Časopis pro pěstování matematiky a fysiky

Jaroslav Janko

Koeficient korelace v homogradní statistice

Časopis pro pěstování matematiky a fysiky, Vol. 55 (1926), No. 2, 160--164

Persistent URL: http://dml.cz/dmlcz/121529

Terms of use:

© Union of Czech Mathematicians and Physicists, 1926

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://project.dml.cz

Koeficient korelace v homogradní statistice.

Napsal Dr. *Jaroslav Janko.*

Uvažujme počet s jednotek, o němž předpokládejme, že je konstantní. Tyto jednotky nechť se třídí podle n alternujících znaků $A, B, C, \ldots$, takže střední počet jejich jest respektive s_1, s_2, s_3, ... s_n.

Jest tedy

$$s_1 + s_2 + s_3 + \ldots + s_n = s. \tag{1}$$

Klademe-li $p_1 = \frac{s_1}{s}$, $p_2 = \frac{s_2}{s}$, ... $p_n = \frac{s_n}{s}$,

pak můžeme považovati p_1, p_2 ... p_n za pravděpodobnosti, že jednotka ze souboru náhodně vybraná patří skupině A, B, C, ... Podle rovnice (1) je zřejmě

$$p_1 + p_2 + \ldots + p_n = 1 \tag{2}$$

Vyberme tedy ze souboru libovolně s jednotek; obdržíme-li m_1, m_2, ... m_n jednotek se znaky A, B, C, ... pak jest

$$m_1 + m_2 + \ldots + m_n = s.$$

Položíme nyní

$$m_i = s p_i + l_i \quad (i = 1, 2, 3, \ldots n)$$

při čemž podle (2)

$$l_1 + l_2 + l_3 + \ldots + l_n = 0. \tag{3}$$

Ptejme se nyní na korrelaci mezi k libovolnými l. Při mnohonásobné korelaci (předpokládáme normální) jest korelační funkce[1])

$$\varphi\,(l_1, l_2, \ldots, l_\nu) = \frac{1}{(2\pi)^{\frac{\nu}{2}} \sigma_1 \sigma_2 \ldots \sigma_\nu \sqrt{R}}\, e^{-\frac{1}{2}\chi^2} \tag{4}$$

kde σ jsou standardní deviace, χ^2 jsou elipsoidy četnosti v n-rozměrném prostoru

$$\chi^2 = \frac{1}{R}\left(\Sigma R_{pp} \frac{l^2_p}{\sigma^2_p} + 2\,\Sigma' R_{st}\, \frac{l_s\, l_t}{\sigma_s\, \sigma_t}\right),$$

---

1) A. Bowley: Elements of Statistics str. 406.

$$R = \begin{vmatrix} 1 & r_{12} & r_{13} & \dots & r_{1\nu} \\ r_{21} & 1 & r_{23} & \dots & r_{2\nu} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_{\nu 1} & r_{\nu 2} & r_{\nu 3} & \dots & 1 \end{vmatrix}$$

R_{pp} a R_{st} jsou příslušné minory k jednotlivým koeficientům korelace r_{pp} resp. r_{st}. Dále jest $r_{st} = r_{ts}$. Omezíme-li se na přibližné vyjádření $n!$ pomocí Stirlingovy formule, jest při zanedbání členů řádu $\frac{1}{\sqrt{s}}$ pravděpodobnost současných úchylek[2] $l_1, l_2, \dots l_n$,

$$P\,(m_1, m_2, m_3, \dots m_n) = \frac{e^{-\frac{1}{2s}\Phi}}{(2\pi s)^{\frac{n-1}{2}}\,(p_1 p_2 \dots p_n)^{\frac{1}{2}}} \tag{5}$$

kde
$$\Phi = \frac{l_1^2}{p_1} + \frac{l_2^2}{p_2} + \dots + \frac{l_n^2}{p_n}.$$

Hledejme tedy korelaci k-násobnou kterýchkoliv úchylek, na př. $l_1, l_2, \dots l_k$; na základě rovnice (3) můžeme vyloučiti třeba l_n, ostatní pak $l_{k+1}, l_{k+2}, \dots l_{n-1}$ nabývají všech hodnot od $-\infty$ do $+\infty$. Je to ovšem metoda přibližná, neboť musíme uvážiti, že pokládáme $l_1, l_2, \dots l_n$ za veličiny malé vzhledem k $sp_1, sp_2, \dots sp_n$. Dostáváme tedy

$$P\,(l_1, l_2, \dots l_k) = \int_{-\infty}^{+\infty} d\,l_{k+1} \int_{-\infty}^{+\infty} d\,l_{k+2} \dots \int_{-\infty}^{+\infty} d\,l_{n-1}\, P\,(m_1, m_2, \dots, m_n). \tag{6}$$

Vyloučíme-li l_n pomocí rovnice (3) jest

$$\Phi_1 = \sum_{i=1}^{n-1} l_i^2 \left(\frac{1}{p_i} + \frac{1}{p_n}\right) + \frac{1}{p_n} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} l_i\, l_j, \tag{7}$$

kdež jest $i \neq j$.

Determinant této kvadratické formy si označíme D, subdeterminant pak D_{ij}, subdeterminanty druhého řádu $D_{ij,\,rs}$ atd., prvky jeho d_{pq}.

Máme-li integrovati $P\,(m_1, m_2, \dots \dots m_n)$ na př. podle l_1, rozložíme Φ_1 takto:

---

[2]) Viz na příklad C. V. Charlier: Vorlesungen über die Grundzüge der mathematischen Statistik.

$$\Phi_1 = d_{11}\, l_1^2 + 2 l_1 \sum_{i=2}^{n-1} d_{1i}\, l_i + \sum_{i=2}^{n-1} \sum_{j=2}^{n-1} d_{ij}\, l_i\, l_j =$$

$$= d_{11}\left(l_1 + \frac{d_{12}}{d_{11}}\, l_2 + \frac{d_{13}}{d_{11}}\, l_3 + \ldots + \frac{d_{1\,n-1}}{d_{11}}\, l_{n-1}\right)^2 +$$

$$+ \sum_{i=2}^{n-1} \sum_{j=2}^{n-1} \frac{d_{ij}\, d_{11} - d_{1i}\, d_{1j}}{d_{11}}\, l_i\, l_j.$$

Po této transformaci lze provésti integraci podle ξ

$$\xi = l_1 + \frac{d_{12}}{d_{11}}\, l_2 + \ldots + \frac{d_{1\,n-1}}{d_{11}}\, l_{n-1}.$$

Před integrál vystoupí faktor $\frac{\sqrt{\pi}}{\sqrt{d_{11}}}$; proměnná l_1 je odstraněna a Φ_1 se redukuje na

$$\sum_{i=2}^{n-1} \sum_{j=2}^{n-1} \frac{d_{ij}\, d_{11} - d_{1i}\, d_{1j}}{d_{11}}\, l_i\, l_j.$$

Použijeme-li známých vztahů odvozených Pearsonem[8])

$$d_{ij} = \frac{1}{2}\, \frac{R_{ij}}{\sigma_i\, \sigma_j\, R}, \quad d_{ii} = \frac{1}{2}\, \frac{R_{ii}}{\sigma_i^2\, R}$$

dostáváme

$$\frac{d_{ij}\, d_{11} - d_{1i}\, d_{1j}}{d_{11}} = \frac{1}{2\, \sigma_i\, \sigma_j\, R_{11}} \cdot \frac{R_{ij}\, R_{11} - R_{1i}\, R_{1j}}{R} = \frac{R_{11,\, ij}}{2\sigma_i\, \sigma_j R_{11}}.$$

Determinant v čitateli $R_{ij}\, R_{11} - R_{1i}\, R_{1j}$ se totiž podle známé věty rovná součinu z hlavního determinantu a subdeterminantu druhého řádu. Redukovaná forma Φ_1 se dá vyjádřiti tedy pomocí determinantu R_{11} stejně jako původní pomocí determinantu R. Třeba jen ještě zdůrazniti, že integrál má smysl, jen když forma Φ_1 je definitní. Tímto způsobem pak pokračujeme dále, takže v našem případě máme podle rovnice (4)

$$P\,(l_1, l_2, \ldots l_k) = \frac{e^{-\frac{1}{2}\chi_k^2}}{(2\pi)^{\frac{k}{2}}\, \sigma_1\, \sigma_2 \ldots \sigma_k \sqrt{R_{k+1,\, k+1;\, \ldots\, n-1,\, n-1}}} \tag{8}$$

[8]) Phil. Trans. 187, který uvádí tam prioritu Edgeworthovu. Viz A. Bowley: Elements of statistics.

kdež

$$\chi_k^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{R_{k+1,\,k+1;\,\ldots\,n-1,\,n-1;\,i,\,j}}{\sigma_i\,\sigma_j\,R_{k+1,\,k+1;\,\ldots\,n-1,\,n-1}}\, l_i l_j \tag{9}$$

Tak dostáváme konečně obecný vztah

$$\begin{aligned}\frac{1}{\sigma_i\,\sigma_j}\frac{R_{k+1,\,k+1;\,\ldots\,n-1,\,n-1;\,i,\,j}}{R_{k+1,\,k+1;\,\ldots\,n-1,\,n-1}} &= \\ = \frac{D_{11,\,11;\,\ldots\,i-1,\,j-1;\,i+1,\,j+1;\,\ldots\,k,\,k}}{D_{11,\,11\,\ldots\,k,\,k}}\end{aligned} \tag{10}$$

Na základě této rovnice můžeme pak napsati všechny potřebné případy zvláštní. Uvedeme zde jen jeden příkladem.

Pro $n = 5$, $k = 2$ máme dle (10) a (7) píšeme-li $1 - p_i = q^i$ tyto rovnice

$$\frac{1}{\sigma_1^2}\frac{R_{33,\,44;\,11}}{R_{33,\,44}} = \frac{1}{\sigma_1^2}\,\frac{1}{1 - r_{12}^2} = \frac{q_2}{s\,p_1(p_3 + p_4 + p_5)},$$

$$\frac{1}{\sigma_1^2}\frac{R_{33,\,44;\,22}}{R_{33,\,44}} = \frac{1}{\sigma_2^2}\,\frac{1}{1 - r_{12}^2} = \frac{q_1}{s\,p_1\,(p_3 + p_4 + p_5)},$$

$$\frac{1}{\sigma_1\,\sigma_2}\frac{R_{33,\,44;\,12}}{R_{33,\,44}} = \frac{1}{\sigma_1\,\sigma_2}\,\frac{r_{12}}{1 - r_{12}^2} = \frac{1}{s\,(p_3 + p_4 + p_5)};$$

z nich pak vyplývá

$$r_{12}\frac{\sigma_1}{\sigma_2} = -\frac{p_1}{q_2},$$

$$r_{12}\frac{\sigma_2}{\sigma_1} = -\frac{p_2}{q_1},$$

$$r_{12}^2 = \frac{p_1\,p_2}{q_1\,q_2}.$$

Z prvních dvou rovnic je patrno, že r je negativní, tedy

$$r_{12} = -\sqrt{\frac{p_1\,p_2}{q_1\,q_2}}.$$

Tato rovnice byla po prvé uvedena Pearsonem[4]) a vyplývá zde ihned jako zvláštní případ z našeho obecného odvození. Nemění se ovšem, jedná-li se o jiný případ n při $k = 2$.

Konečně ještě dostáváme

$$\sigma_1 = \sqrt{s\,p_1\,q_1}, \quad \sigma_2 = \sqrt{s\,p_2\,q_2}.$$

---

[4]) Math. Contrib. to the theory of evolution XIV.

Sur le coefficient de correlation dans la statistique homograde.

(Extrait de l'article précédent.)

Nous résolvons la question d'une correlation multiple d'ordre k entre les écarts quelconques l_i $(i = 1, 2 \ldots . n)$ dans l'ensemble s des individus classés d'après n propriétés alternantes.

En employant l'expression approchée $P(m_1, m_2, \ldots . m_n)$ pour la probabilité des écarts simultanés $l_1, l_2 \ldots . l_n$, nous obtenons l'équation (6), qui peut être ramenée à l'équation (8) à l'aide de l'expression $d_{ij} = \frac{1}{2} \frac{R_{ij}}{\sigma_i \sigma_j R}$; il en suit immédiatement la relation (10) entre les coefficients partiels de correlation et les probabilités fondamentales. Comme un cas spécial nous trouvons aisément

$$r_{1, 2} = -\sqrt{\frac{p_1 p_2}{q_1 q_2}},$$

équation indiquée pour la première fois par Pearson.