Zeszyty Naukowe ASzWoj nr 3(108) 2017
ISSN 2543-6937

DOBÓR ZMIENNYCH NIEZALEŻNYCH W MODELOWANIU PROCESÓW LOGISTYCZNYCH METODĄ ANALIZY GRAFÓW

dr inż. Marek MRÓZ
Akademia Sztuki Wojennej

Streszczenie

Tematem artykułu jest dobór zmiennych niezależnych w modelowaniu procesów logistycznych metodą analizy grafów. Celem głównym jest teoretyczna i empiryczna analiza możliwości stosowania metody analizy grafów do doboru zmiennych niezależnych w modelach ekonometrycznych. Problem badawczy natomiast obejmował swym zasięgiem także poszukiwanie rozwiązań dotyczących uzyskania właściwych kryteriów wyboru zmiennych niezależnych spośród zmiennych kandydujących, przy wykorzystaniu właściwie stawianych i weryfikowanych hipotez zerowych.

Niewątpliwym osiągnięciem zastosowania metody analizy grafów jest możliwość uzyskania macierzy współczynników korelacji liniowej Pearsona, które obrazują powiązania między poszczególnymi zmiennymi oraz pozwalają na ich podstawie wyeliminować te zmienne kandydujące, które nie są statystycznie istotne dla badanego procesu logistycznego.

Słowa kluczowe: metoda analizy grafów, procesy logistyczne, zmienne zależne i niezależne, dobór zmiennych, budowa modelu ekonometrycznego.

Wstęp

Modelowanie procesów logistycznych polega na tworzeniu modeli matematycznych nazywanych w konkretnych przypadkach modelami ekonometrycznymi, które otrzymujemy dzięki realizacji z góry przyjętego algorytmu oraz dokonanych pośrednich czynności i obliczeń. Podobnie jest i w tym przypadku. Algorytm zastosowany do tworzenia modelu matematycznego (ekonometrycznego) liniowego lub nieliniowego w początkowej fazie dotyczy zazwyczaj doboru zmiennych niezależnych[1] do tego modelu, spośród wszystkich zmiennych kandydujących do niego, których na początku badanego zjawiska może być branych pod uwagę kilka, kilkanaście czy

1 Zmienne niezależne w modelach matematycznych i ekonometrycznych w literaturze przedmiotu badań przez różnych autorów nazywane są także zmiennymi objaśniającymi lub egzogenicznymi.

kilkadziesiąt. Zmiennymi kandydującymi możemy nazwać zbiór wszystkich możliwych zmiennych niezależnych tworzących zbiór opisujący badany proces logistyczny. *Zmienna zależna to taka zmienna, której określone liczbowo wartości rosną lub maleją pod wpływem działania innych zmiennych, zmiennych niezależnych. Zmienna niezależna to taka zmienna (cecha), której konkretne wartości liczbowe powodują zmiany w innych wartościach, w zmiennej zależnej*[2].

Niestety przy badaniu nowych procesów, słabo opisanych lub nieopisanych, trudno jest określić kompletny i zupełny zbiór zmiennych kandydujących. Dlatego też w wyjściowych empirycznych modelach ekonometrycznych występuje składnik (czynnik) losowy ξ_t, który ma praktycznie zastępować te zmienne niezależne, które celowo lub niecelowo pominęliśmy w procesie doboru do modelu. Pominięcie tych zmiennych kandydujących, w początkowej fazie modelowania, występuje najczęściej ze względu na nierozpoznany całkowicie przez badacza opisywany proces logistyczny lub nie do końca opisany teoretycznie w literaturze przedmiotu badań.

Aby skupić się na najważniejszych czynnikach opisujących badany proces logistyczny, należy wyeliminować te zmienne kandydujące, które nie są statystycznie istotne dla badanego procesu. Możemy to uzyskać, wykorzystując testowanie hipotez statystycznych w metodzie analizy grafów przy doborze zmiennych niezależnych do modelu ekonometrycznego.

Zatem celem badawczym podjętym w niniejszym artykule jest analiza możliwości stosowania metody analizy grafów do doboru zmiennych niezależnych w modelowaniu procesów logistycznych.

Natomiast problem badawczy jawi się jako pytanie: Jak wyselekcjonować i według jakich kryteriów przyjąć zmienne niezależne spośród zmiennych kandydujących, do empirycznego modelu ekonometrycznego, wykorzystując metodę analizy grafów?

Poddając szczegółowej analizie cel i problem badawczy sformułowano hipotezę badawczą w brzmieniu: Dobór zmiennych niezależnych do modelu ekonometrycznego oparto na wykorzystaniu metody analizy grafów, w której właściwie postawione i zweryfikowane hipotezy zerowe, a także przyjęte kryteria analizy powiązań sieciowych w grupie grafów, eliminują te zmienne kandydujące, które nie są statystycznie istotne dla badanego procesu logistycznego.

Dla tak sformułowanego celu badań, problemu i hipotezy badawczej przyjęto i wykorzystano teoretyczne i empiryczne metody badawcze. Metoda teoretyczna została oparta na krytycznej analizie literatury przedmiotu badań. Natomiast metodę empiryczną oparto na starannie dobranym i wyselekcjonowanym przykładzie, według którego dokonano obliczeń na bazie danych empirycznym (załącznik 1) i dokonano doboru zmiennych niezależnych w modelowanym procesie logistycznym metodą analizy grafów.

2 J.D. Łaniec, *Elementy statystyki dla pedagogów*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn 1999, s. 17.

Metoda analizy grafów, oprócz innych metod doboru zmiennych niezależnych, takich jak np. metoda Hellwiga, jest jedną z metod o dużym stopniu ogólności, wykorzystywaną jedynie w początkowym procesie modelowania ekonometrycznego i należy traktować ją jak swoistego rodzaju drogowskaz wskazujący kierunek dalszych badań. Uzyskiwane wyniki należy traktować informacyjnie, a nie w sposób kategoryczny. Okazuje się bowiem, że na podstawie tych samych danych empirycznych, posługując się różnymi metodami, możemy uzyskać inne, różniące się wyniki.

Dobór zmiennych niezależnych metodą analizy grafów

Metoda analizy grafów doboru zmiennych niezależnych w modelowaniu procesów logistycznych oparta jest na analizie utworzonych grup grafów (rysunek 5), których wierzchołki tworzą zmienne niezależne wyselekcjonowane w procesie analizy testowanych hipotez statystycznych. Natomiast łuki łączące poszczególne wierzchołki w tych grafach stanowią współczynniki korelacji pomiędzy tymi zmiennymi niezależnymi[3]. Podstawową zasadą, którą należy uwzględniać przy doborze zmiennych niezależnych do modelu ekonometrycznego, jest to, aby współczynniki korelacji liniowej Pearsona[4] były w swej wartości jak najmniejsze pomiędzy zmiennymi niezależnymi, a najlepiej, aby dążyły do zera, i jak największe w swej wartości pomiędzy zmienną zależną[5] a zmiennymi niezależnymi. Współczynnik korelacji to liczba wskazująca, w jakim stopniu dwa zjawiska są ze sobą powiązane i z jaką siłą wywołują wzajemne zmiany[6].

W codziennych obserwacjach możemy dostrzegać zależności pomiędzy wieloma zmiennymi. Mierzenie tych zależności nazywamy korelacją, której celem jest określenie częstości współwystępowania zmiennych skorelowanych[7]. Mając na uwadze cechy ilościowe, można mówić o korelacji dodatniej i ujemnej. Korelacja o znaku dodatnim informuje, że wzrost wartości jednej cechy powoduje wzrost wartości drugiej. Natomiast korelacja ujemna charakteryzuje się tym, że wraz ze wzrostem wartości jednej cechy powoduje spadek wartości drugiej[8].

Należy pamiętać, że współczynniki korelacji liniowej Pearsona można obliczać tylko pomiędzy dwiema zmiennymi mierzalnymi typu ilościowego, tzn. tylko zmiennymi, którym można przypisać jednostki miary, takie jak: kg, metry, sekundy, zł itp.

3 *Wprowadzenie do ekonometrii w przykładach i zadaniach*, red. K. Kukuła, Wydawnictwo Naukowe PWN, Warszawa 2004, s. 24.

4 Współczynnik korelacji liniowej Pearsona przyjmuje wartości od -1 do 1, czyli z przedziału zamkniętego $\langle -1, 1 \rangle$.

5 Zmienna zależna w modelach matematycznych i ekonometrycznych w literaturze przedmiotu badań przez różnych autorów nazywana jest także zmienną objaśnianą lub endogeniczną.

6 J.D. Łaniec, dz. cyt., s. 161.

7 W.P. Zaczyński, *Statystyka w pracy badawczej nauczyciela*, Wydawnictwo „Żak”, Warszawa 1997, s. 21.

8 Tamże, s. 156.

Obliczając współczynniki korelacji Pearsona dla szeregów liczbowych utworzonych z wielu zmiennych, należy sprawdzić, czy[9]:

- wszystkie zmienne są ilościowe;
- zmienne posiadają charakterystykę zbliżoną do rozkładu normalnego;
- związek między zmiennymi zachowuje liniową zależność.

Dopiero po spełnieniu tych trzech warunków: zmienne mierzalne, rozkład zmiennych normalny i korelacja liniowa, zasadna i nieobarczona statystycznym błędem formalnym wydaje się możliwość podjęcia próby definicji i obliczania współczynnika korelacji liniowej Pearsona.

Współczynnik korelacji liniowej Pearsona można zdefiniować jako iloraz kowariancji dwóch zmiennych do iloczynu odchyleń standardowych tych zmiennych[10] i zapisać wzorem definicyjnym:

$$r_{x_i x_j} = \frac{cov(x_i, x_j)}{s_{x_i} \cdot s_{x_j}} \qquad i,j = 1,2,\dots,m \qquad [1]$$

gdzie:
$r_{x_i x_j}$ – współczynnik korelacji liniowej Pearsona pomiędzy zmiennymi x_i, x_j,
$cov(x_i, x_j)$ – kowariancja zmiennych x_i, x_j,
$s_{x_i} s_{x_j}$ – odchylenie standardowe odpowiednio zmiennych x_i, x_j.

Jednak do wykonania praktycznych obliczeń na określonych zmiennych zależnych i niezależnych dla posiadanych danych empirycznych, należy stosować przekształconą postać wzoru 1. Współczynniki korelacji, pomiędzy potencjalnymi zmiennymi niezależnymi, obliczane w praktyce są według wzoru 2:

$$r_{x_i x_j} = r_{ij} = \frac{\sum_{t=1}^{n}(x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)}{\sqrt{\sum_{t=1}^{n}(x_{it} - \bar{x}_i)^2 \sum_{t=1}^{n}(x_{jt} - \bar{x}_j)^2}} \qquad i, j = 1, 2, \dots, m \qquad [2]$$

gdzie:
$r_{x_i x_j}$ – współczynnik korelacji liniowej pomiędzy zmiennymi niezależnymi x_i, x_j,
x_{it} – wartość kolejnej zmiennej niezależnej x_i,
x_{jt} – wartość kolejnej zmiennej niezależnej x_j,
$\bar{x}_i$ – średnia arytmetyczna zmiennej niezależnej x_i,
$\bar{x}_j$ – średnia arytmetyczna zmiennej niezależnej x_j,
n – liczba obserwacji danych empirycznych.

Do obliczenia współczynnika korelacji liniowej Pearsona, pomiędzy jedną zmienną zależną i wszystkimi zmiennymi niezależnymi, wykorzystujemy w praktyce podany poniżej wzór 3:

9 J.D. Łaniec, dz. cyt., s. 162.
10 J. Podgórski, *Statystyka dla studiów licencjackich*, PWE, Warszawa 2005, s. 314.

$$r_{yx_j} = r_{0j} = \frac{\sum_{t=1}^{n}(y_t - \bar{y})(x_{jt} - \bar{x}_j)}{\sqrt{\sum_{t=1}^{n}(y_t - \bar{y})^2 \sum_{t=1}^{n}(x_{jt} - \bar{x}_j)^2}} \qquad j = 1, 2, \ldots, m \qquad [3]$$

gdzie:
r_{yxj}, r_{0j} – współczynnik korelacji zmiennej zależnej y ze zmiennymi niezależnymi x_j,
y_t – wartość kolejnej zmiennej zależnej,
$\bar{y}$ – średnia arytmetyczna zmiennej zależnej y,
x_{jt} – wartość kolejnej zmiennej niezależnej x_j,
$\bar{x}_j$ – średnia arytmetyczna zmiennej niezależnej x_j,
n – liczba obserwacji danych empirycznych.

Współczynnik ten określa liniową zależność pomiędzy zmienną zależną a zmiennymi niezależnymi[11].

Najlepszą kandydującą do modelu zmienną niezależną jest zatem zmienna, która jednocześnie jest bardzo wysoko skorelowana ze zmienną zależną i bardzo słabo skorelowana z pozostałymi zmiennymi niezależnymi, dlatego też słuszne wydaje się nazywanie takiej zmiennej – „niezależną", której współczynnik korelacji z innymi zmiennymi niezależnymi dąży do zera.

W metodzie analizy grafów dobór zmiennych niezależnych do modelu ekonometrycznego wykorzystywanego w analizowanych procesach logistycznych realizuje się najczęściej według podanego poniżej algorytmu:

- ustalenie i wyszczególnienie wszystkich znanych danych empirycznych na przestrzeni badanego okresu, najlepiej w formie tabelarycznej (tabela 1);
- utworzenie dwóch macierzy: jednej $\boldsymbol{R}$ z obliczonymi współczynnikami korelacji pomiędzy wszystkimi zmiennymi niezależnymi (wzór 4 i 5) i drugiej macierzy (wektora) $\boldsymbol{R}_0$ współczynników korelacji pomiędzy zmienną zależną i wszystkimi zmiennymi niezależnymi (wzór 7 i 8);
- postawienie ogólnej oraz szczegółowych hipotez zerowych H_0 i alternatywnych H_1 (wzór 10 i 11);
- obliczenie wartości krytycznej współczynnika korelacji przy wykorzystaniu statystyki testu t-Studenta (wzór 12);
- ustalenie warunków przyjęcia i odrzucenia hipotezy zerowej;
- weryfikacja istotności statystycznej współczynników korelacji zmiennych niezależnych na podstawie warunków przyjęcia i odrzucenia hipotezy zerowej;
- utworzenie macierzy $\boldsymbol{R'}$ (wzór 13) poprzez zastąpienie zerami nieistotnych statystycznie współczynników korelacji w macierzy współczynników korelacji zmiennych niezależnych $\boldsymbol{R}$ (wzór 4);
- utworzenie grup grafów obrazujących powiązania pomiędzy zmiennymi niezależnymi na podstawie analizy macierzy $\boldsymbol{R'}$ (wzór 13);

11 H. Spustek, *Wybrane zagadnienia badań operacyjnych i modelowania liniowego*, AON, Warszawa 2002, s. 25.

− analiza utworzonych grafów i wybór do modelu ekonometrycznego najistotniejszych zmiennych niezależnych na podstawie analizy schematu powiązań pomiędzy wierzchołkami zawierającymi zmienne niezależne oraz siły związku korelacyjnego ze zmienną zależną;

− utworzenie empirycznego modelu ekonometrycznego z wybranymi zmiennymi niezależnymi (wzór 15).

Ustalenie i wyszczególnienie wszystkich znanych danych empirycznych

Postępując zgodnie z przedstawionym powyżej algorytmem, na początku należy określić, zgodnie z przyjętym celem, problemem badawczym i hipotezą, najlepiej w formie tabelarycznej, wszystkie zmienne zależne i niezależne (tabela 1).

Tabela 1

Określenie wszystkich badanych zmiennych zależnych i niezależnych

t	y_t	x_{1t}	x_{2t}	x_{3t}	x_{4t}	x_{5t}	x_{6t}	$\cdots$	x_{mt}
1	y_1	x_{11}	x_{21}	x_{31}	x_{41}	x_{51}	x_{61}	$\cdots$	x_{m1}
2	y_2	x_{12}	x_{22}	x_{32}	x_{42}	x_{52}	x_{62}	$\cdots$	x_{m2}
3	y_3	x_{13}	x_{23}	x_{33}	x_{43}	x_{53}	x_{63}	$\cdots$	x_{m3}
4	y_4	x_{14}	x_{24}	x_{34}	x_{44}	x_{54}	x_{64}	$\cdots$	x_{m4}
5	y_5	x_{15}	x_{25}	x_{35}	x_{45}	x_{55}	x_{65}	$\cdots$	x_{m5}
6	y_6	x_{16}	x_{26}	x_{36}	x_{46}	x_{56}	x_{66}	$\cdots$	x_{m6}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	x_{1n}	x_{2n}	x_{3n}	x_{4n}	x_{5n}	x_{6n}	$\cdots$	x_{mn}

Utworzenie dwóch macierzy: R i R_0

Na podstawie wygenerowanych zmiennych (tabela 1): jednej zmiennej zależnej y_t i m zmiennych niezależnych $x_1, ..., x_6, ..., x_m$, zbudowano przykładową teoretyczną macierz R współczynników korelacji liniowej Pearsona r_{ij} (wzór 4 i 5), wykorzystując wszystkie możliwe powiązania pomiędzy badanymi zmiennymi niezależnymi. Zbudowano także teoretyczny wektor R_0 współczynników korelacji zmiennej zależnej y_t z n zmiennymi niezależnymi $x_1, ..., x_6, ..., x_n$ (wzór 7 i 8).

Dla ułatwienia budowy macierzy współczynników korelacji R (wzór 4) dobrze jest dodatkowo, informacyjnie na obrzeżach macierzy, dołączyć wiersz i kolumnę zmiennych $x_1, ..., x_6, ..., x_m$, pomiędzy którymi łatwiej jest wyznaczać współczynniki korelacji. Indeksy dolne poszczególnych współczynników korelacji stanowią zmienne, pomiędzy którymi określono korelację: $r_{x_1x_1}$, $r_{x_1x_6}$, $r_{x_3x_2}$ itp.

$$R = \begin{array}{c|cccccccc|} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & & x_m \\ \hline x_1 & r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} & r_{x_1x_4} & r_{x_1x_5} & r_{x_1x_6} & \cdots & r_{x_1x_m} \\ x_2 & r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} & r_{x_2x_4} & r_{x_2x_5} & r_{x_2x_6} & \cdots & r_{x_2x_m} \\ x_3 & r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} & r_{x_3x_4} & r_{x_3x_5} & r_{x_3x_6} & \cdots & r_{x_3x_m} \\ x_4 & r_{x_4x_1} & r_{x_4x_2} & r_{x_4x_3} & r_{x_4x_4} & r_{x_4x_5} & r_{x_4x_6} & \cdots & r_{x_4x_m} \\ x_5 & r_{x_5x_1} & r_{x_5x_2} & r_{x_5x_3} & r_{x_5x_4} & r_{x_5x_5} & r_{x_5x_6} & \cdots & r_{x_5x_m} \\ x_6 & r_{x_6x_1} & r_{x_6x_2} & r_{x_6x_3} & r_{x_6x_4} & r_{x_6x_5} & r_{x_6x_6} & \cdots & r_{x_6x_m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & r_{x_nx_1} & r_{x_nx_2} & r_{x_nx_3} & r_{x_nx_4} & r_{x_nx_5} & r_{x_nx_6} & \cdots & r_{x_nx_m} \end{array} \qquad [4]$$

Do dalszych obliczeń, w celu uproszczenia oznaczeń, przyjęto oznaczenia współczynników korelacji, w których pierwszy indeks dolny stanowi numer zmiennej w danym wierszu macierzy, a drugi indeks dolny stanowi numer zmiennej w danej kolumnie macierzy: r_{11}, r_{16}, r_{32} itp. (wzór 5):

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} & \cdots & r_{1m} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{26} & \cdots & r_{2m} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{36} & \cdots & r_{3m} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{46} & \cdots & r_{4m} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} & r_{56} & \cdots & r_{5m} \\ r_{61} & r_{62} & r_{63} & r_{64} & r_{65} & r_{66} & \cdots & r_{6m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & r_{n3} & r_{n4} & r_{n5} & r_{n6} & \cdots & r_{nm} \end{bmatrix} \qquad [5]$$

Poniżej przedstawiono teoretyczną macierz $\boldsymbol{R}$ współczynników korelacji r_{ij} z uproszczonymi oznaczeniami, dla przykładowych sześciu zmiennych niezależnych $x_1, ..., x_6$ (wzór 6):

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{26} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{36} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{46} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} & r_{56} \\ r_{61} & r_{62} & r_{63} & r_{64} & r_{65} & r_{66} \end{bmatrix} \qquad [6]$$

Podobne uproszczenia zastosowano do oznaczeń współczynników korelacji w wektorze $\boldsymbol{R}_0$, w którym pierwszy indeks dolny stanowi zmienną zależną y_t, a drugi indeks dolny stanowi zmienną niezależną (jej numer), np.: r_{yx_1}, r_{yx_4}, r_{yx_6} (wzór 7), a po uproszczeniu oznaczeń współczynniki korelacji będą posiadały postać, np.: r_{01}, r_{04}, r_{06} (wzór 8).

$$R_0 = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ \vdots \\ x_n \end{array} \begin{array}{c} y \\ \begin{bmatrix} r_{yx_1} \\ r_{yx_2} \\ r_{yx_3} \\ r_{yx_4} \\ r_{yx_5} \\ r_{yx_6} \\ \vdots \\ r_{yx_n} \end{bmatrix} \end{array} \quad [7] \qquad R_0 = \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \\ r_{04} \\ r_{05} \\ r_{06} \\ \vdots \\ r_{0n} \end{bmatrix} \quad [8]$$

Teoretyczny wektor $\boldsymbol{R}_0$ współczynników korelacji, dla przykładu jednej zmiennej zależnej y_t z sześcioma zmiennymi niezależnymi $x_1, \ldots, x_6$, przedstawia wzór 9 w ogólnej postaci oraz z uproszczonymi oznaczeniami:

$$R_0 = \begin{bmatrix} r_{yx_1} \\ r_{yx_2} \\ r_{yx_3} \\ r_{yx_4} \\ r_{yx_5} \\ r_{yx_6} \end{bmatrix} = \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \\ r_{04} \\ r_{05} \\ r_{06} \end{bmatrix} \quad [9]$$

Ogólne oraz szczegółowe hipotezy zerowe H_0 i alternatywne H_1

Hipotezy są twierdzeniami, które mówią o oczekiwanych wartościach lub związkach między określonymi zmiennymi[12]. W procesie decyzyjnym formułowane są *a priori* przypuszczenia, które dotyczą szacowanej wartości nieznanego parametru czy rozkładu. Nazywane są one hipotezami statystycznymi[13]. Hipotezą statystyczną nazywamy każdy sąd dotyczący badanej populacji bez przeprowadzenia badania wyczerpującego[14].

*Hipotezę, którą sprawdzamy, nazywamy hipotezą zerową i oznaczamy symbolem H_0. Każdą dopuszczaną hipotezę – poza zerową – nazywa się alternatywną i oznacza symbolem H_1. Hipoteza alternatywna jest zatem tą, którą jesteśmy skłonni przyjąć w przypadku odrzucenia hipotezy zerowe*j[15].

Dlatego kolejnym krokiem jest postawienie ogólnej hipotezy zerowej H_0 oraz alternatywnej H_1. Formułujemy hipotezę zerową H_0, stanowiącą o tym, że współczynnik korelacji liniowej r_{ij} nieistotnie różni się od zera, oraz hipotezę alternatywną H_1,

12 S. Nowak, *Metodologia badań społecznych*, PWN, Warszawa 2012, s. 209.

13 M. Szwed, *Metody statystyczne w naukach społecznych. Elementy teorii i zadania*, Wydawnictwo KUL, Lublin 2008, s. 167, 169.

14 M. Sobczyk, *Statystyka*, Wydawnictwo Naukowe PWN, Warszawa 1994, s. 125.

15 Tamże, s. 126.

stanowiącą o tym, że współczynnik korelacji liniowej r_{ij} istotnie różni się od zera. W formalnym zapisie matematycznym można sformułować te hipotezy w następujący sposób (wzór 10):

$$H_0\colon r_{ij} = 0 \qquad \text{dla } i \neq j \qquad H_1\colon r_{ij} \neq 0 \tag{10}$$

Na podstawie określonej powyżej hipotezy ogólnej możemy postawić szczegółowe hipotezy zerowe H_0 i alternatywne H_1, dla konkretnych współczynników korelacji z macierzy $\boldsymbol{R}$ (wzór 6):

$$\begin{array}{lll}
H_0\colon r_{12} = 0 & \text{dla } i \neq j & H_1\colon r_{12} \neq 0 \\
H_0\colon r_{13} = 0 & \text{dla } i \neq j & H_1\colon r_{13} \neq 0 \\
H_0\colon r_{14} = 0 & \text{dla } i \neq j & H_1\colon r_{14} \neq 0 \\
H_0\colon r_{15} = 0 & \text{dla } i \neq j & H_1\colon r_{15} \neq 0 \\
H_0\colon r_{16} = 0 & \text{dla } i \neq j & H_1\colon r_{16} \neq 0 \\
H_0\colon r_{23} = 0 & \text{dla } i \neq j & H_1\colon r_{23} \neq 0 \\
H_0\colon r_{24} = 0 & \text{dla } i \neq j & H_1\colon r_{24} \neq 0 \\
H_0\colon r_{25} = 0 & \text{dla } i \neq j & H_1\colon r_{25} \neq 0 \\
H_0\colon r_{26} = 0 & \text{dla } i \neq j & H_1\colon r_{26} \neq 0 \\
H_0\colon r_{34} = 0 & \text{dla } i \neq j & H_1\colon r_{34} \neq 0 \\
H_0\colon r_{35} = 0 & \text{dla } i \neq j & H_1\colon r_{35} \neq 0 \\
H_0\colon r_{36} = 0 & \text{dla } i \neq j & H_1\colon r_{36} \neq 0 \\
H_0\colon r_{45} = 0 & \text{dla } i \neq j & H_1\colon r_{45} \neq 0 \\
H_0\colon r_{46} = 0 & \text{dla } i \neq j & H_1\colon r_{46} \neq 0 \\
H_0\colon r_{56} = 0 & \text{dla } i \neq j & H_1\colon r_{56} \neq 0
\end{array} \tag{11}$$

Należy pamiętać, że współczynniki korelacji w macierzach symetrycznych (wzór 6) są po przeciwnej stronie diagonalnej, odpowiednio sobie równe, np.: $r_{12} = r_{21}$, $r_{13} = r_{31}$, $r_{26} = r_{62}$. Dlatego też stawiając hipotezy szczegółowe (wzór 11) dla konkretnych współczynników korelacji z macierzy $\boldsymbol{R}$, nie musimy stawiać ich dla wszystkich współczynników korelacji tej macierzy (lecz tylko dla odpowiednio im równym).

Obliczenie wartości krytycznej współczynnika korelacji liniowej Pearsona

Aby zweryfikować postawione szczegółowe hipotezy zerowe (wzór 11), należy obliczyć wartość krytyczną współczynnika korelacji, według wzoru 12 opartego na statystyce t-Studenta:

$$r_\alpha = \sqrt{\frac{t_\alpha^2}{n\text{-}2 + t_\alpha^2}} \tag{12}$$

gdzie:

r_α – wartość krytyczna współczynnika korelacji liniowej Pearsona,

α – poziom istotności statystycznej przyjęty przez badacza,

n – liczba obserwacji danych empirycznych,

$df = n$-2 – liczba stopni swobody,

t_α – wartość statystyki t-Studenta, którą odczytujemy z tablic rozkładu teoretycznego t-Studenta, przy przyjętym poziomie istotności α i obliczonej liczbie stopni swobody df.

Wartość poziomu istotności α przyjmujemy samodzielnie (arbitralnie). Najczęściej w badaniach społecznych przyjmowane są poziomy: $\alpha = 0{,}05$, $\alpha = 0{,}03$ lub $\alpha = 0{,}01$. Ostatecznie odpowiedzialność za wybór odpowiedniego poziomu istotności spoczywa na badaczu[16].

Następnym krokiem jest obliczenie liczby stopni swobody df (ang. *degrees of freedom*), według wzoru: $df = n$-2. Należy dodać, że kolejne liczby stopni swobody ($df = 1, 2, ..., n$) decydują o kształcie rozkładu statystyki t-Studenta. Rozkład ten nosi nazwę t-Studenta od pseudonimu Student, którym posługiwał się Gosset w 1908 roku, kiedy wprowadził statystykę t[17].

Ustalenie warunków przyjęcia i odrzucenia hipotezy zerowej

Po obliczeniu wartości krytycznej współczynnika korelacji liniowej Pearsona (wzór 12) należy także określić warunki przyjęcia i odrzucenia hipotezy zerowej.

Jeżeli $|r_{ij}|>r_\alpha$, to odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1. Zatem współczynniki r_{ij} są statystycznie istotne. Rozpatrujemy tylko zależność dla $i \neq j$[18].

Jeżeli $|r_{ij}|\leq r_\alpha$, to nie ma podstaw do odrzucenia hipotezy zerowej H_0. Pozostaje hipoteza zerowa H_0, a zatem współczynniki r_{ij} są statystycznie nieistotne.

Regułę postępowania, która każdej losowej próbie przyporządkowuje decyzję przyjęcia lub odrzucenia sprawdzanej hipotezy, nazywamy testem statystycznym[19].

Weryfikacja istotności statystycznej współczynników korelacji zmiennych niezależnych na podstawie warunków przyjęcia i odrzucenia hipotezy zerowej

Opierając się na określonych powyżej warunkach przyjęcia i odrzucenia hipotezy zerowej, dokonujemy weryfikacji istotności statystycznej współczynników korelacji zmiennych niezależnych dla przyjętego poziomu istotności α.

Realizujemy to w ten sposób, że porównujemy wartość każdego współczynnika korelacji z macierzy R (wzór 6) z obliczoną wartością krytyczną współczynnika korelacji liniowej Pearsona r_α (wzór 12).

16 M. Szwed, dz. cyt., s. 146, 176.

17 Tamże.

18 Rozpatrujemy tylko zależność dla $i \neq j$, ponieważ dla $i = j$ współczynniki korelacji r_{ij} zawsze równają się jedności.

19 M. Sobczyk, dz. cyt., s. 126.

Jeżeli przykładowo:

$|r_{12}|>r_{\alpha}$ – to odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1. Pozostawiamy w tworzonej macierzy R' (wzór 13) dotychczasową wartość współczynnika korelacji występującego w macierzy R (wzór 6);

$|r_{14}|\leq r_{\alpha}$ – to nie ma podstaw do odrzucenia hipotezy zerowej H_0. Zatem w tworzonej macierzy R' (wzór 13) zastępujemy zerem dotychczasową wartość współczynnika korelacji występującego w macierzy R (wzór 6).

Weryfikację hipotez statystycznych dokonujemy poprzez konfrontację wyników próby z treścią danej hipotezy[20].

Utworzenie macierzy R′ poprzez zastąpienie zerami nieistotnych statystycznie współczynników korelacji w macierzy R

Eliminując wszystkie nieistotne statystycznie współczynniki korelacji z macierzy R (wzór 6), według ustalonych warunków przyjęcia i odrzucenia hipotezy zerowej, tworzymy macierz R' (wzór 13).

Przykładowa teoretyczna macierz R' z istotnymi statystycznie współczynnikami korelacji, dla np. sześciu zmiennych niezależnych $x_1, \ldots, x_6$, będzie prezentowała się następująco:

$$R' = \begin{bmatrix} 1{,}00 & r_{12} & r_{13} & 0 & 0 & r_{16} \\ r_{21} & 1{,}00 & r_{23} & 0 & 0 & r_{26} \\ r_{31} & r_{32} & 1{,}00 & 0 & 0 & r_{36} \\ 0 & 0 & 0 & 1{,}00 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1{,}00 & 0 \\ r_{61} & r_{62} & r_{63} & 0 & 0 & 1{,}00 \end{bmatrix} \qquad [13]$$

Dla ułatwienia tworzenia konstrukcji grafów i sprawniejszego poszukiwania powiązań pomiędzy poszczególnymi zmiennymi niezależnymi dobrze jest dodatkowo, informacyjnie na obrzeżach macierzy R' (wzór 13), dołączyć wiersz i kolumnę zmiennych $x_1, \ldots, x_6$ (wzór 14), pomiędzy którymi łatwiej jest znaleźć interesujące nas powiązania:

$$R' = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{array} \\ \begin{bmatrix} 1{,}00 & r_{12} & r_{13} & 0 & 0 & r_{16} \\ r_{21} & 1{,}00 & r_{23} & 0 & 0 & r_{26} \\ r_{31} & r_{32} & 1{,}00 & 0 & 0 & r_{36} \\ 0 & 0 & 0 & 1{,}00 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1{,}00 & 0 \\ r_{61} & r_{62} & r_{63} & 0 & 0 & 1{,}00 \end{bmatrix} \end{array} \qquad [14]$$

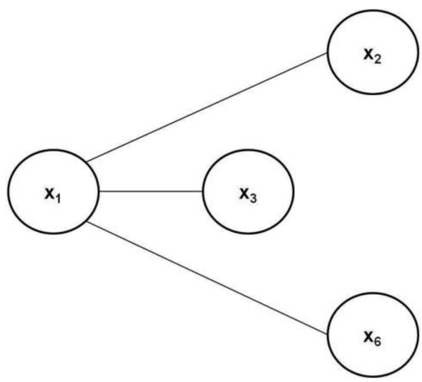
20 Tamże.

Utworzenie grafów obrazujących powiązania pomiędzy zmiennymi niezależnymi

Tworzenie grup grafów jest oparte na analizie macierzy R' (wzór 13) współczynników korelacji zmiennych niezależnych, po wcześniejszej weryfikacji ich istotności statystycznej. Zatem przystępując do tworzenia grafu, analizujemy macierz R' według kolejnych wierszy (lub kolumn) macierzy, szukając powiązań pomiędzy poszczególnymi zmiennymi.

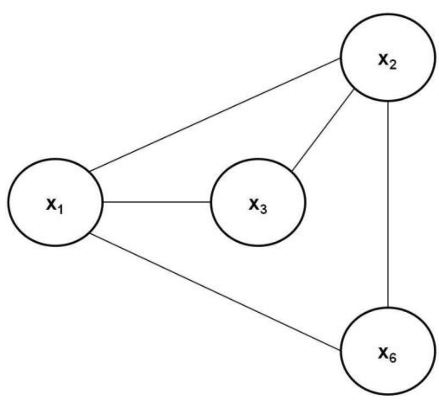
Poczynając od zmiennej x_1 w pierwszym wierszu, szukamy powiązań z kolejnymi zmiennymi $x_2, x_3, ..., x_6$ i tworzymy wierzchołki tych zmiennych w konstruowanym grafie (rysunek 4). Rozpatrujemy i tworzymy wierzchołki tylko tych zmiennych, których współczynniki korelacji z innymi zmiennymi w analizowanej macierz R' są różne od zera, np.:

1) w pierwszym wierszu macierzy R' zmienną x_1 łączymy ze zmienną x_2, x_3 i x_6 i odpowiednio dla nich rysujemy stosowny graf połączeń (rysunek 1):



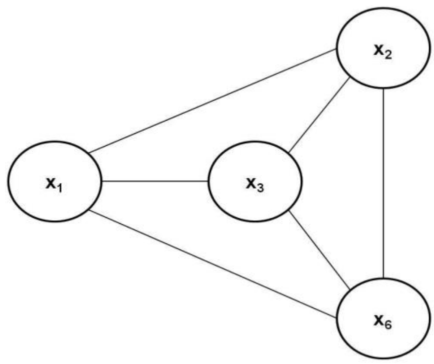
Rys. 1. Wierzchołki grafu utworzone na podstawie analizy pierwszego wiersza macierzy R′

2) w drugim wierszu macierzy zmienną x_2 łączymy ze zmienną x_1, x_3, x_6 (rysunek 2):



Rys. 2. Wierzchołki grafu utworzone na podstawie analizy drugiego wiersza macierzy R′

3) w trzecim wierszu macierzy zmienną x_3 łączymy ze zmienną x_1, x_2 i x_6 (rysunek 3):

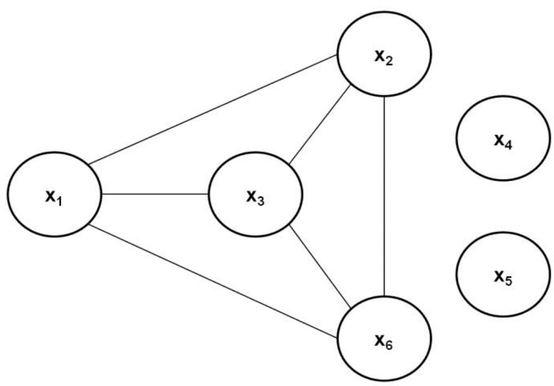


Rys. 3. Wierzchołki grafu utworzone na podstawie analizy trzeciego wiersza macierzy R′

4) w czwartym wierszu macierzy zmiennej x_4 nie łączymy z żadną ze zmiennych, bo jest to zmienna izolowana (rysujemy wierzchołek izolowany);

5) w piątym wierszu macierzy zmiennej x_5 również nie łączymy z żadną ze zmiennych, bo jest to zmienna izolowana (rysujemy wierzchołek izolowany);

6) w szóstym wierszu macierzy zmienną x_6 łączymy ze zmienną x_1, x_2 i x_3 (rysunek 4):



Rys. 4. Wierzchołki grafu utworzone na podstawie analizy czwartego, piątego i szóstego wiersza macierzy R′

Analiza utworzonych grafów i wybór do modelu ekonometrycznego najistotniejszych zmiennych niezależnych

Analizę utworzonych grafów i wybór do modelu ekonometrycznego najistotniejszych zmiennych niezależnych realizujemy na podstawie schematu powiązań pomiędzy wierzchołkami zawierającymi istotne statystycznie zmienne niezależne oraz uwzględniając, w razie potrzeby, zależności korelacyjne ze zmienną zależną.

Zmienne niezależne rozmieszczone w wierzchołkach (rysunek 4) łączy się w grupy grafów, rozpatrując trzy warianty: graf kilkuelementowy, graf dwuelementowy i graf jednoelementowy. Z każdej z tych grup grafów do modelu zostanie zakwalifikowana tylko jedna[21] zmienna niezależna według następujących zasad:

– z grupy grafów z kilkoma wierzchołkami wybieramy do modelu tę zmienną niezależną, która jest wierzchołkiem o najwyższym stopniu, czyli która ma najwięcej powiązań (łuków)[22] z innymi wierzchołkami. W przypadku zmiennych posiadających ten sam, najwyższy stopień powiązań ostatecznym rozstrzygnięciem jest sprawdzenie, która z analizowanych zmiennych jest ponadto najsilniej skorelowana ze zmienną zależną;

– z grupy grafów z dwoma wierzchołkami wybieramy do modelu tę zmienną niezależną, która jest silniej skorelowana ze zmienną zależną. Rozstrzygamy to na podstawie wartości współczynnika korelacji zmiennej zależnej z pozostałymi zmiennymi niezależnymi znajdującymi się w wektorze R_0 (wzór 6);

– z grupy grafów z jednoelementowym wierzchołkiem zawsze wybieramy do modelu zmienną z tego wierzchołka, ponieważ jest ona traktowana jako zmienna izolowana.

Uwzględniając powyższe zasady, po przeprowadzeniu analizy ostatecznej postaci grupy grafów, np. według rysunków 4 i 5, wybieramy do modelu ekonometrycznego następujące zmienne niezależne:

– zmienną x_2 z grafu kilkuelementowego wybieramy jako zmienną o najwyższym stopniu, czyli która ma najwięcej powiązań (łuków) z innymi wierzchołkami i jest równocześnie najsilniej skorelowana ze zmienną zależną[23];

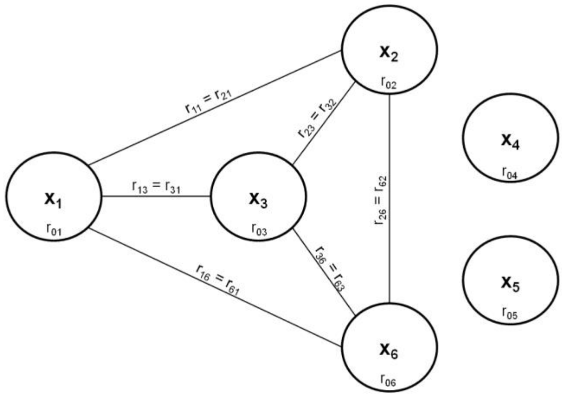
– zmienną x_4 z grafu jednoelementowego wybieramy jako zmienną izolowaną;

– zmienną x_5 z grafu jednoelementowego wybieramy jako zmienną izolowaną.

21 Zakwalifikowanie tylko jednej zmiennej niezależnej z każdej z grup grafów należy traktować jako ogólną regułę. Wyjątek stanowią przypadki, kiedy przy równej liczbie powiązań pomiędzy zmiennymi niezależnymi (wierzchołki o tym samym stopniu) należy rozstrzygać przyjęcie zmiennej do modelu na podstawie współczynnika korelacji zmiennej zależnej z pozostałymi zmiennymi niezależnymi znajdującymi się w wektorze R_0. Jeżeli rozpatrywane wartości tych współczynników są równe, to do modelu wchodzi więcej niż jedna zmienna niezależna z tej grupy wierzchołków.

22 Przy równej liczbie powiązań (łuków) z innymi wierzchołkami ma zastosowanie ogólna reguła z opisanym wyjątkiem.

23 W tym przypadku przyjmujemy *a priori*, że zmienna x_2 jest najsilniej skorelowana ze zmienną zależną.



Rys 5. Grupa grafów z oznaczonymi współczynnikami korelacji liniowej Pearsona

Po dokonaniu wyboru najistotniejszych statystycznie zmiennych niezależnych, np.: x_2, x_4, x_5, metodą analizy grafów spośród analizowanych wszystkich zmiennych kandydujących, możemy utworzyć empiryczny model ekonometryczny, na przykład liniowy z tymi zmiennymi (wzór 15):

$$y_t = \alpha_0 + \alpha_2 x_{2t} + \alpha_4 x_{4t} + \alpha_5 x_{5t} + \xi_t \qquad [15]$$

Przy budowie empirycznego modelu ekonometrycznego niezbędne jest uwzględnienie składnika losowego ξ_t w zależnościach opisujących związki pomiędzy zmiennymi. Wynika to m.in. z następujących przyczyn[24]:

− postępowanie podmiotów gospodarczych i człowieka nie jest deterministyczne;

− występuje niedoskonałość statystycznego sposobu mierzenia zjawisk ekonomicznych (np. procesów logistycznych);

− konstrukcja modelu ekonometrycznego może być wadliwa ze względu na pominięcie zmiennych lub związków między zmiennymi, czy chociażby uwzględnienie niewłaściwych zmiennych.

Model empiryczny może posłużyć do dalszych analiz i rozważań. Można go wykorzystać do budowy teoretycznego modelu ekonometrycznego z oszacowanymi parametrami strukturalnymi[25]. Należy pamiętać, że model to pewna reprezentacja badanego systemu[26], efekt celowo tworzonego i świadomego odwzorowania określonego fragmentu rzeczywistości[27]. Jeżeli model ekonometryczny zostanie pozytywnie zweryfikowany, to można go wykorzystać do opisu ilościowych zależności

24 *Ekonometria i badania operacyjne*, red. M. Gruszczyński, T. Kuszewski, M. Podgórska, Wydawnictwo Naukowe PWN, Warszawa 2009, s. 17.

25 J. Bendkowski, M. Kramarz, W. Kramarz, *Metody i techniki ilościowe w logistyce stosowanej. Wybrane zagadnienia*, Wydawnictwo Politechniki Śląskiej, Gliwice 2010, s. 28.

26 S. Krawczyk, *Metody ilościowe w planowaniu (działalności przedsiębiorstwa)*, Wydawnictwo C.H. Beck, Warszawa 2001, s. 3.

27 K. Kukuła, dz. cyt., s. 63

zachodzących pomiędzy zmiennymi zależnymi i niezależnymi w przeszłości, wykorzystując jego przyczynowo-skutkowe właściwości, a także do prognozowania przyszłych realizacji, na które mają wpływ niezależne zmienne[28].

Zakończenie

Przeprowadzone teoretyczne i empiryczne (załącznik 1) rozważania, dotyczące doboru zmiennych niezależnych w modelowaniu procesów logistycznych, które oparto na konstruowanych modelach ekonometrycznych, a realizowane przy wykorzystaniu metod analizy grafów, wykazały wiele obszarów w literaturze przedmiotu badań jeszcze nieujawnionych i nieopisanych w przystępny sposób dla szerokiej rzeszy odbiorców.

Pogłębione badania teoretyczne, a także szczególnie dokładnie przeanalizowane przykłady empiryczne ujawniły wiele wątpliwości natury teoretycznego podejścia do rozważanego problemu przy wykorzystaniu metody analizy grafów.

Pierwszą, zasadniczą, teoretyczną wątpliwością jest formułowanie (stawianie) hipotezy zerowej H_0 odnoszącej się do tego, że współczynnik korelacji liniowej r_{ij} nieistotnie różni się od zera, wobec alternatywnej hipotezy H_1 stanowiącej o tym, że współczynnik korelacji liniowej r_{ij} istotnie różni się od zera. Problem polega na tym, że zazwyczaj badaczowi zależy na odrzuceniu hipotezy zerowej na rzecz hipotezy alternatywnej, czyli oczekuje tego, że współczynniki korelacji pomiędzy wszystkimi zmiennymi niezależnymi istotnie różniłyby się od zera (H_1: $r_{ij} \neq 0$).

Niestety ogólna zasada doboru zmiennych niezależnych do modelu ekonometrycznego stoi w sprzeczności z tym warunkiem, bowiem najlepszą kandydującą zmienną do modelu jest zmienna niezależna, która jednocześnie jest bardzo wysoko skorelowana ze zmienną zależną i bardzo słabo skorelowana z pozostałymi zmiennymi niezależnymi[29].

Przy tak postawionych hipotezach:

$$H_0 : r_{ij} = 0 \qquad \text{dla } i \neq j \qquad H_1 : r_{ij} \neq 0$$

badacz wolałby, aby hipoteza zerowa nie była odrzucona na rzecz hipotezy alternatywnej. Współczynniki korelacji pomiędzy zmiennymi niezależnymi byłyby bowiem równe zero (H_1: $r_{ij} = 0$), a to oznacza, że nie występowałaby korelacja między nimi (na czym nam bardzo zależy). Należałoby zatem pokusić się o stawianie innej hipotezy zerowej lub ustalić odmienne niż sformułowane warunki przyjęcia i odrzucenia hipotezy zerowej.

Drugą wątpliwością teoretyczną jest końcowy sposób doboru kandydujących zmiennych niezależnych do modelu na podstawie sporządzonego grafu. Kierując

28 Tamże, s. 112.
29 Tamże, s. 16.

się ogólnymi zasadami, w grafie kilkuelementowym należy wybrać tę zmienną, która ma najwięcej powiązań (łuków) z innymi wierzchołkami, czyli jest zmienną o najwyższym stopniu. Znaczy to, że jest skorelowana w większym lub mniejszym stopniu z wieloma innymi zmiennymi niezależnymi, w przeciwieństwie do niewybranych z tego grafu zmiennych o najmniejszym stopniu powiązania[30]. Badaczowi zależy na zmiennych niezależnych w jak najmniejszym stopniu powiązanych między sobą i jak najsłabiej skorelowanych. Zatem zasadne wydaje się, aby w pierwszej kolejności raczej wybierać zmienne o jak najniższym stopniu powiązania z innymi zmiennymi niezależnymi (wierzchołkami) i jednocześnie o jak najwyższym skorelowaniu ze zmienną zależną.

Postawiony na wstępie cel badań dotyczący teoretycznej i empirycznej weryfikacji możliwości stosowania metody analizy grafów do doboru zmiennych niezależnych w modelowaniu procesów logistycznych, a konkretnie w modelach ekonometrycznych, został osiągnięty poprzez dogłębnie analizowane podstawy teoretyczne stosowanej metody oraz poprzez rozwiązanie i przeanalizowanie wybranej empirycznej egzemplifikacji (załącznik 1). Natomiast problem badawczy został rozwiązany poprzez wyselekcjonowanie i określenie kryteriów przyjęcia i odrzucenia zmiennych niezależnych w metodzie analizy grafów na podstawie analizy statystycznej sformułowanych współczynników korelacji liniowej Pearsona. Ponadto poddana analizie hipoteza badawcza została częściowo pozytywnie zweryfikowana. Należy tutaj wskazać kilka zdiagnozowanych powyżej wątpliwości w obszarze teorii problemu, a uzyskanych w wyniku pogłębionej analizy literatury przedmiotu badań.

Powyższe teoretyczne rozważania i wątpliwości naukowe autora mogą stać się przyczynkiem do podjęcia dalszych badań oraz analiz teoretycznych i empirycznych dotyczących sposobów doboru zmiennych niezależnych w modelowaniu procesów logistycznych metodą analizy grafów i innymi znanymi i jeszcze nieznanymi metodami.

Bibliografia

Pozycje zwarte

Bendkowski J., Kramarz M., Kramarz W., *Metody i techniki ilościowe w logistyce stosowanej. Wybrane zagadnienia*, Wydawnictwo Politechniki Śląskiej, Gliwice 2010.

Ekonometria i badania operacyjne, red. M. Gruszczyński, T. Kuszewski, M. Podgórska, Wydawnictwo Naukowe PWN, Warszawa 2009.

Krawczyk S., *Metody ilościowe w planowaniu (działalności przedsiębiorstwa)*, Wydawnictwo C.H. Beck, Warszawa 2001.

Łaniec J.D., *Elementy statystyki dla pedagogów*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn 1999.

Nowak S., *Metodologia badań społecznych*, PWN, Warszawa 2012.

30 Nie uwzględniamy tutaj jeszcze wartości współczynników korelacji zmiennych niezależnych ze zmienną zależną, która przy równych stopniach powiązań decyduje o wyborze.

Podgórski J., *Statystyka dla studiów licencjackich*, PWE, Warszawa 2005.
Sobczyk M., *Statystyka*, Wydawnictwo Naukowe PWN, Warszawa 1994.
Spustek H., *Wybrane zagadnienia badań operacyjnych i modelowania liniowego*, AON, Warszawa 2002.
Szwed M., *Metody statystyczne w naukach społecznych. Elementy teorii i zadania*, Wydawnictwo KUL, Lublin 2008.
Wprowadzenie do ekonometrii w przykładach i zadaniach, red. K. Kukuła, Wydawnictwo Naukowe PWN, Warszawa 2004.
Zaczyński W.P., *Statystyka w pracy badawczej nauczyciela*, Wydawnictwo „Żak", Warszawa 1997.

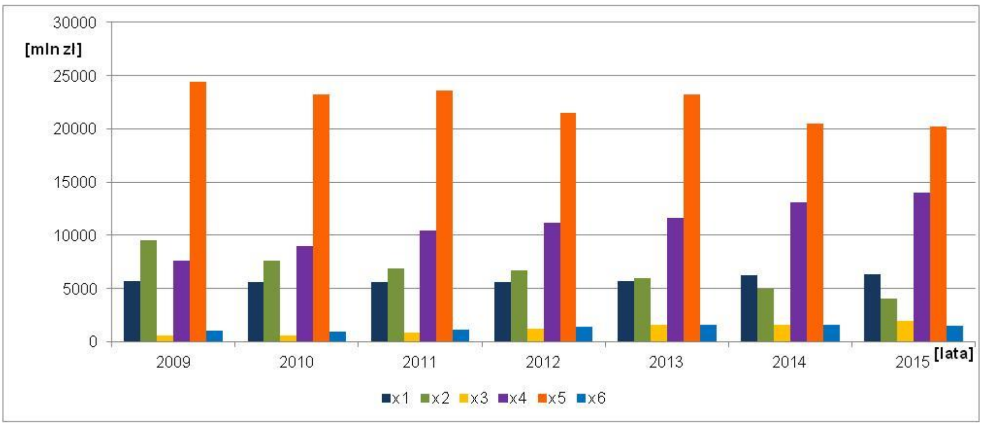
Zasoby internetowe

Rocznik Statystyczny Rzeczypospolitej Polskiej 2016, Zakład Wydawnictw Statystycznych, Warszawa 2016, ISSN 1506-0632 (www.stat.gov.pl).
Rocznik Statystyczny Rzeczypospolitej Polskiej 2015, Zakład Wydawnictw Statystycznych, Warszawa 2015, ISSN 1506-0632 (www.stat.gov.pl).
Rocznik Statystyczny Rzeczypospolitej Polskiej 2014, Zakład Wydawnictw Statystycznych, Warszawa 2014, ISSN 1506-0632 (www.stat.gov.pl).
Rocznik Statystyczny Rzeczypospolitej Polskiej 2013, Zakład Wydawnictw Statystycznych, Warszawa 2013, ISSN 1506-0632 (www.stat.gov.pl).
Rocznik Statystyczny Rzeczypospolitej Polskiej 2012, Zakład Wydawnictw Statystycznych, Warszawa 2012, ISSN 1506-0632 (www.stat.gov.pl).
Rocznik Statystyczny Rzeczypospolitej Polskiej 2011, Zakład Wydawnictw Statystycznych, Warszawa 2011, ISSN 1506-0632 (www.stat.gov.pl).
Rocznik Statystyczny Rzeczypospolitej Polskiej 2010, Zakład Wydawnictw Statystycznych, Warszawa 2010, ISSN 1506-0632 (www.stat.gov.pl).

Załącznik 1

Dobór zmiennych niezależnych na przykładzie metody analizy grafów

Dobór zmiennych niezależnych do modelu ekonometrycznego metodą analizy grafów realizowany w procesach logistycznych oparto na przykładzie (rysunek 1) analizy przychodów ze sprzedaży wyrobów i usług działalności pocztowej i kurierskiej oraz telekomunikacji w Polsce w latach 2009–2015.



Źródło: opracowanie własne na podstawie danych empirycznych zestawionych w tabeli 1.

Rys 1. Przychody ze sprzedaży wyrobów i usług działalności pocztowej i kurierskiej oraz telekomunikacji w Polsce w latach 2009–2015

Przeprowadzona koncepcyjna analiza danych empirycznych implikuje sformułowanie celu badawczego, którym będzie określenie metodą analizy grafów, spośród wszystkich zmiennych kandydujących do modelu ekonometrycznego, tych zmiennych niezależnych, które najistotniej statystycznie wpływają na przychody ze sprzedaży wyrobów i usług działalności pocztowej i kurierskiej oraz telekomunikacji w Polsce, która realizowana była w latach 2009–2015.

Problem badawczy natomiast można sformułować w formie następującego pytania: Jakie zmienne niezależne oznaczające przychody mają najistotniejszy statystycznie wpływ na zmienną zależną oznaczającą przychody ze sprzedaży wyrobów i usług działalności pocztowej i kurierskiej oraz telekomunikacji w Polsce?

Hipoteza badawcza zakłada, że na przychody y_t ze sprzedaży wyrobów i usług (tabela 1) największy wpływ mają: przychody z usług pocztowych x_1, przychody z usług telefonii stacjonarnej x_2, przychody z usług transmisji telewizyjnej i radiofonicznej oraz radia i telewizji kablowej x_3, przychody z usług udostępniania Internetu, transmisji danych i poczty elektronicznej x_4, przychody z usług telefonii komórkowej x_5 oraz przychody z usług transportowych, w tym usług kurierskich x_6.

Dla identyfikacji i wyszczególnienia badanych danych empirycznych na przestrzeni lat 2009–2015 (tabela 1) przyjęto następujące zmienne: jedną zmienną zależną y_t i sześć zmiennych niezależnych $x_1, \ldots, x_6$, których opis podano poniżej:

– y_t – przychody ze sprzedaży wyrobów i usług działalności pocztowej i kurierskiej oraz telekomunikacji w Polsce (w latach 2009–2015);

– x_{1t} – przychody usług pocztowych;

– x_{2t} – przychody usług telefonii stacjonarnej;

– x_{3t} – przychody usług transmisji telewizyjnej i radiofonicznej, radia i telewizji kablowej;

– x_{4t} – przychody usług udostępniania Internetu, transmisji danych i poczty elektronicznej;

– x_{5t} – przychody usług telefonii komórkowej;

– x_{6t} – przychody usług transportowych, w tym usług kurierskich.

Tabela 1

Przychody ze sprzedaży wyrobów i usług działalności pocztowej i kurierskiej oraz telekomunikacji w Polsce w latach 2009–2015 (w mln zł)

Rok	t	y_t	x_{1t}	x_{2t}	x_{3t}	x_{4t}	x_{5t}	x_{6t}
2009	1	50 926,0	5 692,4	9 572,6	581,4	7 604,6	24 423,2	1 044,1
2010	2	49 766,3	5 578,4	7 661,9	628,1	9 005,2	23 214,3	944,4
2011	3	50 579,1	5 604,8	6 916,6	837,8	10 434,6	23 566,4	1 174,9
2012	4	52 053,5	5 602,3	6 720,9	1 249,3	11 193,3	21 534,9	1 453,0
2013	5	49 467,1	5 714,3	5 949,2	1 573,1	11 645,8	23 214,3	1 616,2
2014	6	48 989,1	6 291,4	4 993,5	1 598,4	13 105,9	20 507,3	1 633,6
2015	7	48 710,0	6 354,3	4 079,7	1 960,4	13 984,0	20 261,2	1 551,9

Źródło: opracowanie własne na podstawie Roczników Statystycznych Rzeczypospolitej Polskiej 2010–2016 (www.stat.gov.pl).

Na podstawie danych empirycznych z siedmiu kolejnych lat (tabela 1) wyników przychodu ze sprzedaży wyrobów i usług (w mln zł) obliczono współczynniki korelacji sześciu zmiennych niezależnych i zestawiono w macierzy R (wzór 1):

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{26} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{36} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{46} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} & r_{56} \\ r_{61} & r_{62} & r_{63} & r_{64} & r_{65} & r_{66} \end{bmatrix} = \begin{bmatrix} 1{,}00 & -0{,}76 & 0{,}77 & 0{,}78 & -0{,}80 & 0{,}63 \\ -0{,}76 & 1{,}00 & -0{,}97 & -0{,}99 & 0{,}86 & -0{,}82 \\ 0{,}77 & -0{,}97 & 1{,}00 & 0{,}95 & -0{,}82 & 0{,}93 \\ 0{,}78 & -0{,}99 & 0{,}95 & 1{,}00 & -0{,}89 & 0{,}88 \\ -0{,}80 & 0{,}86 & -0{,}82 & -0{,}89 & 1{,}00 & -0{,}72 \\ 0{,}63 & -0{,}82 & 0{,}93 & 0{,}88 & -0{,}72 & 1{,}00 \end{bmatrix} \quad [1]$$

Obliczono także współczynniki korelacji pomiędzy zmienną zależną y_t i wszystkimi sześcioma zmiennymi niezależnymi $x_1, ..., x_6$, które zestawiono w wektorze R_0 (wzór 2):

$$R_0 = \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \\ r_{04} \\ r_{05} \\ r_{06} \end{bmatrix} = \begin{bmatrix} -0{,}73 \\ 0{,}64 \\ -0{,}56 \\ -0{,}57 \\ 0{,}42 \\ -0{,}38 \end{bmatrix} \qquad [2]$$

Po utworzeniu macierzy *R* współczynników korelacji zmiennych niezależnych oraz wektora R_0 współczynników korelacji jednej zmiennej zależnej z sześcioma zmiennymi niezależnymi przystępujemy do postawienia ogólnej hipotezy zerowej H_0 oraz alternatywnej H_1 (wzór 3):

$$H_0: r_{ij} = 0 \qquad \text{dla } i \neq j \qquad H_1: r_{ij} \neq 0 \qquad [3]$$

Na podstawie sformułowanej hipotezy ogólnej tworzymy szczegółowe hipotezy zerowe H_0 i alternatywne H_1 dla poszczególnych kolejnych współczynników korelacji (wzór 1) z macierzy *R*:

$$\begin{array}{llll}
1) & H_0: r_{12} = 0 & \text{dla } i \neq j & H_1: r_{12} \neq 0 \\
2) & H_0: r_{13} = 0 & \text{dla } i \neq j & H_1: r_{13} \neq 0 \\
3) & H_0: r_{14} = 0 & \text{dla } i \neq j & H_1: r_{14} \neq 0 \\
4) & H_0: r_{15} = 0 & \text{dla } i \neq j & H_1: r_{15} \neq 0 \\
5) & H_0: r_{16} = 0 & \text{dla } i \neq j & H_1: r_{16} \neq 0 \\
6) & H_0: r_{23} = 0 & \text{dla } i \neq j & H_1: r_{23} \neq 0 \\
7) & H_0: r_{24} = 0 & \text{dla } i \neq j & H_1: r_{24} \neq 0 \\
8) & H_0: r_{25} = 0 & \text{dla } i \neq j & H_1: r_{25} \neq 0 \\
9) & H_0: r_{26} = 0 & \text{dla } i \neq j & H_1: r_{26} \neq 0 \\
10) & H_0: r_{34} = 0 & \text{dla } i \neq j & H_1: r_{34} \neq 0 \\
11) & H_0: r_{35} = 0 & \text{dla } i \neq j & H_1: r_{35} \neq 0 \\
12) & H_0: r_{36} = 0 & \text{dla } i \neq j & H_1: r_{36} \neq 0 \\
13) & H_0: r_{45} = 0 & \text{dla } i \neq j & H_1: r_{45} \neq 0 \\
14) & H_0: r_{46} = 0 & \text{dla } i \neq j & H_1: r_{46} \neq 0 \\
15) & H_0: r_{56} = 0 & \text{dla } i \neq j & H_1: r_{56} \neq 0
\end{array} \qquad [4]$$

Po postawieniu piętnastu hipotez szczegółowych dla wszystkich, niepowtarzalnych zależnościowo, współczynników korelacji z macierzy kwadratowej, symetrycznej *R* rozmiaru $6x6$, przystępujemy do obliczenia wartości krytycznej współczynnika korelacji liniowej Pearsona, na podstawie poniższego wzoru 5:

$$r_\alpha = \sqrt{\frac{t_\alpha^2}{n\text{-}2 + t_\alpha^2}} \qquad [5]$$

gdzie:
r_α – wartość krytyczna współczynnika korelacji liniowej Pearsona,
α – poziom istotności statystycznej przyjęty przez badacza,
n – liczba obserwacji danych empirycznych,
$df = n$-2 – liczba stopni swobody,
t_α – wartość statystyki t-Studenta, którą odczytujemy z tablic rozkładu teoretycznego t-Studenta, przy przyjętym poziomie istotności α i obliczonej liczbie stopni swobody df.

Przyjmujemy samodzielnie poziom istotności $\alpha = 0{,}01$ (według merytorycznej oceny badacza) oraz obliczamy liczbę stopni swobody według wzoru: $df = n$-2.

Dla liczby siedmiu obserwacji ($n = 7$) danych empirycznych z lat 2009–2015 (tabela 1) liczba stopni swobody wynosi: $df = n$-2 = 7-2 = 5.

Na podstawie przyjętego poziom istotności $\alpha = 0{,}01$ i obliczonej liczby stopni swobody $df = 5$ odczytujemy, z tablic rozkładu teoretycznego t-Studenta, wartość statystyki t-Studenta, która wynosi $t_\alpha = 4{,}0321$.

Otrzymane dane podstawiamy do wzoru 5:

$$r_{\alpha=0{,}01} = \sqrt{\frac{t_\alpha^2}{n\text{-}2 + t_\alpha^2}} = \sqrt{\frac{4{,}0321^2}{7\text{-}2 + 4{,}0321^2}} = 0{,}87$$

Po obliczeniu wartości krytycznej współczynnika korelacji liniowej Pearsona przystępujemy kolejno do przyjęcia bądź odrzucenia postawionych powyżej piętnastu hipotez szczegółowych, według ustalonych warunków przyjęcia i odrzucenia hipotezy zerowej.

Jeżeli $|r_{ij}| > r_\alpha$, to odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1. Zatem współczynniki r_{ij} są statystycznie istotne. Rozpatrujemy tylko zależność dla $i \neq j$.

Jeżeli $|r_{ij}| \leq r_\alpha$, to nie ma podstaw do odrzucenia hipotezy zerowej H_0. Pozostaje hipoteza zerowa H_0, a zatem współczynniki r_{ij} są statystycznie nieistotne.

Weryfikacja istotności statystycznej poszczególnych współczynników korelacji zmiennych niezależnych z macierzy R (wzór 1) polega na porównaniu modułu[31] wartości każdego z tych współczynników z obliczoną wartością krytyczną współczynnika korelacji liniowej Pearsona $r_\alpha = 0{,}87$ (tabela 2).

31 Moduł – wartość bezwzględna rozpatrywanej wielkości.

Tabela 2

Weryfikacja istotności statystycznej współczynników korelacji

Lp.	H_0	H_1	Analiza kryterium	Wniosek
1.	$H_0: r_{12} = 0$	$H_1: r_{12} \neq 0$	$\lvert r_{12}\rvert \leq r_\alpha$	nie ma podstaw do odrzucenia hipotezy zerowej H_0
2.	$H_0: r_{13} = 0$	$H_1: r_{13} \neq 0$	$\lvert r_{13}\rvert \leq r_\alpha$	nie ma podstaw do odrzucenia hipotezy zerowej H_0
3.	$H_0: r_{14} = 0$	$H_1: r_{14} \neq 0$	$\lvert r_{14}\rvert \leq r_\alpha$	nie ma podstaw do odrzucenia hipotezy zerowej H_0
4.	$H_0: r_{15} = 0$	$H_1: r_{15} \neq 0$	$\lvert r_{15}\rvert \leq r_\alpha$	nie ma podstaw do odrzucenia hipotezy zerowej H_0
5.	$H_0: r_{16} = 0$	$H_1: r_{16} \neq 0$	$\lvert r_{16}\rvert \leq r_\alpha$	nie ma podstaw do odrzucenia hipotezy zerowej H_0
6.	$H_0: r_{23} = 0$	$H_1: r_{23} \neq 0$	$\lvert r_{23}\rvert > r_\alpha$	odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1
7.	$H_0: r_{24} = 0$	$H_1: r_{24} \neq 0$	$\lvert r_{24}\rvert > r_\alpha$	odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1
8.	$H_0: r_{25} = 0$	$H_1: r_{25} \neq 0$	$\lvert r_{25}\rvert \leq r_\alpha$	nie ma podstaw do odrzucenia hipotezy zerowej H_0
9.	$H_0: r_{26} = 0$	$H_1: r_{26} \neq 0$	$\lvert r_{26}\rvert \leq r_\alpha$	nie ma podstaw do odrzucenia hipotezy zerowej H_0
10.	$H_0: r_{34} = 0$	$H_1: r_{34} \neq 0$	$\lvert r_{34}\rvert > r_\alpha$	odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1
11.	$H_0: r_{35} = 0$	$H_1: r_{35} \neq 0$	$\lvert r_{35}\rvert \leq r_\alpha$	nie ma podstaw do odrzucenia hipotezy zerowej H_0
12.	$H_0: r_{36} = 0$	$H_1: r_{36} \neq 0$	$\lvert r_{36}\rvert > r_\alpha$	odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1
13.	$H_0: r_{45} = 0$	$H_1: r_{45} \neq 0$	$\lvert r_{45}\rvert > r_\alpha$	odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1
14.	$H_0: r_{46} = 0$	$H_1: r_{46} \neq 0$	$\lvert r_{46}\rvert > r_\alpha$	odrzucamy hipotezę zerową H_0 na rzecz hipotezy alternatywnej H_1
15.	$H_0: r_{56} = 0$	$H_1: r_{56} \neq 0$	$\lvert r_{56}\rvert \leq r_\alpha$	nie ma podstaw do odrzucenia hipotezy zerowej H_0

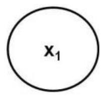
Z dokonanej weryfikacji statystycznej wynika, że sześć hipotez zerowych H_0 odrzucamy na rzecz hipotezy alternatywnej H_1. W dziewięciu przypadkach nie ma podstaw do odrzucenia hipotezy zerowej H_0, a zatem w tworzonej nowej macierzy R' (wzór 6), w tym przypadku wszystkie dotychczasowe wartości współczynników korelacji macierzy R (wzór 1), jako nieistotne statystycznie, zastępujemy zerami. Na tej podstawie utworzono macierz R' z istotnymi statystycznie współczynnikami korelacji dla sześciu zmiennych niezależnych x_1, ..., x_6:

$$R' = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{array}{c} \begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{array} \\ \begin{bmatrix} 1{,}00 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1{,}00 & -0{,}93 & -0{,}99 & 0 & 0 \\ 0 & -0{,}93 & 1{,}00 & 0{,}95 & 0 & 0{,}93 \\ 0 & -0{,}99 & 0{,}95 & 1{,}00 & -0{,}89 & 0{,}88 \\ 0 & 0 & 0 & -0{,}89 & 1{,}00 & 0 \\ 0 & 0 & 0{,}93 & 0{,}88 & 0 & 1{,}00 \end{bmatrix} \end{array} \qquad [6]$$

Podczas analizy macierzy R' kolejnym etapem jest utworzenie stosownych grafów obrazujących wszystkie istotne statystycznie powiązania pomiędzy zmiennymi niezależnymi. Aby łatwiej przeprowadzić analizę macierzy R' i zbudować odpowiedni graf, dobrze jest dodatkowo dołączyć do tej macierzy wiersz i kolumnę ze zmiennymi $x_1, ..., x_6$, pomiędzy którymi łatwiej jest odnaleźć istniejące powiązania zmiennych (wzór 6).

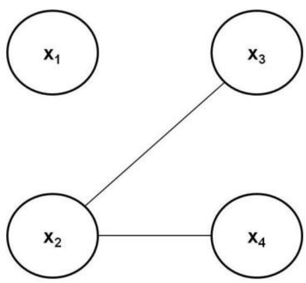
Tworząc graf, analizujemy macierz R' według kolejnych wierszy lub kolumn tej macierzy, szukając powiązań pomiędzy poszczególnymi zmiennymi, poczynając od zmiennej x_1 z kolejnymi zmiennymi $x_2, x_3, ..., x_6$ i tworząc wierzchołki tych zmiennych w konstruowanym grafie (rysunek 5). Tworzymy wierzchołki tylko tych zmiennych, których współczynniki korelacji z innymi zmiennymi w analizowanej macierzy R' (wzór 6) są różne od zera, czyli:

1) w pierwszym wierszu macierzy zmiennej x_1 nie łączymy z żadną ze zmiennych – jest ona zmienną izolowaną (rysunek 2):



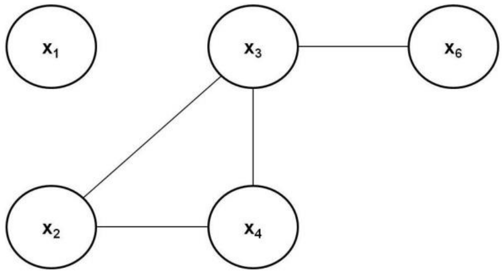
Rys. 2. Zmienna izolowana wynikająca z analizy pierwszego wiersza macierzy R′

2) w drugim wierszu macierzy zmienną x_2 łączymy ze zmienną x_3 i x_4 (rysunek 3):



Rys. 3. Zmienne tworzące wierzchołki grafu po analizie drugiego wiersza macierzy R′

3) w trzecim wierszu macierzy zmienną x_3 łączymy ze zmienną x_2, x_4 i x_6 (rysunek 4):

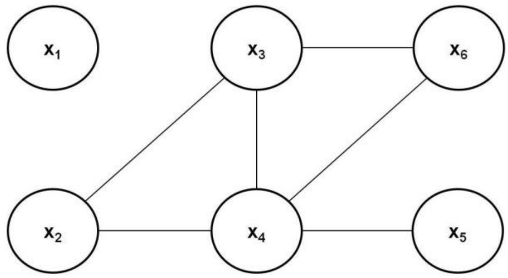


Rys. 4. Zmienne tworzące wierzchołki grafu po analizie trzeciego wiersza macierzy R′

4) w czwartym wierszu macierzy zmienną x_4 łączymy ze zmienną x_2, x_3, x_5 i x_6 (rysunek 5);

5) w piątym wierszu macierzy zmienną x_5 łączymy ze zmienną x_4 (rysunek 5);

6) w szóstym wierszu macierzy zmienną x_6 łączymy ze zmienną x_3, i x_4 (rysunek 5):

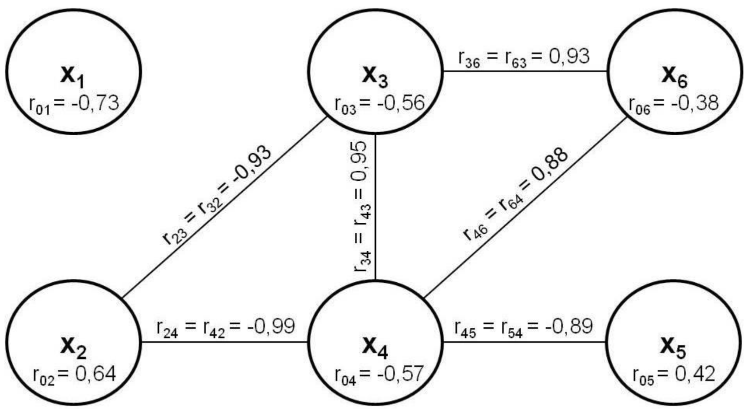


Rys. 5. Zmienne tworzące wierzchołki grafu po analizie czwartego, piątego i szóstego wiersza macierzy R′

Przeprowadzając końcową analizę utworzonej grupy grafów rysunkach 5 i 6, wybieramy do modelu ekonometrycznego następujące zmienne niezależne:

− zmienną x_1, z grafu jednoelementowego, jako zmienną izolowaną;

− zmienną x_4, z grafu kilkuelementowego, jako zmienną o najwyższym stopniu, która ma najwięcej powiązań (łuków) z innymi wierzchołkami.



Rys. 6. Zestawienie zmiennych niezależnych w postaci grupy grafów z istotnymi statystycznie współczynnikami korelacji liniowej Pearsona

Dokonanie wyboru najistotniejszych zmiennych niezależnych x_1 i x_4 metodą analizy grafów, spośród wszystkich analizowanych zmiennych kandydujących, pozwala utworzyć empiryczny liniowy model ekonometryczny z dwiema zmiennymi (wzór 7):

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_4 x_{4t} + \xi_t \qquad [7]$$

Postać funkcyjna modelu może oczywiście przybrać również postać nieliniową (wzór 8), np.: potęgową, logarytmiczną, wykładniczą.

Uzyskane wyniki wyboru zmiennych kandydujących do modelu ekonometrycznego metodą analizy grafów wskazały na zmienne x_1 i x_4. Zmienna x_1 reprezentuje przychody z usług pocztowych, co nie jest zaskoczeniem w jej wyborze, ponieważ wiąże się ona z podstawową działalnością realizowaną przez rozpatrywaną branżę. Natomiast zmienna x_4, reprezentuje przychody z usług udostępniania Internetu, transmisji danych i poczty elektronicznej. Jest to branża od dłuższego czasu bardzo dynamicznie się rozwijająca i z roku na rok zwiększająca przychody (rysunek 1).

Hipoteza badawcza postawiona na wstępie opisywanego przykładu jedynie częściowo została pozytywnie zweryfikowana. Spośród sześciu zmiennych kandydujących tylko dwie najistotniej statystycznie wpływają na przychody ze sprzedaży wyrobów i usług działalności pocztowej i kurierskiej oraz telekomunikacji w Polsce. Zatem w przykładowym modelowaniu procesów logistycznych przy wykorzystaniu metody analizy grafów do doboru zmiennych niezależnych postawiony cel badań empirycznych został osiągnięty, a problem badawczy rozwiązany.

Uwzględniając również merytoryczną analizę danych empirycznych (tabela 1), należy zauważyć, że nie byłoby błędem dołączenie do konstruowanego empirycznego modelu ekonometrycznego trzeciej zmiennej x_5, która reprezentuje branżę o najwyższych rocznych przychodach (mimo zaobserwowanej w ostatnich latach spadkowej tendencji rozwojowej). Mógłby być to zatem także potęgowy empiryczny nieliniowy model ekonometryczny z trzema zmiennymi (wzór 8):

$$y_t = \alpha_0 \cdot x_{1t}^{\alpha_1} \cdot x_{4t}^{\alpha_4} \cdot x_{5t}^{\alpha_5} \cdot 10^{\xi_t} \qquad [8]$$

Dopuszcza się, na tym etapie modelowania procesów logistycznych, uzupełnianie wyjściowych modeli empirycznych o dodatkowe kandydujące zmienne niezależne, np. x_5, na podstawie merytorycznej decyzji badacza, ponieważ w dalszym postępowaniu weryfikacyjnym i tak wszystkie zmienne niezależne modelu będą poddane badaniu statystycznej istotności (parametrów strukturalnych), np. testem t-Studenta.

SELECTION OF INDEPENDENT VARIABLES IN THE MODELLING OF LOGISTICS PROCESSES USING THE ANALYSIS GRAPHS METHOD

Abstract

The theme of the article is the selection of independent variables in the modelling of logistic processes using the analysis graphs method. The main objective is theoretical and empirical analysis of the applicability of the analysis graphs method for the selection of independent variables in econometric models. The research included seeking solutions for obtaining the appropriate criteria for the selection of independent variables of the variables candidate, using properly posed and verified hypotheses.

The unquestionable achievement of the analysis graphs method is rendering it possible to obtain a matrix of correlation Pearson coefficients which illustrate the relationships between the candidate variables and allow these variables to be eliminated, which is not statistically significant for the test of the logistics process.

Key words: analysis graphs method, logistics processes, dependent and independent variables, selection of variables, selection of variables.