PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1096 — 2005
Ekonometria 15

Marek Walesiak

UOGÓLNIONA MIARA ODLEGŁOŚCI GDM A WSPÓŁCZYNNIK KORELACJI LINIOWEJ PEARSONA I COSINUS KĄTA MIĘDZY WEKTORAMI

1. Wstęp

Do podstawowych pojęć statystycznej analizy wielowymiarowej zalicza się pojęcie obiektu i zmiennej. W artykule przez obiekt rozumie się „najmniejszy element poddany obserwacji, który dostarcza podstawowej z punktu widzenia sformułowanej hipotezy informacji" (por. [Steczkowski, Zeliaś 1981, s. 19-20]). Obiekty są rozumiane w sensie zarówno dosłownym, jak i przenośnym. Obiektem jest w badaniach określona rzecz, osoba, kategoria abstrakcyjna lub zdarzenie. Konkretnymi przykładami obiektów są: konsument X, produkt Y, respondent R, przedsiębiorstwo F, rynek testowy T, dom towarowy D, koncepcja (idea) produktu I, rynek zbytu Z, gospodarstwo domowe G. Zbiór obiektów badania oznaczamy przez $A=\{A_i\}_1^n=\{A_1, A_2, \ldots, A_n\}$. Zmienna w statystycznej analizie wielowymiarowej jest charakterystyką opisującą zbiorowość obiektów. W ujęciu formalnym zmienna M_j to odwzorowanie: M_j: $A \rightarrow R$ ($j=1,2,\ldots,m$). W analizie statystycznej znajomość zbioru obiektów i zmiennych pozwala zapisać macierz danych

$$[x_{ij}]=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}, \tag{1}$$

gdzie: x_{ij} – wartość j-tej zmiennej zaobserwowana w i-tym obiekcie,
$i=1,2,\ldots,n$ – numer obiektu,
$j=1,2,\ldots,m$ – numer zmiennej.

ISSN 0324-8445
ISSN 1507-3866

W artykule zakładamy, że zmienne opisujące obiekty badania mierzone są na skali przedziałowej lub ilorazowej. W celu doprowadzenia zmiennych do porównywalności zachodzi potrzeba pozbawienia wartości zmiennych mian i ujednolicenia rzędów wielkości. Operacja ta nosi nazwę transformacji normalizacyjnej. Zakładamy, że normalizację przeprowadzono z wykorzystaniem jednej z formuł:
a) standaryzacja (dla $j = 1, 2, \ldots, m$)

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \tag{2}$$

gdzie: z_{ij} – znormalizowana wartość j-tej zmiennej zaobserwowana w i-tym obiekcie,

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}; \; s_j = \left[\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2\right]^{0,5};$$

b) przekształcenie ilorazowe (dla $j = 1, 2, \ldots, m$)

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}}. \tag{3}$$

W artykule, na podstawie wykazanych w literaturze związków istniejących między kwadratem odległości euklidesowej a współczynnikiem korelacji liniowej Pearsona i cosinusem kąta między wektorami, wykazane zostaną analogiczne związki dla uogólnionej miary odległości GDM (por. [Walesiak 2002]).

2. Kwadrat odległości euklidesowej a współczynnik korelacji liniowej Pearsona i cosinus kąta między wektorami

Kwadrat odległości euklidesowej dany jest wzorem:

$$d_{jk}^2 = \sum_{i=1}^{n}\left(z_{ij} - z_{ik}\right)^2, \tag{4}$$

gdzie: d_{jk} – odległość między j-tą i k-tą zmienną, $j, k = 1, 2, \ldots, m$.

Na podstawie pracy M.R. Anderberga [1973, s. 113] w pracy K. Jajugi i M. Walesiaka [2004] pokazano, że dla zmiennych standaryzowanych zgodnie

z formułą (2) między kwadratem odległości euklidesowej a współczynnikiem korelacji liniowej Pearsona zachodzi związek:

$$d_{jk}^2 = \sum_{i=1}^{n}\left(z_{ij} - z_{ik}\right)^2 = 2n\left(1 - r_{jk}\right). \tag{5}$$

Dowód 1

$$d_{jk}^2 = \sum_{i=1}^{n}\left(z_{ij} - z_{ik}\right)^2 = \sum_{i=1}^{n}\left[\frac{x_{ij} - \bar{x}_j}{s_j} - \frac{x_{ik} - \bar{x}_k}{s_k}\right]^2 =$$

$$= \sum_{i=1}^{n}\frac{\left(x_{ij} - \bar{x}_j\right)^2}{s_j^2} - 2\sum_{i=1}^{n}\frac{x_{ij} - \bar{x}_j}{s_j}\cdot\frac{x_{ik} - \bar{x}_k}{s_k} + \sum_{i=1}^{n}\frac{\left(x_{ik} - \bar{x}_k\right)^2}{s_k^2} =$$

$$= n\frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}{s_j^2} - 2\sum_{i=1}^{n}\frac{x_{ij} - \bar{x}_j}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}}\cdot\sum_{i=1}^{n}\frac{x_{ik} - \bar{x}_k}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ik} - \bar{x}_j\right)^2}} +$$

$$+ n\frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ik} - \bar{x}_k\right)^2}{s_k^2} = n - 2n\sum_{i=1}^{n}\frac{x_{ij} - \bar{x}_j}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}}\cdot\frac{x_{ik} - \bar{x}_k}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ik} - \bar{x}_j\right)^2}} + n =$$

$$= 2n - 2nr_{jk} = 2n\left(1 - r_{jk}\right).$$

W artykule K. Jajugi i M. Walesiaka [2004] pokazano ogólną formułę związku istniejącego między kwadratem metryki Minkowskiego a ogólnym współczynnikiem powiązania. Szczególnym przypadkiem tej formuły jest związek między kwadratem odległości euklidesowej a współczynnikiem korelacji liniowej Pearsona określony we wzorze (5).

Jeśli we wzorze (4) przeprowadzona zostanie normalizacja zgodnie z formułą (3), to na podstawie pracy M.R. Anderberga [1973, s. 114] można wykazać, że między kwadratem odległości euklidesowej a cosinusem kąta między wektorami obserwacji j-tego i k-tego obiektu istnieje następujący związek:

$$d_{jk}^2 = \sum_{i=1}^{n}\left(z_{ij} - z_{ik}\right)^2 = 2\left(1 - \cos\alpha_{jk}\right). \tag{6}$$

Dowód 2

$$d_{jk}^2 = \sum_{i=1}^{n}\left(z_{ij} - z_{ik}\right)^2 = \sum_{i=1}^{n}\left[\frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} - \frac{x_{ik}}{\sqrt{\sum_{i=1}^{n} x_{ik}^2}}\right]^2 = \sum_{i=1}^{n}\left[\frac{x_{ij}\sqrt{\sum_{i=1}^{n} x_{ik}^2} - x_{ik}\sqrt{\sum_{i=1}^{n} x_{ij}^2}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}\cdot\sqrt{\sum_{i=1}^{n} x_{ik}^2}}\right]^2 =$$

$$= \frac{1}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}\cdot\sqrt{\sum_{i=1}^{n} x_{ik}^2}} \sum_{i=1}^{n}\left[x_{ij}\sqrt{\sum_{i=1}^{n} x_{ik}^2} - x_{ik}\sqrt{\sum_{i=1}^{n} x_{ij}^2}\right]^2 =$$

$$= \frac{1}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}\cdot\sqrt{\sum_{i=1}^{n} x_{ik}^2}} \sum_{i=1}^{n}\left[x_{ij}^2\sqrt{\sum_{i=1}^{n} x_{ik}^2} - 2x_{ij}x_{ik}\sqrt{\sum_{i=1}^{n} x_{ik}^2}\sqrt{\sum_{i=1}^{n} x_{ij}^2} + x_{ik}^2\sqrt{\sum_{i=1}^{n} x_{ij}^2}\right] =$$

$$= 2\left[1 - \frac{\sum_{i=1}^{n} x_{ij}x_{ik}}{\sqrt{\sum_{i=1}^{n} x_{ik}^2}\sqrt{\sum_{i=1}^{n} x_{ij}^2}}\right] = 2\left(1 - \cos\alpha_{jk}\right).$$

3. GDM a współczynnik korelacji liniowej Pearsona i cosinus kąta między wektorami

GDM dla zmiennych mierzonych na skali przedziałowej i (lub) ilorazowej określa wzór (por. [Walesiak 2002, s. 36]):

$$d_{jk} = \left(1 - s_{jk}\right)/2 = \frac{1}{2} - \frac{\sum_{i=1}^{n}\left(z_{ij} - z_{ik}\right)\left(z_{ik} - z_{ij}\right) + \sum_{i=1}^{n}\sum_{l=1}^{m}\left(z_{ij} - z_{il}\right)\left(z_{ik} - z_{il}\right)}{2\left[\sum_{i=1}^{n}\sum_{l=1}^{m}\left(z_{ij} - z_{il}\right)^2 \cdot \sum_{i=1}^{n}\sum_{l=1}^{m}\left(z_{ik} - z_{il}\right)^2\right]^{\frac{1}{2}}}, \quad (7)$$

gdzie $d_{jk}(s_{jk})$ – miara odległości (podobieństwa: $s_{jk} \in [-1; 1]$) GDM między j-tą i k-tą zmienną.

Dla zmiennych standaryzowanych zgodnie z formułą (2) można wykazać, że między GDM a współczynnikami korelacji liniowej Pearsona istnieje związek:

$$d_{jk} = \frac{1}{2} - \frac{-4 + m(r_{jk} + 1) - \sum_{\substack{l=1 \\ l \neq j,\, k}}^{m} r_{jl} - \sum_{\substack{l=1 \\ l \neq j,\, k}}^{m} r_{kl}}{4 \cdot \left[\left(m - \sum_{l=1}^{m} r_{jl} \right) \cdot \left(m - \sum_{l=1}^{m} r_{kl} \right) \right]^{0,5}}. \tag{8}$$

Dowód 3

$$\sum_{i=1}^{n} (z_{ij} - z_{ik})(z_{ik} - z_{ij}) = -\sum_{i=1}^{n} (z_{ij} - z_{ik})^2 - 2n(1 - r_{jk}) - \text{zob. dowód 1;}$$

$$\sum_{i=1}^{n} \sum_{\substack{l=1 \\ l \neq j,\, k}}^{n} (z_{ij} - z_{il})(z_{ik} - z_{il}) = \sum_{i=1}^{n} \sum_{\substack{l=1 \\ l \neq j,\, k}}^{n} \left(\frac{x_{ij} - \bar{x}_j}{s_j} - \frac{x_{il} - \bar{x}_l}{s_l} \right) \left(\frac{x_{ik} - \bar{x}_k}{s_k} - \frac{x_{il} - \bar{x}_l}{s_l} \right) =$$

$$= \sum_{i=1}^{n} \sum_{\substack{l=1 \\ l \neq j,k}}^{n} \left[\frac{x_{ij} - \bar{x}_j}{s_j} \cdot \frac{x_{ik} - \bar{x}_k}{s_k} - \frac{x_{ij} - \bar{x}_j}{s_j} \cdot \frac{x_{il} - \bar{x}_l}{s_l} - \frac{x_{il} - \bar{x}_l}{s_l} \cdot \frac{x_{ik} - \bar{x}_k}{s_k} + \frac{x_{il} - \bar{x}_l}{s_l} \cdot \frac{x_{il} - \bar{x}_l}{s_l} \right] =$$

$$= n \sum_{\substack{l=1 \\ l \neq j,\, k}}^{n} \left[\frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^{n} (x_{ik} - \bar{x}_k)^2}} - \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l)}{\sqrt{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^{n} (x_{il} - \bar{x}_l)^2}} - \right.$$

$$\left. - \frac{\sum_{i=1}^{n} (x_{il} - \bar{x}_l)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{n} (x_{il} - \bar{x}_l)^2} \sqrt{\sum_{i=1}^{n} (x_{ik} - \bar{x}_k)^2}} + \frac{\sum_{i=1}^{n} (x_{il} - \bar{x}_l)(x_{il} - \bar{x}_l)}{\sqrt{\sum_{i=1}^{n} (x_{il} - \bar{x}_l)^2} \sqrt{\sum_{i=1}^{n} (x_{il} - \bar{x}_l)^2}} \right] =$$

$$= \sum_{\substack{l=1 \\ l \neq j,\, k}}^{m} \left[r_{jk} - r_{jl} - r_{kl} + 1 \right] = n(m-2) r_{jk} - n \sum_{\substack{l=1 \\ l \neq j,\, k}}^{m} r_{jl} - n \sum_{\substack{l=1 \\ l \neq j,\, k}}^{m} r_{kl} + n(m-2) =$$

$$= n(m-2)(r_{jk} + 1) - n \sum_{\substack{l=1 \\ l \neq j,\, k}}^{m} r_{jl} n \sum_{\substack{l=1 \\ l \neq j,\, k}}^{m} r_{kl};$$

$$\sum_{i=1}^{n} \sum_{l=1}^{m} (z_{ij} - z_{il})^2 = \sum_{i=1}^{n} \left[\sum_{l=1}^{m} (z_{ij} - z_{il})^2 \right] = \sum_{i=1}^{n} 2n(1 - r_{jl}) = 2n \sum_{i=1}^{n} (1 - r_{jl}).$$

Po podstawieniu do wzoru (7) otrzymuje się prawą stronę równania (8):

$$d_{jk} = \frac{1}{2} - \frac{-2n(1-r_{jk}) + n(m-2)(r_{jk}+1) - n\sum_{\substack{l=1 \\ l\neq j,\,k}}^{m} r_{jl} - n\sum_{\substack{l=1 \\ l\neq j,\,k}}^{m} r_{kl}}{4n\cdot\left[\sum_{l=1}^{m}(1-r_{jl})\cdot\sum_{l=1}^{m}(1-r_{kl})\right]^{0,5}} =$$

$$= \frac{1}{2} - \frac{-2(1-r_{jk}) + (m-2)(r_{jk}+1) - \sum_{\substack{l=1 \\ l\neq j,\,k}}^{m} r_{jl} - \sum_{\substack{l=1 \\ l\neq j,\,k}}^{m} r_{kl}}{4\cdot\left[\sum_{l=1}^{m}(1-r_{jl})\cdot\sum_{l=1}^{m}(1-r_{kl})\right]^{0,5}} =$$

$$= \frac{1}{2} - \frac{-4 + m(r_{jk}+1) - \sum_{\substack{l=1 \\ l\neq j,\,k}}^{m} r_{jl} - \sum_{\substack{l=1 \\ l\neq j,\,k}}^{m} r_{kl}}{4\cdot\left[\left(m-\sum_{l=1}^{m} r_{jl}\right)\cdot\left(m-\sum_{l=1}^{m} r_{kl}\right)\right]^{0,5}}.$$

Jeśli we wzorze (7) przeprowadzona zostanie normalizacja zgodnie z formułą (3), to między GDM a cosinusem kąta między wektorami istnieje następujący związek:

$$d_{jk} = \frac{1}{2} - \frac{-4 + m(\cos\alpha_{jk}+1) - \sum_{\substack{l=1 \\ l\neq j,\,k}}^{m}\cos\alpha_{jl} - \sum_{\substack{l=1 \\ l\neq j,\,k}}^{m}\cos\alpha_{kl}}{4\cdot\left[\left(m-\sum_{l=1}^{m}\cos\alpha_{jl}\right)\cdot\left(m-\sum_{l=1}^{m}\cos\alpha_{kl}\right)\right]^{0,5}}, \qquad (9)$$

gdzie $\cos\alpha_{jk}$ – cosinus kąta między wektorami obserwacji na j-tej i k-tej zmiennej.

Dowód 4

$$\sum_{i=1}^{n}(z_{ij}-z_{ik})(z_{ik}-z_{ij}) = -\sum_{i=1}^{n}(z_{ij}-z_{ik})^2 = -2(1-\cos\alpha_{jk}) - \text{zob. dowód 2;}$$

$$\sum_{i=1}^{n}\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}(z_{ij}-z_{il})(z_{ik}-z_{il})=\sum_{i=1}^{n}\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\left(\frac{x_{ij}}{\sqrt{\sum_{i=1}^{n}x_{ij}^2}}-\frac{x_{il}}{\sqrt{\sum_{i=1}^{n}x_{il}^2}}\right)\left(\frac{x_{ik}}{\sqrt{\sum_{i=1}^{n}x_{ik}^2}}-\frac{x_{il}}{\sqrt{\sum_{i=1}^{n}x_{il}^2}}\right)=$$

$$=\sum_{i=1}^{n}\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\left[\frac{x_{ij}}{\sqrt{\sum_{i=1}^{n}x_{ij}^2}}\cdot\frac{x_{ik}}{\sqrt{\sum_{i=1}^{n}x_{ik}^2}}-\frac{x_{ij}}{\sqrt{\sum_{i=1}^{n}x_{ij}^2}}\cdot\frac{x_{il}}{\sqrt{\sum_{i=1}^{n}x_{il}^2}}-\frac{x_{il}}{\sqrt{\sum_{i=1}^{n}x_{il}^2}}\cdot\frac{x_{ik}}{\sqrt{\sum_{i=1}^{n}x_{ik}^2}}+\frac{x_{il}}{\sqrt{\sum_{i=1}^{n}x_{il}^2}}\cdot\frac{x_{il}}{\sqrt{\sum_{i=1}^{n}x_{il}^2}}\right]=$$

$$=\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\left[\frac{\sum_{i=1}^{n}x_{ij}x_{ik}}{\sqrt{\sum_{i=1}^{n}x_{ij}^2}\sqrt{\sum_{i=1}^{n}x_{ik}^2}}-\frac{\sum_{i=1}^{n}x_{ij}x_{il}}{\sqrt{\sum_{i=1}^{n}x_{ij}^2}\sqrt{\sum_{i=1}^{n}x_{il}^2}}-\frac{\sum_{i=1}^{n}x_{il}x_{ik}}{\sqrt{\sum_{i=1}^{n}x_{il}^2}\sqrt{\sum_{i=1}^{n}x_{ik}^2}}+\frac{\sum_{i=1}^{n}x_{il}x_{il}}{\sqrt{\sum_{i=1}^{n}x_{il}^2}\sqrt{\sum_{i=1}^{n}x_{il}^2}}\right]=$$

$$=\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\left[\cos\alpha_{jk}-\cos\alpha_{jl}-\cos\alpha_{kl}+1\right]=$$

$$=(m-2)\cos\alpha_{jk}-\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\cos\alpha_{jl}-\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\cos\alpha_{kl}+(m-2)=$$

$$=(m-2)(\cos\alpha_{jk}+1)\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\cos\alpha_{jl}-\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\cos\alpha_{kl};$$

$$\sum_{i=1}^{n}\sum_{l=1}^{m}(z_{ij}-z_{il})^2=\sum_{i=1}^{n}\left[\sum_{l=1}^{m}(z_{ij}-z_{il})^2\right]=\sum_{l=1}^{m}2(1-\cos\alpha_{jl})=2\sum_{l=1}^{m}(1-\cos\alpha_{jl}).$$

Po podstawieniu do wzoru (7) otrzymuje się prawą stronę równania (9):

$$d_{jk}=\frac{1}{2}-\frac{-2(1-\cos\alpha_{jk})+(m-2)(\cos\alpha_{jk}+1)-\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\cos\alpha_{jl}-\sum_{\substack{l=1\\ l\neq j,\,k}}^{m}\cos\alpha_{kl}}{4\cdot\left[\sum_{l=1}^{m}(1-\cos\alpha_{jl})\cdot\sum_{l=1}^{m}(1-\cos\alpha_{kl})\right]^{0,5}}=$$

$$= \frac{1}{2} - \frac{-4 + m(\cos\alpha_{jk} + 1) - \sum_{\substack{l=1 \\ l \neq j,\, k}}^{m} \cos\alpha_{jl} - \sum_{\substack{l=1 \\ l \neq j,\, k}}^{m} \cos\alpha_{kl}}{4 \cdot \left[\left(m - \sum_{l=1}^{m} \cos\alpha_{jl} \right) \cdot \left(m - \sum_{l=1}^{m} \cos\alpha_{kl} \right) \right]^{0,5}}.$$

Aby uniknąć zera w mianowniku miar (8) i (9), należy przyjąć założenie, że w zbiorze zmiennych istnieje przynajmniej jedna para takich, dla których obserwacje po normalizacji zgodnie z formułą (2) lub (3) nie są identyczne.

4. Podsumowanie

Na podstawie wykazanych związków między uogólnioną miarą odległości GDM a współczynnikiem korelacji liniowej Pearsona (cosinusem kąta między wektorami) można sformułować kilka spostrzeżeń:

- znając macierz korelacji (cosinusów kąta między wektorami), można obliczyć odległości między zmiennymi,
- odległość między zmiennymi j, k zależy od ich skorelowania (cosinusa kąta) oraz ich korelacji (cosinusów kątów) z pozostałymi zmiennymi,
- dla zbioru zawierającego dwie zmienne $d_{jk} = 1$, jeśli obserwacje po normalizacji nie są identyczne,
- rozważania w artykule dotyczyły odległości między zmiennymi; analogiczne wzory można wyznaczyć, gdy przedmiotem badania są obiekty (por. [Anderberg 1973, s. 113-114]). Wtedy jednak normalizacja dana wzorami (2) i (3) będzie przeprowadzana według obiektów.

Literatura

Anderberg M.R. (1973), *Cluster Analysis for Applications*, Academic Press, New York–San Francisco–London.

Jajuga K., Walesiak M. (2004), *Remarks on the Dependence Measures and the Distance Measures*, [w:] K. Jajuga, M. Walesiak (red.), *Klasyfikacja i analiza danych – teoria i zastosowania*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1022, AE, Wrocław, s. 348-354.

Steczkowski J., Zeliaś A. (1981), *Statystyczne metody analizy cech jakościowych*, PWE, Warszawa.

Walesiak M. (2002), *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, AE, Wrocław.

THE GENERALISED DISTANCE MEASURE GDM AND PEARSON CORRELATION COEFFICIENT AND THE COSINE OF THE ANGLE BETWEEN VECTORS

Summary

The paper gives based on relation between squared Euclidean distance and Pearson correlation coefficient (the cosine of the angle between vectors), similar proposals for Generalised Distance Measure GDM.

Prof. dr hab. Marek Walesiak jest pracownikiem Katedry Ekonometrii i Informatyki Akademii Ekonomicznej we Wrocławiu.