

koeficientov $m_A = m_{y0} \sqrt{Q_{AA}}$, $m_B = m_{y0} \sqrt{Q_{BB}}$, keď váhové koeficienty Q_{AA} a Q_{BB} sa nachádzajú na diagonále inverznej matice normálnych rovníc a $m_{y0} = \sqrt{\frac{\mathbf{\varepsilon}^T \mathbf{P} \mathbf{\varepsilon}}{n-2}}$.

Ak vydelíme prvú rovnicu (9.58) hodnotou $\sum p$ a druhú rovnicu hodnotou $\sum px$ dostaneme

$$\hat{A}_y + \frac{\sum px}{\sum p} \hat{B}_y - \frac{\sum py}{\sum p} = 0, \quad (9.59)$$

$$\hat{A}_y + \frac{\sum pxx}{\sum px} \hat{B}_y - \frac{\sum pxy}{\sum px} = 0.$$

Dosiahli sme, že vyrovnnacia priamka prechádza ťažiskom bodov T a tzv. ťažiskom ťažkých bodov U so súradnicami

$$x_T = \frac{\sum px}{\sum p}, \quad y_T = \frac{\sum py}{\sum p}, \quad x_U = \frac{\sum pxx}{\sum px}, \quad y_U = \frac{\sum pxy}{\sum px}. \quad (9.60)$$

Ak sú ťažiská T a U od seba dostatočne vzdialené, vypočítame ich smernice priamky

$$\operatorname{tg} \alpha_y = \hat{B}_y = \frac{y_U - y_T}{x_U - x_T}. \quad (9.61)$$

Zjednodušenie rovníc (9.58) docielime redukciou súradníc na ťažisko $T \left(x_T = \frac{\sum px}{\sum p}, y_T = \frac{\sum py}{\sum p} \right)$, $x'_i = x_i - x_T$, $y'_i = y_i - y_T$, vtedy bude $\sum px' = \sum py' = 0$.

Rovnica priamky (9.57) po redukcii na ťažisko má posunutý počiatok do ťažiska, vtedy regresný koeficient $\hat{A}_y = 0$ a

$$y'_i + \varepsilon_{yi} = \hat{B}_y x'_i, \quad \varepsilon_{yi} = \hat{B}_y x'_i - y'_i. \quad (9.62)$$

Regresný koeficient B_y vypočítame z druhej rovnice (9.59), keď do nej dosadíme redukované súradnice na ťažisko

$$\hat{B}_y = \frac{\sum px'y'}{\sum px'x'} \quad \text{a} \quad y' = \hat{B}_y x' = \frac{\sum px'y'}{\sum px'x'} x' = \operatorname{tg} \alpha_y x'. \quad (9.63)$$

Regresná priamka pre druhú podmienku (9.56) bude

$$y' = \hat{B}_x x' = \operatorname{tg} \alpha_x x'. \quad (9.64)$$

Z rovnice (9.64)

$$x' = \cot g \alpha_x y' = \hat{B}_x^* y'. \quad (9.65)$$

Regresné koeficienty \hat{B} v smere osí Y a X sú

$$\hat{B}_y = \frac{\sum px'y'}{\sum px'x'}, \quad \hat{B}_x = \frac{1}{\hat{B}_x^*} = \frac{\sum py'y'}{\sum px'y'}. \quad (9.66)$$

Keby všetky dvojice (x_i, y_i) odpovedali lineárnemu funkčnému vzťahu (ležali na priamke), obidve regresné priamky by splynuli do jednej priamky. Vtedy by obidve smernice priamok boli rovnaké a súčin veličín B_y a B_x^* bol rovný 1:

$$\operatorname{tg} \alpha_y = \operatorname{tg} \alpha_x \quad \text{a} \quad \operatorname{tg} \alpha_y \cot \alpha_x = B_y B_x^* = 1. \quad (9.67)$$

Je to len teoretický predpoklad, ktorý v meračskej praxi prakticky nikdy nenastane, pretože účinkom meračských chýb súčasne nenastane predpoklad, aby $\boldsymbol{\varepsilon}_y^T \boldsymbol{\varepsilon}_y = 0$ a $\boldsymbol{\varepsilon}_x^T \boldsymbol{\varepsilon}_x = 0$. Preto bude pomer

$$r = \sqrt{\frac{\operatorname{tg} \alpha_y}{\operatorname{tg} \alpha_x}} = \sqrt{\frac{\hat{B}_y}{\hat{B}_x}} = \sqrt{B_y B_x^*} \quad (9.68)$$

miera tesnosti náhodného vzťahu veličín y_i a x_i , r sa nazýva **koeficient korelácie**.

Koeficient korelácie r je odmocnina z podielu smerníc oboch regresných priamok alebo geometrický priemer oboch koeficientov regresie. Koeficient korelácie r vypočítame z rovníc (9.66) a vzťahu (9.68)

$$r = \sqrt{B_y B_x^*} = \sqrt{\frac{\sum p x' y'}{\sum p x' x'} \frac{\sum p x' y'}{\sum p y' y'}} = \frac{\sum p x' y'}{\sqrt{\sum p x' x' \sum p y' y'}}. \quad (9.69)$$

Koeficient korelácie pri rovnakých váhach ($p = 1$) bude

$$r = \frac{\sum x' y'}{\sqrt{\sum x' x' + \sum y' y'}}. \quad (9.70)$$

Koeficient korelácie pre lineárnu koreláciu je vhodné počítať zo vzťahu

$$r = \sqrt{1 - \frac{\boldsymbol{\varepsilon}_y^T \boldsymbol{\varepsilon}_y}{\sum y' y'}} = \sqrt{1 - \frac{\boldsymbol{\varepsilon}_x^T \boldsymbol{\varepsilon}_x}{\sum x' x'}}. \quad (9.71)$$

Dôkaz vzťahu (9.71) naznačíme pre argument x . Rovnice opráv sú

$$\varepsilon_i = A + B x_i - y_i. \quad (9.72)$$

V maticovom tvare

$$\boldsymbol{\varepsilon}_{(n,1)} = \mathbf{D}_{(n,2)} \mathbf{d}\mathbf{x}_{(2,1)} - \boldsymbol{\ell}_{(n,1)} \quad (9.73)$$

kde

$$\mathbf{D}_{(n,2)} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{d}\mathbf{x}_{(2,1)} = \begin{bmatrix} A \\ B \end{bmatrix}, \quad \boldsymbol{\ell}_{(n,1)} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Zostavíme funkciu MNS a upravíme:

$$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{d}\mathbf{x}^T \mathbf{D}^T - \boldsymbol{\ell}^T) (\mathbf{D} \mathbf{d}\mathbf{x} - \boldsymbol{\ell}) = \mathbf{d}\mathbf{x}^T \mathbf{D}^T \mathbf{D} \mathbf{d}\mathbf{x} - \mathbf{d}\mathbf{x}^T \mathbf{D}^T \boldsymbol{\ell} - \boldsymbol{\ell}^T \mathbf{D} \mathbf{d}\mathbf{x} + \boldsymbol{\ell}^T \boldsymbol{\ell}. \quad (9.74)$$

Nájďme extrém funkcie deriváciou a porovnaním s nulou

$$\frac{\partial(\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon})}{\partial \mathbf{d}\mathbf{x}} = \mathbf{D}^T \mathbf{D} \mathbf{d}\mathbf{x} - \mathbf{D}^T \boldsymbol{\ell} = 0. \quad (9.75)$$

Regresné koeficienty \hat{A}, \hat{B} vypočítame z rovnice

$$\mathbf{dx} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \ell. \quad (9.76)$$

V rovnici (9.74) vytkneme \mathbf{dx}^T a dostaneme

$$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \mathbf{dx}^T (\mathbf{D}^T \mathbf{D} \mathbf{dx} - \mathbf{D}^T \ell) + \ell^T \ell - \ell^T \mathbf{D} \mathbf{dx}. \quad (9.77)$$

Výraz v zátvorke je rovnica (9.75), ktorá je rovná nule. Zostávajúce členy v rovnici (9.77) majú význam

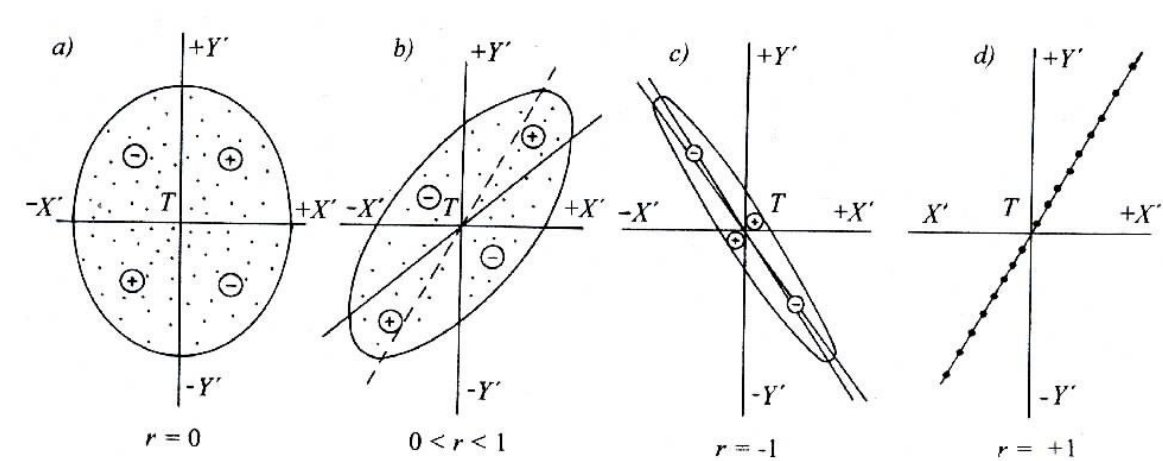
$$\ell^T \ell = \begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{vmatrix} = \sum y y, \quad \ell^T \mathbf{D} \mathbf{dx} = \begin{vmatrix} 1 & x_1 \\ \vdots & x_2 \\ 1 & x_n \end{vmatrix} \begin{vmatrix} \hat{A} \\ \hat{B} \end{vmatrix} = \begin{vmatrix} y_1 & x_1 y_1 \\ y_2 & x_2 y_2 \\ \vdots & \vdots \\ y_n & x_n y_n \end{vmatrix} \begin{vmatrix} \hat{A} \\ \hat{B} \end{vmatrix} = \hat{A} \sum y + \hat{B} \sum x y.$$

Keď súradnice zredukujeme na ťažisko $x'_i = x_i - x_T$ a $y'_i = y_i - y_T$, kde $x_T = \frac{\sum x}{n}$ a

$$y_T = \frac{\sum y}{n}, \text{ vtedy } \hat{A} = 0 \text{ a podľa rovnice (9.63) je } \hat{B} = \frac{\sum x' y'}{\sum x' x'}.$$

Súčet štvorcov opráv vypočítame z rovnice (9.77)

$$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \sum y' y' - \frac{(\sum x' y')^2}{\sum x' x'}. \quad (9.78)$$



Obr. 9.4. Stupne korelácie

Súčet štvorcov opráv dosadíme do rovnice (9.71) a upravíme

$$\begin{aligned}
 r &= \sqrt{1 - \frac{\mathbf{\varepsilon}^T \mathbf{\varepsilon}}{\sum y'y'}} = \sqrt{1 - \frac{\sum y'y' - \frac{\sum x'y' \sum x'y'}{\sum x'x'}}{\sum y'y'}} = \\
 &= \sqrt{\frac{\sum y'y' - \sum y'y' + \frac{\sum x'y' \sum x'y'}{\sum x'x'}}{\sum y'y'}} = \frac{\sum x'y'}{\sqrt{\sum x'x' + \sum y'y'}}. \quad (9.79)
 \end{aligned}$$

Dokázali sme tak rovnosť vzťahov (9.69) a (9.71) na výpočet koeficienta korelácie r .

Koeficient korelácie nadobúda hodnoty z intervalu $0, \pm 1$. V rozsahu intervalu korelácie hodnotíme korelačné vzťahy nasledovne (obr. 9.4).

1. $r = 0$, medzi premennými x_i, y_i nie je lineárny vzťah (korelácia). $\sum x'y' = 0$ a hodnoty v bodkovanom grafe (9.4a) sú symetrické k osám redukovaných súradníc x'_i, y'_i . Obidve regresné priamky sú na seba kolmé a stotožňujú sa s osami $X'Y'$.
2. $r = 1$, stochastický vzťah medzi premennými x'_i, y'_i prechádza na lineárny funkčný vzťah a obidve regresné priamky splynú do jednej priamky (obr. 9.4d).
3. $0 < |r| < 1$. Ak sa rozptylový obrazec vzťahu x'_i, y'_i sústreďuje do I. a III. Kvadrantu, súčin $\sum x'y'$ je kladný a koeficient regresie nadobúda hodnoty v intervale $0 < |r| < 1$. A naopak, ak je rozptylový obrazec v II. a IV. kvadrante súčin, $\sum x'y'$ je záporný a koeficient korelácie bude mať záporné hodnoty v intervale $0 > r > -1$.

Variancia koeficienta korelácie

Každá štatistická charakteristika, teda aj koeficient korelácie má tú vlastnosť, že so znižujúcim sa počtom združených dvojíc (x_i, y_i) sa znižuje aj spoľahlivosť vypočítaných regresných koeficientov \hat{A}, \hat{B} . S výpočtom koeficienta korelácie r sa zaoberáme aj otázkou akú veľkú hodnotu výberového koeficienta korelácie považujeme za postačujúcu na rozhodnutie, že dve premenné sú v stochastickom (náhodnom) vzťahu a naopak, že sú v korelačnom vzťahu. K uvedeným problémom potrebujeme poznať rozdelenie empirického koeficienta korelácie pri jeho teoretickej hodnote ρ v základnom súbore združených dvojíc (x_i, y_i) .

Variancie koeficienta korelácie vyjadrujeme vzťahom

$$\sigma_r^2 = \frac{(1 - \rho^2)^2}{n - 1}, \quad (9.80)$$

kde 1 vyjadruje funkčný vzťah, keď obidve regresné priamky splynú do jednej priamky,

$$\rho = E(r),$$

n je počet dvojíc (x_i, y_i) .

Na testovanie spoľahlivosti určenia koeficienta korelácie používame kritickú hodnotu koeficienta korelácie r_α vo výbere zo základného súboru pri hypotéze, že koeficient korelácie (ρ) v základnom súbore je $\rho = 0$. Kritické hodnoty r_α sú uvedené v tab. IX. na posúdenie, že výberový koeficient r vo výbere zo základného súboru s $\rho = 0$ prekročí svojou absolútnou hodnotou údaj r_α s pravdepodobnosťou α , čo zapisujeme

$$P\{|r| > r_\alpha\} = \alpha \quad \text{pri } E(r) = \rho = 0.$$

Na praktické využitie je vhodné testovanie relačnými vzťahmi

$$|r| < t_{\alpha} \sigma_r \quad (9.81)$$

je nepreukázaná korelácia. t_{α} je kritická hodnota, ktorú nájdeme v tabuľke Studentovho rozdelenia s $n-2$ stupňami voľnosti a hladine významnosti α .

$$t_{\alpha} \sigma_r < |r| < 0,40 \quad \text{malá korelácia (veľmi voľný vzťah),} \quad (9.82)$$

$$0,40 < |r| < 0,85 \quad \text{dobrá korelácia (preukázaná korelácia),} \quad (9.83)$$

$$0,85 < |r| < 1 \quad \text{významná korelácia.} \quad (9.84)$$

9.6 Nelineárna korelácia

Keď sa body na korelačnom grafe zoskupujú okolo krivky použijeme najvhodnejšiu nelineárnu funkciu $y = f(x)$. Koeficient korelácie vypočítame podľa vzťahu (9.62)

$$I_{xy}^2 = 1 - \frac{(\mathbf{\epsilon}^T \mathbf{\epsilon})_y}{\sum y' y'} \quad (9.85)$$

Regresný koeficient r^2 hodnotí len lineárny vzťah. Jeho hodnotu určujeme aj v prípade nelineárnej funkcie. Koeficient I sa nazýva tiež aj **index korelácie**.