ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 2005
Seria: TRANSPORT z. 58 Nr kol. 1688

Jolanta ŻUROWSKA[1]

MODEL LOGITOWY DO SZACOWANIA WIELKOŚCI KOLEJOWEGO POTOKU PASAŻERSKIEGO KRAKÓW – BALICE PORT LOTNICZY

Streszczenie. W artykule zaprezentowano przykład wykorzystania modelu logitowego do określenia prawdopodobieństwa korzystania z planowanego do uruchomienia kolejowego połączenia Kraków Główny - Międzynarodowy Port Lotniczy w Balicach. Przykład poprzedzają rozważania teoretyczne dotyczące szacowania parametrów modelu i badania jego jakości.

LOGIT MODEL FOR ESTIMATION THE VOLUME OF PASSENGER FLOW FROM CRACOW TO BALICE AIRPORT

Summary. The case of logit model application for estimation a probability of usage the planned rail junction from Kraków Główny to International Airport Balice has been presented in the paper. The theoretical considerations on model parameters estimation and research of its quality have been also included.

1. WPROWADZENIE

W przypadku gdy zmienna prognozowana jest zmienną dychotomiczną, która opisuje fakt wystąpienia bądź niewystąpienia prognozowanego zjawiska, lub zmienną opisującą częstość wystąpienia tego zjawiska, to wartości teoretyczne tej zmiennej oraz prognozę można traktować jako oszacowania warunkowych prawdopodobieństw realizacji określonego wariantu (określonej wartości) tej zmiennej przy określonych wartościach zmiennych objaśniających modelu. Do opisu zależności między częstościami występowania poszczególnych wariantów zmiennej objaśnianej a wybranymi zmiennymi objaśniającymi można wykorzystać modele probitowe i logitowe. Są to modele oparte na modelach ekonometrycznych [1, 2, 3, 5].

W przypadku prawdopodobieństw zastosowanie klasycznej metody najmniejszych kwadratów (KMNK) do estymacji parametrów modelu jest niewłaściwe i nieprzydatne, ponieważ nie wyklucza otrzymania wartości teoretycznej prawdopodobieństwa leżącej poza przedziałem [0, 1], co jest sprzeczne z definicją prawdopodobieństwa.

Można jednak dokonać monotonicznego przekształcenia prawdopodobieństw z przedziału [0, 1] na przedział $(-\infty, \infty)$. Skorzystanie z tego przekształcenia wymaga wprowadzenia pewnej liczby kategorii zmiennej objaśnianej, tak aby można było mierzyć częstość

[1] Zakład Organizacji i Ekonomiki Transportu, Politechnika Krakowska, 31-155 Kraków, ul. Warszawska 24, tel. (+48 12) 6283292, jmzur@pk.edu.pl

wystąpienia wariantu zmiennej objaśnianej w każdej kategorii. Najbardziej znanymi przekształceniami są transformacja probitowa i transformacja logitowa.

W obu przypadkach transformacji, zamiast prawdopodobieństw empirycznych określonych na przedziale [0, 1] otrzymuje się jednoznacznie im odpowiadające wartości probitu lub logitu. Wykorzystanie dystrybuanty standaryzowanego rozkładu normalnego do przekształcenia częstości w probity ogranicza możliwości stosowania transformacji probitowej do zmiennych o rozkładzie normalnym lub zbliżonym do normalnego.

2. MODEL LOGITOWY

Transformacja logitowa to przekształcenie poszczególnych prawdopodobieństw empirycznych p_j w logity L_j według formuły [2]:

$$L_j = \log \frac{p_j}{1 - p_j}, \qquad (j = 1,2,\ldots,n), \tag{1}$$

lub wg [1, 3, 5]

$$L_j = \ln \frac{p_j}{1 - p_j}, \qquad (j = 1,2,\ldots,n), \tag{2}$$

gdzie:

j - kategoria zmiennej objaśnianej,

n - liczba wyróżnionych kategorii zmiennej objaśnianej.

Kolejnym krokiem w konstrukcji prognoz prawdopodobieństw za pomocą modeli logitowych jest oszacowanie związków między empirycznymi wartościami logitów a uwzględnianymi zmiennymi objaśniającymi. W przypadku modeli liniowych związki te mają w zapisie macierzowym postać:

$$\hat{\mathbf{L}} = \mathbf{X} \cdot \boldsymbol{\beta} + \xi, \tag{3}$$

gdzie:

$\mathbf{X}$ - macierz zmiennych objaśniających,

$\boldsymbol{\beta}$ - wektor parametrów strukturalnych modelu,

ξ - oznacza wektor składnika losowego.

Do wyznaczenia parametrów modelu należy zastosować uogólnioną metodę najmniejszych kwadratów (UMNK). Wektory ocen parametrów wyznacza się z zależności:

$$\mathbf{b} = \left(\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}^{-1} \mathbf{L} \tag{4}$$

Macierz $\mathbf{W}$ jest macierzą diagonalną, tak więc macierz do niej odwrotna ma postać:

$$\mathbf{W}^{-1} = \begin{bmatrix} \frac{1}{w_1} & 0 & \dots & 0 \\ 0 & \frac{1}{w_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{w_n} \end{bmatrix}, \tag{5}$$

przy czym wartości w_j wyznacza się z zależności:

$$w_j = \frac{1}{n_j \cdot p_j \cdot (1 - p_j)}, \qquad (j = 1,2,\dots,n), \tag{6}$$

gdzie:

p_j - empiryczne prawdopodobieństwa,

n_j - liczność j-tej kategorii zmiennej.

Teoretyczne wartości prawdopodobieństw szacuje się, w zależności od przyjętej definicji logitu, na podstawie następujących wzorów:

- definicja logitu na podstawie wzoru (1):

$$\hat{p}_j = 1 + \frac{10^{\hat{L}_j}}{1 + 10^{\hat{L}_j}} \qquad (j = 1,2,\dots,n), \tag{7}$$

- definicja logitu na podstawie wzoru (2):

$$\hat{p}_j = 1 + \frac{1}{1 + \exp(-\hat{L}_j)} \qquad (j = 1,2,\dots,n). \tag{8}$$

Wariancję resztową modelu szacuje się na podstawie wzoru:

$$s^2 = \frac{1}{n - m - 1} \mathbf{e}^{-1} \mathbf{W}^{-1} \mathbf{e} = \frac{1}{n - m - 1} \sum_{j=1}^{n} e_j^2 \cdot \frac{1}{w_j}, \tag{9}$$

gdzie:

s - odchylenie standardowe reszt,

$\mathbf{e}$ - oszacowany wektor reszt modelu $(e_j = L_j - \hat{L}_j)$,

m - liczba zmiennych objaśniających modelu.

Standardowe błędy ocen parametrów oblicza się na podstawie macierzy wariancji i kowariancji ocen parametrów, którą wyznacza się na podstawie wzoru:

$$\mathbf{D}^2(\mathbf{b}) = s^2 \cdot (\mathbf{X}^T \mathbf{W}^{-1} \mathbf{X})^{-1} \tag{10}$$

Zgodność pomiędzy prawdopodobieństwami empirycznymi (p_j) i oszacowanymi na podstawie modelu $(\hat{p}_j)$ pozwala ocenić współczynnik determinacji, który można wyznaczyć z zależności:

$$R^2 = 1 - \frac{\sum_{j=1}^{m}(p_j - \hat{p}_j)^2}{\sum_{j=1}^{m}(p_j - \bar{p})^2} \tag{11}$$

3. PRZYKŁAD

W 2004 roku, w prowadzonej pod kierunkiem autorki pracy dyplomowej [4], przeprowadzone zostały badania sondażowe mające na celu poznanie opinii mieszkańców obszaru oddziaływania Międzynarodowego Portu Lotniczego w Balicach o istniejących i planowanych połączeniach portu z centrum miasta. Jedno z pytań dotyczyło deklaracji respondentów odnośnie do korzystania z planowanego do uruchomienia połączenia kolejowego. Załączona w kwestionariuszu metryczka pozwoliła wyróżnić różne grupy respondentów w zależności od wieku, miejsca zamieszkania i statusu. Na podstawie wyników otrzymanych z badania oszacowano model umożliwiający określenie prawdopodobieństwa, że osoba danej grupy będzie korzystała z planowanego połączenia. Do tego celu wykorzystano model logitowy dla zmiennej zero-jedynkowej Y, która przyjmuje wartości: 1 – gdy osoba będzie korzystała z połączenia oraz 0 – w przeciwnym przypadku. Najbardziej dopasowany model uzyskano przy przyjęciu dwóch zmiennych objaśniających: X_1 - opisującej miejsce zamieszkania respondenta oraz X_2 - opisującej jego status. W kwestionariuszu respondenci mieli do wyboru 3 odpowiedzi odnośnie do miejsca zamieszkania, a mianowicie: gmina Kraków, województwo małopolskie, inne, oraz 4 odpowiedzi odnośnie do statusu: pasażer linii lotniczych, odprowadzający, pracownik portu, inny. Z uwagi jednak na niezbyt liczną próbę w modelu przyjęto, że:

- X_1 przyjmuje wartości: 1 – jeżeli osoba zamieszkuje gminę Kraków i 0 w przeciwnym przypadku,
- X_2 przyjmuje wartości: 1 – jeżeli osoba jest pasażerem linii lotniczych i 0 w przeciwnym przypadku.

Otrzymano więc 4 grupy respondentów (kategorie zmiennej objaśnianej).

W tablicy 1 zestawiono odpowiedzi respondentów poszczególnych grup na pytanie dotyczące deklaracji korzystania z planowanego połączenia kolejowego do portu lotniczego w Balicach.

Tablica 1

Wyniki badania ankietowego

Grupa	Zamieszkanie	Status	Liczba osób ankietowanych	Liczba osób deklarujących korzystanie z połączenia
1	gmina Kraków	pasażer linii lotniczych	22	15
2	gmina Kraków	inny	68	28
3	inne	pasażer linii lotniczych	14	10
4	inne	inny	21	8
		suma	**125**	**61**

Źródło: opracowanie własne na podstawie wyników badania [4]

Dane liczbowe do modelu wraz z obliczonymi prawdopodobieństwami empirycznymi p_j, logitami L_j (1) oraz wartościami w_j (6) przedstawia tabl. 2.

Tablica 2

Dane liczbowe i obliczenia pomocnicze

j	X_1	X_2	n_j	m_j	p_j	$\frac{p_j}{1-p_j}$	L_j	w_j
1	1	1	22	15	0,6818	2,1429	0,3310	0,2095
2	1	0	68	28	0,4118	0,7000	-0,1549	0,0607
3	0	1	14	10	0,7143	2,5000	0,3979	0,3500
4	0	0	21	8	0,3810	0,6154	-0,2109	0,2019
				średnia	0,5472			

Źródło: obliczenia własne

Macierze X i W^{-1} potrzebne do oszacowania parametrów modelu (3) mają postać:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{W}^{-1} = \begin{bmatrix} 4{,}7727 & 0 & 0 & 0 \\ 0 & 16{,}4706 & 0 & 0 \\ 0 & 0 & 2{,}8671 & 0 \\ 0 & 0 & 0 & 4{,}9524 \end{bmatrix}$$

Wykonując, przy wykorzystaniu funkcji MACIERZ.ILOCZYN i MACIERZ.ODW, obliczenia:

$$\mathbf{X}^{\mathrm{T}} \cdot \mathbf{W}^{-1} \cdot \mathbf{X} = \begin{bmatrix} 29{,}0528 & 21{,}2433 & 7{,}6299 \\ 21{,}2433 & 21{,}2433 & 4{,}7727 \\ 7{,}6299 & 4{,}7727 & 7{,}6299 \end{bmatrix},$$

$$\left(\mathbf{X}^{\mathrm{T}} \cdot \mathbf{W}^{-1} \cdot \mathbf{X}\right)^{-1} = \begin{bmatrix} 0{,}1523 & -0{,}1374 & -0{,}0664 \\ -0{,}1374 & 0{,}1787 & 0{,}0256 \\ -0{,}0664 & 0{,}0256 & 0{,}1814 \end{bmatrix}, \quad \mathbf{X}^{\mathrm{T}} \cdot \mathbf{W}^{-1} \cdot \mathbf{L} = \begin{bmatrix} -0{,}8788 \\ -0{,}9716 \\ 2{,}7167 \end{bmatrix},$$

otrzymano wektor **b** ocen parametrów:

$$\mathbf{b} = \begin{bmatrix} -0{,}1807 \\ 0{,}0167 \\ 0{,}5263 \end{bmatrix}$$

Tak więc oszacowany model ma postać:

$$\hat{L}_j = -0{,}1807 + 0{,}0167 \cdot X_{1j} + 0{,}5263 \cdot X_{2j} + \xi$$

Wyniki oszacowania modelu przedstawiono w tabl. 3. Zawiera ona wartości teoretyczne logitów, prawdopodobieństwa teoretyczne (7) oraz obliczenia pomocnicze do oszacowania wariancji resztowej i współczynnika determinacji.

Tablica 3

Wyniki oszacowania modeli i obliczenia pomocnicze

$\hat{L}_j$	$\hat{p}_j$	$e_j = L_j - \hat{L}_j$	$1/w_j$	$e_j^2 \cdot 1/w_j$	$p_j - \overline{p}$	$p_j - \hat{p}_j$
0,3623	0,6973	-0,0313	4,7727	0,0047	0,1346	-0,0154
-0,1640	0,4067	0,0091	16,4706	0,0014	-0,1354	0,0051
0,3456	0,6891	0,0523	2,8571	0,0078	0,1671	0,0252
-0,1807	0,3975	-0,0302	4,9524	0,0045	-0,1663	-0,0165
			suma	0,0184		
sumy kwadratów		0,0047			0,0920	0,0012

Źródło: obliczenia własne

Na podstawie oszacowanego modelu prawdopodobieństwo korzystania z kolejowego połączenia jest najwyższe w grupie pasażerów linii lotniczych zamieszkałych w gminie Kraków.

Wariancja resztowa (9) oraz współczynnik determinacji (11) wynoszą:

$$s^2 = \frac{0{,}0184}{4-2-1} = 0{,}0184$$

$$R^2 = 1 - \frac{0{,}0012}{0{,}0920} = 0{,}9873$$

Wartość współczynnika determinacji świadczy o dobrym dopasowaniu modelu.

4. PODSUMOWANIE

W ostatnich latach w transporcie przeprowadza się często badania marketingowe. Przedstawiony przykład pokazuje możliwość wykorzystania danych uzyskanych z tych badań do prognozowania.

Literatura

1. Dittmann P.: Prognozowanie w przedsiębiorstwie. Metody i ich zastosowanie. Oficyna Ekonomiczna, Kraków 2003.
2. Prognozowanie gospodarcze. Metody, modele, zastosowania, przykłady. Redakcja Nowak E. Agencja Wydawnicza PLACET, Warszawa 1998.
3. Prognozowanie gospodarcze. Metody i zastosowania. Redakcja Cieślak M. Wydanie drugie, rozszerzone i poprawione. Wydawnictwo Naukowe PWN, Warszawa 2002.
4. Suder P.: Badanie opinii potencjalnych pasażerów odnośnie projektowanego kolejowego połączenia Kraków Główny – Balice. Praca dyplomowa magisterska, Politechnika Krakowska 2005.
5. Zeliaś A., Pawełek B., Wanat S.: Prognozowanie ekonomiczne. Teoria, przykłady, zadania. Wydawnictwo Naukowe PWN, Warszawa 2003.