



EKONOMETRIA

Temat wykładu:

*Co to jest model ekonometryczny?
Dobór zmiennych objaśniających
w modelu ekonometrycznym*

Prowadzący: dr inż. Zbigniew TARAPATA
e-mail: Zbigniew.Tarapata@isi.wat.edu.pl
http:// zbigniew.tarapata.akcja.pl/p_ekonometria/
tel.: 0-606-45-54-80



CZYM ZAJMUJE SIĘ EKONOMETRIA?

Ekonometria jest zbiorem modeli i metod ekonomii, matematyki i wnioskowania statystycznego wykorzystywanych do analizy zjawisk ekonomicznych.

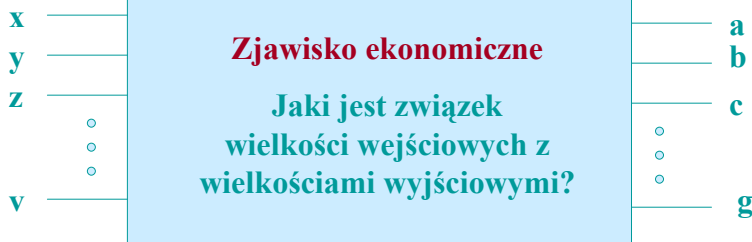
Podstawowym celem ekonometrii jest wyposażenie ekonomii w narzędzia analizy jej zjawisk.

Ekonometria dzieli się na dwa działy: teorie i jej zastosowania.

W ostatnich latach obserwujemy szybki wzrost zastosowań ekonometrii głównie dzięki gwałtownemu rozwojowi techniki komputerowej i informatyki.



CZYM ZAJMUJE SIĘ EKONOMETRIA?



Ekonometria zajmuje się badaniem związków ilościowych i jakościowych pomiędzy wielkościami wejściowymi i wyjściowymi zjawisk ekonomicznych, przy założeniu, że znane są wartości kolejnych pomiarów $(x, y, z, \dots, v)$ oraz odpowiadających im $(a, b, c, \dots, g)$.

Związki te najczęściej opisuje się w postaci równań wektorowo-macierzowych.



SCHEMAT EKONOMETRYCZNY – CELE I PROCEDURA

- opis mechanizmu kształtowania się zjawisk ekonomicznych (cel poznawczy) – ekonometria w wąskim sensie
- przewidywanie dalszego w sensie czasowym przebiegu zjawisk ekonomicznych (cel predykcyjny) – ekonometria sensu stricto
- sterowanie dalszym, w sensie czasowym, przebiegiem zjawisk ekonomicznych (cel decyzyjny) – ekonometria sensu largo
- **Procedura ekonometryczna:**
- Specyfikacja zmiennych → Konstrukcja modelu → Estymacja parametrów → Weryfikacja modelu → Zastosowanie modelu



PRZYKŁADY ZALEŻNOŚCI EKONOMETRYCZNYCH

- J.M. Keynes 'Ogólna teoria zatrudnienia, procentu i pieniądza'

- Dochody (Y) i wydatki na konsumpcję (C) ;

- $f: \mathbb{R} \rightarrow \mathbb{R}, C=f(Y)$

- Założenia i wnioski:

- Spadek dochodów pociąga spadek konsumpcji

$$\frac{dC}{dY} \in (0,1)$$

- Krańcowa skłonność do konsumpcji jest dodatnia i mniejsza od jedności

$$\frac{d(C/Y)}{dY} < 0$$

- Udział konsumpcji w dochodach maleje wraz ze wzrostem dochodów

$$\frac{d(C/Y)}{dY} = \frac{(dC/dY)Y - C}{Y^2} < 0$$

- Przykład funkcji konsumpcji

warunki : $C = \alpha_0 + \alpha_1 Y$



PRZYKŁADY ZALEŻNOŚCI EKONOMETRYCZNYCH

- J.M. Keynes – dwurównaniowy model opisujący zależność między dochodami (Y), wydatkami na konsumpcję (C) i inwestycjami (I):

$$C_t = \alpha_0 + \alpha_1 \cdot Y_t + \alpha_2 \cdot C_{t-1} + \zeta_t$$

$$Y_t = C_t + I_t$$

gdzie:

C_t – wydatki konsumpcyjne w chwili t ;

Y_t – dochody w chwili t ;

I_t – inwestycje w chwili t ;

ζ_t – składnik losowy.



ROWNOWAŻNE POJĘCIA EKONOMETRYCZNE

- Zmienna C nazywana jest :
 - Zmienną zależną
 - Zmienną objaśnianą
 - Regresantem (w teorii estymacji)
 - Zmienną endogeniczną
- Zmienna Y nazywana jest
 - Zmienną niezależną
 - Zmienną objaśniającą
 - Regresorem
 - Zmienną egzogeniczną



SKŁADNIK LOSOWY

- Wykresy funkcji f pokazują niedopasowanie do funkcji liniowej: $C = \alpha_0 + \alpha_1 Y$
- Propozycja : $C = f(Y, \xi)$
- Rzeczywista zależność konsumpcji od dochodu ma charakter stochastyczny :
$$C_t = \alpha_0 + \alpha_1 Y_t + \xi_t, t = 1, \dots, T$$
- Obserwacje na zmiennych C, Y mają charakter szeregów czasowych



PRZYCZYNY LOSOWOŚCI

- Zachowanie ludzi (podmiotów ekonomicznych) cechuje indeterminizm – przy tych samych warunkach podejmowane są różne decyzje;
- Pomiar zjawisk jest niedokładny – błędy obserwacji;
- „Wadliwość” modelu:
 - brak zmiennych objaśniających ,
 - zły typ funkcji f (np. powinny być skokowe a są ciągłe),
 - modele jednej zmiennej a powinny być wielu zmiennych,
 - modele nieliniowe – związane z tym interpolacje,
 - problemy autokorelacji,
 - współzależność równań.



WPROWADZENIE DO MODELOWANIA EKONOMETRYCZNEGO

Szukamy związku zmiennej objaśnianej Y z wartościami zmiennych objaśniających $X_i, i=1,...,m$.

Szukamy zatem związku f postaci:

$$Y = f(X_1, X_2, \dots, X_m)$$

Gdyby założyć, że jest on liniowy, wtedy:

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_mX_m + \xi$$

gdzie ξ obrazuje nam losowe odkształcenie modelu liniowego. Oczywiście, nie musimy przyjmować modelu liniowego zależności zmiennych objaśnianych od zmiennych objaśniających.



WPROWADZENIE DO MODELOWANIA EKONOMETRYCZNEGO

Typowe problemy powstające w trakcie budowy modelu ekonometrycznego:

- jakie wartości mają parametry równania wiążącego zmienną Y od zmiennych X_i ? (typowe zagadnienie ekonometryczne)
- jak zweryfikować postać funkcji f ? (typowe zagadnienie ekonometryczne)
- zakładając przybliżenie funkcji f funkcją liniową, jak ocenić jej przystawanie do rzeczywistości? (adekwatność modelu – typowe zagadnienie ekonometryczne)
- które zmienne objaśniające mamy przyjąć do modelu, a jakie odrzucić? (jest przecież ich olbrzymia liczba i na pewno nie należy uwzględniać wszystkich)



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Zmienne objaśniające (wartości cech mierzalnych) w modelu ekonometrycznym z formalnego punktu widzenia powinny się odznaczać następującymi **własnościami**:

1. Mieć odpowiednio wysoką zmienność,
2. Być silnie skorelowane ze zmienną objaśnianą,
3. Być słabo skorelowane między sobą,
4. Być silnie skorelowane z innymi zmiennymi nie pełniącymi roli zmiennych objaśniających, które zmienne objaśniające reprezentują (chodzi o to, aby zmienne objaśniające były dobrymi reprezentantami zmiennych, które weszły do zbioru zmiennych objaśniających).



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Eliminowanie zmiennych quasi-stałych

Niech:

$$\bar{X}_i = \frac{1}{n} \sum_{t=1}^n X_{ti}, \quad i = \overline{1, m}$$

średnia z próby

$$S_i = \left[\frac{1}{n} \sum_{t=1}^n (X_{ti} - \bar{X}_i)^2 \right]^{\frac{1}{2}}, \quad i = \overline{1, m}$$

odchylenie standardowe



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Eliminowanie zmiennych quasi-stałych, c.d.

Poziom zmienności (miara):

$$v_i = \frac{S_i}{\bar{X}_i}, \quad i = \overline{1, m}$$

v^* - krytyczna wartość,

$$v^* = 0,1 \div 0,15$$

Zmienne X_k uznajemy za quasi-stałe, gdy

$$v_k < v^*, \text{ dla wybranych } k \text{ z zakresu } i = \overline{1, m}$$

i zmienne te eliminujemy ze zbioru potencjalnych zmiennych objaśniających.



WEKTORY I MACIERZE WSPÓŁCZYNNIKÓW KORELACJI

Zmienne objaśniające muszą być silnie skorelowane ze zmienną objaśnianą oraz słabo skorelowane między sobą.

Jak to mierzyć? Współczynnik korelacji między zmiennymi objaśniającymi X_i a zmienną objaśnianą Y (jeśli nie ma korelacji między pewnymi zmiennymi X_i i Y , to od razu je wyrzucmy!):

$$r_i = \frac{\sum_{t=1}^n (y_t - \bar{y})(x_{ti} - \bar{X}_i)}{\sqrt{\sum_{t=1}^n (y_t - \bar{y})^2 \sum_{t=1}^n (x_{ti} - \bar{X}_i)^2}}, \quad i = \overline{1, m}$$

$$R_0 = \begin{bmatrix} r_1 \\ \bullet \\ r_i \\ \bullet \\ r_m \end{bmatrix}$$



WEKTORY I MACIERZE WSPÓŁCZYNNIKÓW KORELACJI, c.d.

Współczynnik korelacji między zmiennymi objaśniającymi:

$$r_{ij} = \frac{\sum_{t=1}^n (x_{ti} - \bar{X}_i)(x_{tj} - \bar{X}_j)}{\sqrt{\sum_{t=1}^n (x_{ti} - \bar{X}_i)^2 \sum_{t=1}^n (x_{tj} - \bar{X}_j)^2}}, \quad i, j = \overline{1, m}$$

współczynnik
korelacji między
zmiennymi
objaśniającymi

$$R = \begin{bmatrix} 1 & r_{12} & \bullet & r_{1m} \\ r_{21} & 1 & \bullet & r_{2m} \\ \bullet & \bullet & \bullet & \bullet \\ r_{m1} & r_{m2} & \bullet & 1 \end{bmatrix}$$

macierz
współczynników
korelacji pomiędzy
zmiennymi
objaśniającymi



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda Hellwiga

- Idea metody polega na tym, że spośród potencjalnych zmiennych objaśniających tworzy się wszystkie możliwe kombinacje tych zmiennych, a następnie bada się tzw. integralną pojemność informacji jaką niesie za sobą zastosowanie każdej z możliwych kombinacji tych zmiennych.



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda Hellwiga, c.d.

Procedurę wyboru zmiennych objaśniających do modelu według metody Hellwiga możemy zapisać następująco:

1. Tworzymy wszystkie możliwe kombinacje potencjalnych zmiennych objaśniających, które można utworzyć z m zmiennych. Ich liczba jest równa liczbie wszystkich możliwych podzbiorów zbioru m -elementowego, czyli $L=2^m-1$. Dla przykładu, jeżeli $m=2$, to można utworzyć $2^2-1=3$ kombinacje zmiennych: $K_1=\{X_1\}$, $K_2=\{X_2\}$, $K_3=\{X_1, X_2\}$, gdzie K_l -ta kombinacja zmiennych, X_i - i -ta potencjalna zmienna objaśniająca.



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda Hellwiga, c.d.

2. Następnie dla każdej j -tej potencjalnej zmiennej objaśniającej, w każdej kombinacji l -tej ($l = \overline{1, L}$) oblicza się indywidualną pojemność informacji h_{lj} tej zmiennej w kombinacji według wzoru:

$$h_{lj} = \frac{r_j^2}{\sum_{i \in I_l} |r_{ij}|}$$

gdzie:

$I_l = \{i : X_i \in K_l\}$ - zbiór indeksów zmiennych wchodzących w skład l -tej kombinacji, tj. kombinacji K_l ;

h_{lj} - indywidualna pojemność j -tej zmiennej w l -tej kombinacji;

r_j - współczynnik korelacji j -tej potencjalnej zmiennej objaśniającej z zmienną endogeniczną (objaśnianą);

$\sum_{i \in I_l} |r_{ij}|$ - suma bezwzględnych wartości współczynników korelacji j -tej potencjalnej zmiennej objaśniającej z pozostałymi, występującymi w danej kombinacji K_l .



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda Hellwiga, c.d.

3. Kolejnym krokiem jest obliczenie integralnej pojemności potencjalnych zmiennych objaśniających (H_l) jako sumy pojemności indywidualnych w ramach każdej kombinacji, według wzoru:

$$H_l = \sum_{j \in I_l} h_{lj}, \quad l = \overline{1, L}$$

4. Jako zmienne objaśniające wybieramy zmienne należące do tej kombinacji l^* , której odpowiada największa wartość H_l , tzn.

$$H_{l^*} = \max_{l \in \{1, \dots, L\}} H_l$$



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda Hellwiga – przykład obliczeniowy

Przykład 1

Na wartości pewnej zmiennej objaśnianej Y mają wpływ wartości dwóch zmiennych X_1 i X_2 . Na podstawie zebranych danych (15-tu) oszacowano macierz R_0 współczynników korelacji zmiennej objaśnianej Y ze zmiennymi X_1 i X_2 oraz pomiędzy zmiennymi X_1 i X_2 (macierz R):

$$R = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}$$

$$R_0 = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 0.7 \\ -0.3 \end{bmatrix}$$

Dokonać wyboru zmiennych objaśniających do modelu według metody Hellwiga.



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda Hellwiga – przykład obliczeniowy

Tworzymy wszystkie możliwe kombinacje zmiennych:

$$K_1 = \{X_1\}, \quad K_2 = \{X_2\}, \quad K_3 = \{X_1, X_2\}$$

$$\text{Stąd: } I_1 = \{1\}, \quad I_2 = \{2\}, \quad I_3 = \{1, 2\}$$

Dla każdej j -tej zmiennej w każdej l -tej kombinacji musimy policzyć indywidualne pojemności informacji h_{jl} tej zmiennej w kombinacji oraz integralną pojemność informacji H_l .

Dla kombinacji K_1 mamy:

$$h_{11} = \frac{r_1^2}{\sum_{i \in I_1} |r_{i1}|} = \frac{r_1^2}{|r_{11}|} = \frac{0.49}{1.0} = 0.49$$

oraz

$$H_1 = h_{11} = 0.49$$



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda Hellwiga – przykład obliczeniowy

Dla kombinacji K_2 mamy:

$$h_{22} = \frac{r_2^2}{\sum_{i \in I_2} |r_{i2}|} = \frac{r_2^2}{|r_{22}|} = \frac{0.09}{1.0} = 0.09 \quad \text{oraz} \quad H_2 = h_{22} = 0.09$$

Dla kombinacji K_3 mamy:

$$h_{31} = \frac{r_1^2}{\sum_{i \in I_3} |r_{i1}|} = \frac{r_1^2}{|r_{11}| + |r_{21}|} = \frac{0.49}{1.0 + 0.5} = 0.33$$
$$h_{32} = \frac{r_2^2}{\sum_{i \in I_3} |r_{i2}|} = \frac{r_2^2}{|r_{12}| + |r_{22}|} = \frac{0.09}{0.5 + 1.0} = 0.06$$

oraz $H_3 = h_{31} + h_{32} = 0.39$

Największą wartość integralnej pojemności informacji ($H_3=0.39$) otrzymano dla kombinacji $K_3=\{X_1\}$, stąd optymalne jest przyjęcie za zmienną objaśniającą jedynie zmiennej X_1 .



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda analizy grafu powiązań korelacyjnych

- Metoda ta polega na budowie grafu powiązań korelacyjnych pomiędzy potencjalnymi zmiennymi objaśniającymi charakteryzującymi się statystycznie istotnymi wartościami współczynników korelacji pomiędzy sobą. Następnie, poprzez odpowiednią analizę grafu, wybiera się te zmienne, które powinny być zmiennymi objaśniającymi.



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda analizy grafu powiązań korelacyjnych, c.d.

1. Przeprowadzamy weryfikację istotności współczynników korelacji pomiędzy zmiennymi, tzn. musimy sprawdzić hipotezę:

$$H_0 : r_{ij} = 0 \quad \text{dla } i \neq j$$

W tym celu obliczamy wartość krytyczną współczynnika korelacji r^* według wzoru:

$$r^* = \sqrt{\frac{t_{\alpha, v}^2}{n - 2 + t_{\alpha, v}^2}}$$

gdzie n jest liczbą obserwacji, na podstawie których obliczono współczynniki korelacji, a $t_{\alpha, v}$ jest wartością krytyczną rozkładu t-Studenta dla przyjętego poziomu istotności α i $v=n-2$ stopni swobody, odczytywaną z tablic statystycznych.



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda analizy grafu powiązań korelacyjnych, c.d.

2. Tworzymy macierz $R'' = [r''_{ij}]_{m \times m}$, gdzie elementami tej macierzy są następujące liczby:

$$r''_{ij} = \begin{cases} 0, & \text{gdy } |r_{ij}| \leq r^* \text{ dla } i \neq j \\ r_{ij}, & \text{w przeciwnym przypadku} \end{cases}$$

3. Na podstawie macierzy R'' budujemy graf G powiązań pomiędzy zmiennymi. Wierzchołkami grafu są poszczególne zmienne, natomiast krawędzie łączą te pary wierzchołków (zmiennych), którym w macierzy R'' odpowiadają niezerowe wartości. Formalnie możemy graf G zdefiniować następująco:

$$G = \langle W, U \rangle$$

gdzie:

W - zbiór wierzchołków grafu, $W = \{1, \dots, m\}$;

U - zbiór krawędzi grafu, $U \subset W \times W = \{i, j\} \in W \times W : r''_{ij} \neq 0\}$.



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda analizy grafu powiązań korelacyjnych, c.d.

4. Sposób wyboru zmiennych objaśniających do modelu jest następujący:

Z każdej składowej spójności [1] grafu wybieramy te wierzchołki (zmienne), które mają największy stopień [2]. Jeżeli w składowej spójności znajduje się kilka wierzchołków, które mają ten sam stopień, to wybieramy taki, że zmienna z nim związana jest najsilniej skorelowana ze zmienną objaśnianą spośród wszystkich pozostałych wierzchołków o tym samym stopniu. Jeśli mimo to wartości współczynników korelacji są równe dla kilku zmiennych (wierzchołków) o tym samym stopniu, to wszystkie one wchodzi do zbioru zmiennych objaśniających.

[1] Składowa spójność grafu G jest to taki maksymalny (ze względu na liczbę wierzchołków) podgraf grafu G , który jest grafem spójnym, tzn. między każdą parą wierzchołków w tym podgrafie istnieje marszruta (możliwość przejścia z jednego wierzchołka do drugiego w sposób bezpośredni lub pośredni (poprzez inne wierzchołki i gałęzie je łączące)). Przypominamy również, że podgraf grafu tworzymy w ten sposób, że wybieramy dowolne wierzchołki grafu G i wszystkie gałęzie, które między tymi wierzchołkami istniały w grafie G . W konsekwencji - maksymalny podgraf spójny jest to taki podgraf grafu G , że nie można już dodać żadnego innego wierzchołka do tego podgrafu, aby nowy podgraf, który w ten sposób by powstał był nadal grafem spójnym.

[2] Stopień wierzchołka grafu jest to liczba wszystkich gałęzi incydentnych z tym wierzchołkiem, tzn. jest to suma luków "wchodzących" do danego wierzchołka, luków "wychodzących" z danego wierzchołka, krawędzi dochodzących do wierzchołka i pętli znajdujących się przy tym wierzchołku.



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda analizy grafu powiązań korelacyjnych – przykład obliczeniowy

Przykład 2

Na podstawie 15-tu obserwacji dotyczących zmiennej endogenicznej i 7-miu potencjalnych zmiennych objaśniających obliczono współczynniki korelacji, które zestawiono w macierzy R i R_0 . Stosując metodę analizy grafu powiązań korelacyjnych, wybrać optymalną kombinację zmiennych objaśniających do modelu ekonometrycznego. Weryfikację istotności współczynników korelacji przeprowadzić na poziomie istotności $\alpha=0.05$.

$$R = \begin{bmatrix} 1.0 & -0.45 & -0.92 & 0.52 & 0.21 & 0.23 & 0.11 \\ & 1.0 & 0.15 & -0.30 & 0.15 & 0.21 & -0.38 \\ & & 1.0 & 0.21 & -0.72 & 0.17 & 0.06 \\ & & & 1.0 & 0.31 & 0.08 & -0.15 \\ & & & & 1.0 & -0.65 & 0.46 \\ & & & & & 1.0 & -0.71 \\ & & & & & & 1.0 \end{bmatrix} \quad R_0 = \begin{bmatrix} 0.61 \\ -0.52 \\ -0.71 \\ 0.45 \\ 0.82 \\ 0.07 \\ 0.09 \end{bmatrix}$$



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda analizy grafu powiązań korelacyjnych – przykład obliczeniowy, c.d.

1. Z tablic rozkładu t-Studenta obliczamy wartość krytyczną $t_{\alpha, v}$ dla $v=15-2=13$ stopni swobody i $\alpha=0.05$ otrzymując $t_{\alpha, v=0.05, 13}=2.160$.

Wyliczamy wartość krytyczną współczynnika korelacji r^* :

$$r^* = \sqrt{\frac{t_{\alpha, v}^2}{n-2+t_{\alpha, v}^2}} = \sqrt{\frac{2.160^2}{15-2+2.160^2}} = \sqrt{\frac{4.67}{17.67}} = 0.51$$

Tworzymy macierz R'' zgodnie z punktem 2 procedury:

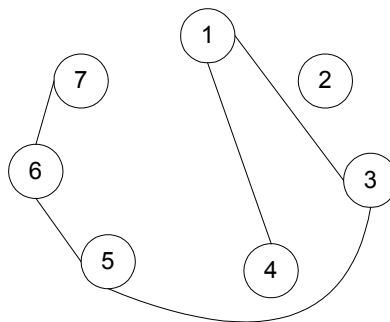
$$R'' = \begin{bmatrix} 1.0 & 0 & -0.92 & 0.52 & 0 & 0 & 0 \\ & 1.0 & 0 & 0 & 0 & 0 & 0 \\ & & 1.0 & 0 & -0.72 & 0 & 0 \\ & & & 1.0 & 0 & 0 & 0 \\ & & & & 1.0 & -0.65 & 0 \\ & & & & & 1.0 & -0.71 \\ & & & & & & 1.0 \end{bmatrix}$$



DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda analizy grafu powiązań korelacyjnych – przykład obliczeniowy, c.d.

Zgodnie z punktem 3 procedury budujemy graf G powiązań korelacyjnych:





DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU

Metoda analizy grafu powiązań korelacyjnych – przykład obliczeniowy, c.d.

Zauważmy, że w grafie G mamy dwie składowe spójności: pierwsza tworzona jest przez pojedynczy wierzchołek 2, tzn. $S_1 = \{2\}$, natomiast drugą tworzy podgraf zbudowany na pozostałych wierzchołkach, tzn. $S_2 = \{1, 3, 4, 5, 6, 7\}$. Wobec tego, zgodnie z punktem 4 procedury, jedną ze zmiennych objaśniających będzie zmienna nr 2. Kolejne zmienne wybieramy poprzez analizę składowej spójności S_2 . Ponieważ w grafie G mamy 4 wierzchołki o tym samym, maksymalnym stopniu równym 2 (tzn. wierzchołki: 1, 3, 5, 6), wobec tego musimy sprawdzić, która ze zmiennych skojarzonych z tymi wierzchołkami jest silniej skorelowana ze zmienną objaśnianą. Z macierzy R_θ mamy, że :

$$r_1=0.61, \quad r_3=-0.71, \quad r_5=0.82, \quad r_6=0.07$$

Z tego wynika, że zmienna X_5 jest najsilniej skorelowana ze zmienną objaśnianą i to ona zostanie wybrana ze składowej S_2 do zbioru zmiennych objaśniających.

Otrzymujemy zatem, że **zmiennymi objaśniającymi w modelu będą zmienne X_2 i X_5 .**