Uwagi o współczynniku zależności prostoliniowej i wielośredniej

Andrzej Wilkowski
Uniwersytet Ekonomiczny we Wrocławiu

- In this paper we talk about new statistic tools, which enable more precise economics data analysis. Firstly, we define line dependent coefficient as a cosine of angle made of the cross of regression lines. It is the base, thanks to which we can define other nonlinear relation coefficients. Just like the classic correlation coefficient, line dependent coefficient is also asymptotically normal.

- The second part of this article is about multiaverage, generalization of the classic expected value of the random variable idea. The average may be considered as root-mean-square average approximation of the random variable with one point. Multiaverage is approximation of the variable with more than just one point at a time (which is important when we talk about random variables, which distributions are mixtures, or about multimodal densities). While defining multiaverage we use standard moments method.

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}},$$

$$R(X,Y) = \sup_{f,g} r\left(f(X), g(Y)\right),$$

$$R\left(\sum_{j=1}^{m} X_j, \sum_{j=l+1}^{n} X_j\right) = \frac{m-l}{\sqrt{m(n-l)}},$$

$1 \le l+1 \le m \le n$, Yaming (2008)

Dane są proste regresji

$$y = a_1 x + b_1 \ ,$$
$$x = a_2 y + b_2 \ .$$

Współczynnik zależności prostoliniowej k, zmiennych X, Y (Antoniewicz 1988) jest to kosinus kąta pod jakim przecinają się proste regresji.

$$k(X,Y) = \cos\alpha = \frac{a_1 + a_2}{\sqrt{a_1^2 + 1}\sqrt{a_2^2 + 1}}$$

$$k(Var(X),Var(Y),r) = \frac{\left(Var(X)+Var(Y)\right)r}{\sqrt{Var(X)+r^2Var(Y)}\sqrt{Var(Y)+r^2Var(X)}}$$

$$\max_{Var(X),Var(Y)>0,r\in[-1,1]} \left|k(Var(X),Var(Y),r)-r\right| = \frac{\sqrt{10\sqrt{5}-22}}{2}$$

Maksimum jest osiągane dla $r = \pm\sqrt{\sqrt{5}-2}$

oraz Var(X) = Var(Y) , (Wilkowski 1994).

$$\hat{k}_n = \frac{\left(\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2\right)\hat{r}_n}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \hat{r}_n^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}\ \sqrt{\hat{r}_n^2 \sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Próbkowy współczynnik zależności prostoliniowej $\hat{k}_n$ jest asymptotycznie normalny (Wilkowski 2008).

$\hat{k}_n$ jest AN(k , n^{-1} $\delta S \delta^T$)

S jest macierzą kowariancji wektora (X, Y, X^2, Y^2, XY), a wektory

$$\boldsymbol{\delta} = \left(\frac{\partial g}{\partial z_1}|\boldsymbol{z} = E(\boldsymbol{V}), \ldots, \frac{\partial g}{\partial z_5}|\boldsymbol{z} = E(\boldsymbol{V})\right)$$

$$\mathbf{V} = \left(\overline{X}, \overline{Y}, \frac{1}{n}\sum_{i=1}^{n} X_i^2, \frac{1}{n}\sum_{i=1}^{n} Y_i^2, \frac{1}{n}\sum_{i=1}^{n} X_i \ Y_i\right),$$

Funkcja g , dana jest wzorem :

$$g(z_1, z_2, z_3, z_4, z_5) = \frac{(z_5 - z_1 z_2)\left(\sqrt{\frac{z_3 - z_1^2}{z_4 - z_2^2}} + \sqrt{\frac{z_4 - z_2^2}{z_3 - z_1^2}}\right)}{\sqrt{z_3 - z_1^2 + \frac{(z_5 - z_1 z_2)^2}{z_3 - z_1^2}}\sqrt{z_4 - z_2^2 + \frac{(z_5 - z_1 z_2)^2}{z_4 - z_2^2}}}$$

$$\min_{a \in R} E(X-a)^2 = E(X-E(X))^2 = V_1(X).$$

$$\min_{a,b,\dots,c \in R} E(X^n + aX^{n-1} + bX^{n-2} + \dots + c)^2 = E(p_n(X))^2$$

$$p_n(x) = K \begin{vmatrix} 1 & m_1 & \dots & m_n \\ \dots & \dots & \dots & \dots \\ m_{n-1} & m_n & \dots & m_{2n-1} \\ 1 & x & \dots & x^n \end{vmatrix}$$

gdzie , $K \neq 0$, $E(X^k) = m_k < \infty$, $k = 1,2,\dots,2n-1$.

Ponieważ p_n jest wielomianem ortogonalnym stopnia n , więc:

$$p_n(x) = (x - s_1) \ldots (x - s_n) \text{ , gdzie } s_1 < \ldots < s_n .$$

Uporządkowaną n-kę **(s_1, … , s_n)** nazywamy **n-średnią (wielośrednią)** zmiennej losowej X (Antoniewicz 2005).

Wektor ten jest aproksymacją zmiennej losowej n punktami.
Analogonami wariancji i odchylenia standardowego będą wyrażenia:

$$V_n(X) = E((X - s_1)\ldots(X - s_n))^2$$

Dwumodalny rozkład Webera (Antoniewicz, Wilkowski 2004).

Zmienna losowa X o tym rozkładzie $(X \sim W(\alpha, \beta, \gamma))$

ma gęstość postaci

$$g_{\alpha,\beta,\gamma}(x) = \frac{1}{z(\alpha, \beta)} e^{\alpha(x-\gamma)^2 - \beta(x-\gamma)^4}$$

$$x, \gamma \in R \;;\; \alpha, \beta > 0 .$$

$$z(\alpha,\beta)=\int_{\Re} e^{\alpha x^2-\beta x^4}dx = \exp(\frac{\alpha^2}{8\beta})\frac{\sqrt{\Pi}}{\sqrt[4]{2\beta}}D_{-\frac{1}{2}}(-\frac{\alpha}{\sqrt{2\beta}})$$

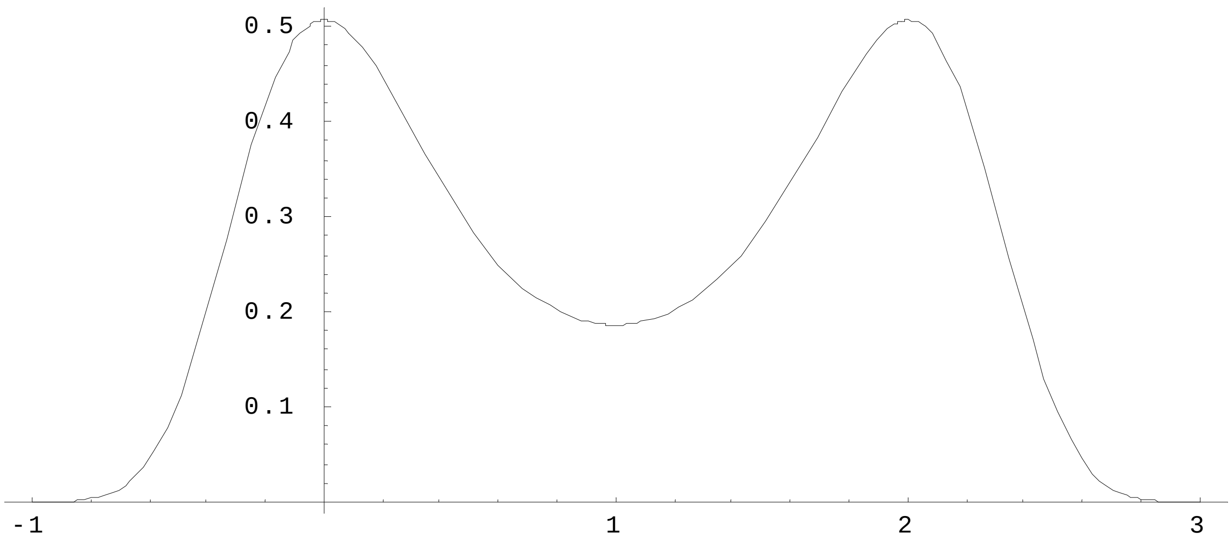
Z definicji widać, że funkcja g ma dwie mody :

$$Mo_1 = -\sqrt{\frac{\alpha}{2\beta}}+\gamma \qquad Mo_2 = \sqrt{\frac{\alpha}{2\beta}}+\gamma$$

a wartość oczekiwana zmiennej losowej X wynosi

$$E(X)=\gamma$$

Funkcja gęstości rozkładu W(2, 1, 1)



$$(s_1, s_2) = (-\ 0{,}08746\ ;\ 1{,}91253)$$

$$(s_1, s_2, s_3) = (-\ 0{,}76371\ ;\ 0{,}82532\ ;\ 2{,}69894)$$

$$V_1(X) = E(X - E(X))^2 = 0{,}83274$$

$$V_2(X) = E((X - s_1)(X - s_2))^2 = 0{,}38928$$

$$V_3(X) = E((X - s_1)(X - s_2)(X - s_3))^2 = 6{,}66794$$

Dziękuję za uwagę