Ekonometria – wykład do kurs. DOiT na kierunku LiT

W2

Temat: Wektor i macierz współczynników korelacji oraz metoda wskaźników pojemności informacyjnej jako realizacja procedury III pkt 2 i 3.

1°. Wektor i macierz wsp. korelacji

Założenia ogólne:

Dla MEL
$$Y=\alpha_0+\alpha_1X_1+\alpha_2X_2+\ldots+\alpha_mX_m+\varepsilon,$$
gdzie $X_1,\ldots,X_m$ dobrane zostały wg. procedury III pkt 1 na podstawie obserwacji n-krotnej:

$$y=\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \qquad X=\begin{array}{c} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \\ \begin{array}{cccc} X_1 & X_2 & \cdots & X_m \end{array} \end{array}$$

Do oceny liniowej zależności Y od $X_1, \ldots, X_m$,

a więc (i) miary „powiązania" Y z każdą X_j

(ii) miary „powiązania" $X_1, \ldots, X_m$

wykorzystuje się:

a) macierz typu kolumny, czyli wektor R_0, gdzie

$$R_0 = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$$ oraz formuła na każde r_i

jest następująca:

$$(*) \quad r_i = \frac{\sum_{j=1}^{n} (y_j - \bar{y})(x_{ji} - \bar{x}_i)}{\sqrt{\left(\sum_{j=1}^{n} (y_j - \bar{y})^2\right)\left(\sum_{j=1}^{n} (x_{ji} - \bar{x}_i)^2\right)}},$$

$i = 1, \ldots, m$

Dla lepszego zrozumienia

wzór (*) rozpiszmy dla przypadku $n = 2$ (dwa pomiary

Mamy system

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & & x_{2m} \end{bmatrix}$$

$$X_1 \quad X_2 \quad \dots \quad X_m$$

Wtedy wsp. $R_0 = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$, gdzie zgodnie z (*)

$$r_1 = \frac{(y_1 - \bar{y})(x_{11} - \bar{X}_1) + (y_2 - \bar{y})(x_{21} - \bar{X}_1)}{\sqrt{\left((y_1 - \bar{y})^2 + (y_2 - \bar{y})^2\right)\left((x_{11} - \bar{X}_1)^2 + (x_{21} - \bar{X}_1)^2\right)}}$$

$$\bar{y} = \tfrac{1}{2}(y_1 + y_2), \quad \bar{X}_1 = \tfrac{1}{2}(x_{11} + x_{21})$$

ZADANIE 1

Proszę dla powyższej sytuacji rozpisać „(*)” dla r_m.

Uwaga 1

W przykładzie podanym dalej pokażemy sposób wyznaczania R_0.

ZADANIE 2

Proszę wykorzystując MS Excel zaprogramować arkusz do wyliczenia R_0

Drugim narzędziem pozwalającym dokonać analizy zbioru $\{Y, X_1, X_2, \ldots, X_m\}$ pod kątem (i) & (ii) jest tzw. MACIERZ KORELACJI dla ciągu $(X_1, X_2, \ldots, X_m)$ oznaczana jako $\mathbb{R}$, gdzie

$$\mathbb{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \vdots & & & \vdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{bmatrix} \in M_{m \times m}$$

gdzie na jej diagonali są same 1 oraz $\mathbb{R} = \mathbb{R}^t$ (symetria), czyli $r_{ij} = r_{ji}$, gdzie

$$(**) \quad r_{ij} = \frac{\sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\left(\sum_{k=1}^{n} (x_{ki} - \bar{x}_i)^2\right)\left(\sum_{k=1}^{n} (x_{kj} - \bar{x}_j)^2\right)}}$$

Uwaga 2

① Wynika z (xx) widać, że $r_{ij} = r_{ji}$

② Niech $i = j$, to licznik = mianownik w (xx), zatem $r_{ii} = 1$.

Podobnie jak dla R_o, założmy na chwilę, że $n=2$, czyli

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \end{bmatrix}$$

$$\quad\quad\quad\quad\quad\quad\quad x_1 \quad x_2$$

Napiszmy postać r_{12} z (xx) wynika, że

$$r_{12} = \frac{(x_{11}-\bar{x}_1)(x_{12}-\bar{x}_2) + (x_{21}-\bar{x}_1)(x_{22}-\bar{x}_2)}{\sqrt{\left((x_{11}-\bar{x}_1)^2 + (x_{21}-\bar{x}_1)^2\right)\left((x_{12}-\bar{x}_2)^2 + (x_{22}-\bar{x}_2)^2\right)}}$$

ZADANIE 3

Dla powyższej wydruku „rozpisać" (xx) dla r_{24}

ZADANIE 4

Wykorzystując MS Excel zaprogramować (xx)

Zeskanowane w CamScanner

Zadanie 5

Nauczyć się posługiwać makrami MS Excel dla R_0 i R.

Zadanie 6

Metodą makr oraz przy użyciu wzorów zad. 2 i 4 ~~[illegible]~~ wyliczyć R_0 i R dla poniższego przykładu.

Przykład

Do opisu kształtowania się wartości sprzedaży usług w mln zł (Y) pewnego przedsiębiorstwa zaproponowano 3 potencjalne zmienne objaśniające:

X_1 – zatrudnienie w os.

X_2 – wartość produkcyjnego majątku trwałego w mln

X_3 – średni czas przestoju maszyn z powodu ich awarii w dniach

Kwartalne dane statystyczne (#n) podano w tabeli

#n	Y	X_1	X_2	X_3
1	11	2,1	11	30
2	13	2,2	11	30
3	16	2,3	12	31
4	24	2,4	13	28
5	24	2,5	15	26
6	26	2,5	17	24
7	27	2,6	17	22
8	28	2,7	18	23
9	29	2,6	18	20
10	33	2,7	19	16
11	33	2,7	19	14
12	36	2,7	22	12

Celem ustalenia rozw. zadań, podamy odpowiedzi: $\bar{y}=25$, $\bar{x}_1=2,5$, $\bar{x}_2=16$, $\bar{x}_3=23$

$$R_0=\begin{bmatrix}0,973\\0,968\\-0,932\end{bmatrix}$$ – korelacja (Y,X_1), (Y,X_2), (Y,X_3)

Natomiast

$$R=\begin{bmatrix}1 & 0,939 & -0,863\\0,939 & 1 & -0,958\\-0,863 & -0,958 & 1\end{bmatrix}.$$

Na kolejnym wykładzie zaprezentujemy w jaki sposób wykorzystując R_0 i R ze zbioru $\{Y, X_1, \ldots, X_m\}$ można

(i) wybrać wszystkie X_i „silnie" skorelowane z Y

(ii) a jednocześnie „słabo" skorelowane między sobą.