(a+b\right)\cdot\left(c+d\right)\cdot\left(a+c\right)\cdot\left(b+d\right)}$$

Statistika χ^2 má χ^2 -rozdělení pro [(2 – 1) · (2 – 1)] = 1 stupeň volnosti.

Fisherův test

$$p_i = \frac{\left(a+b\right)!\,\left(c+d\right)!\,\left(a+c\right)!\,\left(b+d\right)!}{n!\,a!\,b!\,c!\,d!}$$

testové kritérium $p = \sum p_i$

Hodnota p se porovnává se zvolenou hladinou významnosti α.

2. Kontingence (r x s)

Síla závislosti

Pearsonův koeficient kontingence $$C = \sqrt{\frac{\chi^2}{n+\chi^2}}$$

Normovaný koeficient kontingence $$C_n = \frac{C}{C_{max}}, \quad C_{max} = \sqrt{\frac{h-1}{h}}, \quad \text{kde } h = \min(r,s)$$

Čuprovův koeficient kontingence $$K = \sqrt{\frac{\chi^2}{n\sqrt{(r-1)(s-1)}}}$$

Cramérův koeficient kontingence $$V = \sqrt{\frac{\chi^2}{n\left(h-1\right)}}$$

Testování kontingenční tabulky

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij}-o_{ij}\right)^2}{o_{ij}}, \qquad \text{kde } o_{ij} = \frac{n_{i\bullet}\cdot n_{\bullet j}}{n}$$

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{n \cdot (n_{ij})^2}{n_{i\bullet} \cdot n_{\bullet j}} - n$$

Testové kritérium má χ^2-rozdělení pro $[(r-1)\cdot(s-1)]$ stupeň volnosti.

III. Analýza časových řad

Vyrovnání časové řady přímkou

$$u_i = a + b \cdot t_i$$

jestliže $\sum \mathbf{t_i} \neq \mathbf{0}$

$$an + b\sum t_i = \sum y_i$$
$$a\sum t_i + b\sum t_i^2 = \sum t_i y_i$$

jestliže $\sum \mathbf{t_i} = \mathbf{0}$

$$a = \frac{\sum y_i}{n}, \qquad b = \frac{\sum t_i y_i}{\sum t_i^2}$$

Další typy trendových funkcí viz. kapitola III. Závislost kvantitativních znaků.

Index korelace

$$I = \sqrt{1 - \frac{\sum (y_i - u_i)^2}{\sum (y_i - \bar{y})^2}} = \sqrt{1 - \frac{s_\varepsilon^2}{s_y^2}}, \qquad \text{kde} \quad \varepsilon_i = y_i - u_i, \qquad s_\varepsilon^2 = \frac{\sum (y_i - u_i)^2}{n}$$

Náhodné kolísání

absolutní průměrná odchylka v časových řadách s trendem

$$\bar{d} = \sum \frac{|y_i - u_i|}{n}$$

relativní průměrná odchylka v časových řadách s trendem

$$\bar{d}' = \frac{\sum \frac{|y_i - u_i|}{u_i}}{n}$$

absolutní průměrná odchylka ve stacionárních časových řadách

$$\bar{d} = \sum \frac{|y_i - \bar{y}|}{n}$$

relativní průměrná odchylka ve stacionárních časových řadách

$$\bar{d}' = \frac{\sum \frac{|y_i - \bar{y}|}{\bar{y}}}{n}$$

Intervalová predikce

$$u_{n+1} - t_{\alpha(n-2)} \cdot s'_{y_{n+1}} \leq y'_{n+1} \leq u_{n+1} + t_{\alpha(n-2)} \cdot s'_{y_{n+1}},$$

kde u_{n+1} ... bodová předpověď na období (n + 1)

$s'_{y_{n+1}}$ ... směrodatná chyba předpovídané hodnoty

$s'_{y_{n+1}} = s_y \cdot \sqrt{(1 - I^2) \cdot \frac{n(n^2 - 1) + 12\,k^2}{(n^2 - 1)(n - 2)}}$, kde k je počet kroků dopředu

IV. Indexní analýza

1. Individuální indexy

$$\frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{\sum p_0 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} \cdot \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_1}{\sum q_1}}$$

$$\frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_0 q_0}{\sum q_0}} = \frac{\frac{\sum p_1 q_0}{\sum q_0}}{\frac{\sum p_0 q_0}{\sum q_0}} \cdot \frac{\frac{\sum p_1 q_1}{\sum q_1}}{\frac{\sum p_1 q_0}{\sum q_0}}$$

2. Souhrnné indexy

a) indexy množství

$$ {}_{c_0}I^{(q)}_{1/0} = \frac{\sum q_1 c_0}{\sum q_0 c_0}; \qquad \Delta\left[{}_{c_0}I^{(q)}_{1/0}\right] = \sum q_1 c_0 - \sum q_0 c_0 $$

$$ {}_{c_1}I^{(q)}_{1/0} = \frac{\sum q_1 c_1}{\sum q_0 c_1}; \qquad \Delta\left[{}_{c_1}I^{(q)}_{1/0}\right] = \sum q_1 c_1 - \sum q_0 c_1 $$

b) indexy úrovně

Laspeyresův index $\quad {}_{q_0}I^{(c)}_{1/0} = \frac{\sum q_0 c_1}{\sum q_0 c_0}$

Paascheho index $\quad {}_{q_1}I^{(c)}_{1/0} = \frac{\sum q_1 c_1}{\sum q_1 c_0}$

Fisherův index $\quad {}_{q_0 q_1}I^{(c)}_{1/0} = \sqrt{\frac{\sum q_0 c_1}{\sum q_0 c_0} \cdot \frac{\sum q_1 c_1}{\sum q_1 c_0}}$

Loweho index $\quad {}_{q}I^{(c)}_{1/0} = \frac{\sum q.c_1}{\sum q c_0},$ $\qquad$ kde q je stálé množství

rozklad indexů dvoufaktorových ukazatelů

$$ \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_1 p_1}{\sum q_1 p_0} \cdot \frac{\sum q_1 p_0}{\sum q_0 p_0} $$

$$ \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_1 p_1}{\sum q_0 p_1} \cdot \frac{\sum q_0 p_1}{\sum q_0 p_0} $$

rozklad indexů třífaktorových ukazatelů, např.

$$ \frac{\sum q_1 p_1 c_1}{\sum q_0 p_0 c_0} = \frac{\sum q_1 p_0 c_0}{\sum q_0 p_0 c_0} \cdot \frac{\sum q_1 p_1 c_0}{\sum q_1 p_0 c_0} \cdot \frac{\sum q_1 p_1 c_1}{\sum q_1 p_1 c_0} \cdot z $$

$$ \frac{\sum q_1 p_1 c_1}{\sum q_0 p_0 c_0} = \frac{\sum q_1 p_0 c_0}{\sum q_0 p_0 c_0} \cdot \frac{\sum q_0 p_1 c_0}{\sum q_0 p_0 c_0} \cdot \frac{\sum q_0 p_0 c_1}{\sum q_0 p_0 c_0} \cdot z $$