

**PRACE NAUKOWE**

Uniwersytetu Ekonomicznego we Wrocławiu

**RESEARCH PAPERS**

of Wrocław University of Economics

**278**

# **Taksonomia 20**

## **Klasyfikacja i analiza danych – teoria i zastosowania**



Redaktorzy naukowi

**Krzysztof Jajuga**

**Marek Walesiak**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

[www.ibuk.pl](http://www.ibuk.pl), [www.ebscohost.com](http://www.ebscohost.com),

The Central and Eastern European Online Library [www.ceeol.com](http://www.ceeol.com),

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

[http://kangur.uek.krakow.pl/bazy\\_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

[www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2013

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

## Spis treści

<b>Wstęp</b> .....	9
<b>Józef Pocięcha:</b> Wskaźniki finansowe a klasyfikacyjne modele predykcji upadłości firm .....	15
<b>Eugeniusz Gatnar:</b> Analiza miar adekwatności rezerw walutowych .....	23
<b>Marek Walesiak:</b> Zagadnienie doboru liczby klas w klasyfikacji spektralnej .....	33
<b>Joanicjusz Nazarko, Joanna Ejdyś, Anna Kononiuk, Anna M. Olszewska:</b> Analiza strukturalna jako metoda klasyfikacji danych w badaniach foresight .....	44
<b>Andrzej Bąk:</b> Metody porządkowania liniowego w polskiej taksonomii – pakiet <code>pllord</code> .....	54
<b>Aleksandra Łuczak, Feliks Wysocki:</b> Zastosowanie mediany przestrzennej Webera i metody TOPSIS w ujęciu pozycyjnym do konstrukcji syntetycznego miernika poziomu życia .....	63
<b>Ewa Roszkowska:</b> Zastosowanie rozmytej metody TOPSIS do oceny ofert negocjacyjnych .....	74
<b>Jacek Batóg:</b> Analiza wrażliwości metody ELECTRE III na obserwacje nietypowe i zmianę wartości progowych .....	85
<b>Jerzy Korzeniewski:</b> Modyfikacja metody HINoV selekcji zmiennych w analizie skupień .....	93
<b>Małgorzata Markowska, Danuta Strahl:</b> Wykorzystanie referencyjnego systemu granicznego do klasyfikacji europejskiej przestrzeni regionalnej ze względu na filar inteligentnego rozwoju – kreatywne regiony .....	101
<b>Elżbieta Sobczak:</b> Inteligentne struktury pracujących a efekty strukturalne zmian zatrudnienia w państwach Unii Europejskiej.....	111
<b>Elżbieta Gołata, Grażyna Dehnel:</b> Rozbieżności szacunków NSP 2011 i BAEL.....	120
<b>Iwona Foryś:</b> Wykorzystanie analizy historii zdarzeń do badania powtórnego sprzedaży na lokalnym rynku mieszkaniowym .....	131
<b>Hanna Dudek, Joanna Landmesser:</b> Wpływ relatywnej deprivacji na subiektywne postrzeganie dochodów.....	142
<b>Grażyna Łaska:</b> Syntaksonomia numeryczna w klasyfikacji, identyfikacji i analizie przemian zbiorowisk roślinnych .....	151
<b>Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz:</b> Analiza zależności między procesami fundamentalnymi a rynkiem kapitałowym w Chinach .....	161

<b>Andrzej Bąk, Tomasz Bartłomowicz:</b> Mikroekonometryczne modele wielomianowe i ich zastosowanie w analizie preferencji z wykorzystaniem programu R .....	169
<b>Andrzej Dudek, Bartosz Kwaśniewski:</b> Przetwarzanie równoległe algorytmów analizy skupień w technologii CUDA .....	180
<b>Michał Trzęsiok:</b> Wycena rynkowej wartości nieruchomości z wykorzystaniem wybranych metod wielowymiarowej analizy statystycznej .....	188
<b>Joanna Trzęsiok:</b> Wybrane symulacyjne techniki porównywania nieparametrycznych metod regresji.....	197
<b>Artur Mikulec:</b> Kryterium Mojeny i Wisharta w analizie skupień – przypadek skupień o różnych macierzach kowariancji .....	206
<b>Artur Zaborski:</b> Analiza <i>unfolding</i> z wykorzystaniem modelu grawitacji ....	216
<b>Justyna Wilk:</b> Identyfikacja obszarów problemowych i wzrostowych w województwie dolnośląskim w zakresie kapitału ludzkiego .....	225
<b>Karolina Bartos:</b> Analiza ryzyka odejścia studenta z uczelni po uzyskaniu dyplomu licencjata – zastosowanie sieci MLP .....	236
<b>Ewa Genge:</b> Segmentacja uczestników Industriady z wykorzystaniem analizy klas ukrytych .....	246
<b>Izabela Kurzawa:</b> Wielomianowy model logitowy jako narzędzie identyfikacji czynników wpływających na sytuację mieszkaniową polskich gospodarstw domowych .....	254
<b>Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej:</b> Modele eksploracji danych niezbilansowanych – procedury klasyfikacji dla zadania analizy ryzyka operacyjnego.....	262
<b>Aleksandra Łuczak:</b> Zastosowanie rozmytej hierarchicznej analizy w tworzeniu strategii rozwoju jednostek administracyjnych .....	271
<b>Marcin Pelka:</b> Rozmyta klasyfikacja spektralna <i>c</i> -średnich dla danych symbolicznych interwałowych .....	282
<b>Małgorzata Machowska-Szewczyk:</b> Klasyfikacja obiektów reprezentowanych przez różnego rodzaju cechy symboliczne .....	290
<b>Ewa Chodakowska:</b> Indeks Malmquista w klasyfikacji podmiotów gospodarczych według zmian ich względnej produktywności działania .....	300
<b>Beata Bieszk-Stolorz, Iwona Markowicz:</b> Wykorzystanie modeli proporcjonalnego i nieproporcjonalnego hazardu Coxa do badania szansy podjęcia pracy w zależności od rodzaju bezrobocia .....	311
<b>Marcin Salamaga:</b> Weryfikacja teorii poziomego rozwoju gospodarczego J.H. Dunninga w ujęciu sektorowym w wybranych krajach Unii Europejskiej .....	321
<b>Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik:</b> Sytuacja społeczno-gospodarcza jako determinanta migracji wewnętrznych w Polsce. ....	330
<b>Hanna Gruchociak:</b> Delimitacja lokalnych rynków pracy w Polsce na podstawie danych z badania przepływów ludności związanych z zatrudnieniem .....	343

<b>Radosław Pietrzyk:</b> Efektywność inwestycji polskich funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych .....	351
<b>Sabina Denkowska:</b> Procedury testowań wielokrotnych .....	362

## Summaries

<b>Józef Pocięcha:</b> Financial ratios and classification models of bankruptcy prediction .....	22
<b>Eugeniusz Gatnar:</b> Analysis of FX reserve adequacy measures .....	32
<b>Marek Walesiak:</b> Automatic determination of the number of clusters using spectral clustering .....	43
<b>Joanicjusz Nazarko, Joanna Ejdys, Anna Kononiuk, Anna M. Olszewska:</b> Structural analysis as a method of data classification in foresight research .....	53
<b>Andrzej Bąk:</b> Linear ordering methods in Polish taxonomy – pllord package .....	62
<b>Aleksandra Łuczak, Feliks Wysocki:</b> The application of spatial median of Weber and the method TOPSIS in positional formulation for the construction of synthetic measure of standard of living .....	73
<b>Ewa Roszkowska:</b> Application of the fuzzy TOPSIS method to the estimation of negotiation offers.....	84
<b>Jacek Batóg:</b> Sensitivity analysis of ELECTRE III method for outliers and change of thresholds .....	92
<b>Jerzy Korzeniewski:</b> Modification of the HINoV method of selecting variables in cluster analysis .....	100
<b>Małgorzata Markowska, Danuta Strahl:</b> Implementation of reference limit system for the European regional space classification regarding smart growth pillar – creative regions .....	110
<b>Elżbieta Sobczak:</b> Smart workforce structures versus structural effects of employment changes in the European Union countries .....	119
<b>Elżbieta Gołata, Grażyna Dehnel:</b> Divergence in National Census 2011 and LFS estimates.....	130
<b>Iwona Foryś:</b> Event history analysis in the resale study on the local housing market .....	141
<b>Hanna Dudek, Joanna Landmesser:</b> Impact of the relative deprivation on subjective income satisfaction .....	150
<b>Grażyna Łaska:</b> Numerical syntaxonomy in classification, identification and analysis of changes of secondary communities .....	160
<b>Magdalena Osińska, Marcin Faldziński, Tomasz Zdanowicz:</b> Analysis of relations between fundamental processes and capital market in China.....	166
<b>Andrzej Bąk, Tomasz Bartłomowicz:</b> Microeconomic polynomial models and their application in the analysis of preferences using R program.....	179

<b>Andrzej Dudek, Bartosz Kwaśniewski:</b> Parallel processing of clustering algorithms in CUDA technology .....	187
<b>Michał Trzęsiok:</b> Real estate market value estimation based on multivariate statistical analysis .....	196
<b>Joanna Trzęsiok:</b> On some simulative procedures for comparing nonparametric methods of regression.....	205
<b>Artur Mikulec:</b> Mojena and Wishart criterion in cluster analysis – the case of clusters with different covariance matrices .....	215
<b>Artur Zaborski:</b> Unfolding analysis by using gravity model .....	224
<b>Justyna Wilk:</b> Determination of problem and growth areas in Dolnośląskie Voivodship as regards human capital.....	235
<b>Karolina Bartos:</b> Risk analysis of bachelor students' university abandonment – the use of MLP networks .....	245
<b>Ewa Genge:</b> Clustering of industrial holiday participants with the use of latent class analysis.....	253
<b>Izabela Kurzawa:</b> Multinomial logit model as a tool to identify the factors affecting the housing situation of Polish households.....	261
<b>Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Jerzy Kołodziej:</b> Modelling class imbalance problems: comparing classification approaches for surgical risk analysis .....	270
<b>Aleksandra Łuczak:</b> The application of fuzzy hierarchical analysis to the evaluation of validity of strategic factors in administrative districts.....	281
<b>Marcin Pełka:</b> A spectral fuzzy c-means clustering algorithm for interval-valued symbolic data .....	289
<b>Małgorzata Machowska-Szewczyk:</b> Clustering algorithms for mixed-feature symbolic objects .....	299
<b>Ewa Chodakowska:</b> Malmquist index in enterprises classification on the basis of relative productivity changes .....	310
<b>Beata Bieszk-Stolorz, Iwona Markowicz:</b> Using proportional and non proportional Cox hazard models to research the chances for taking up a job according to the type of unemployment .....	320
<b>Marcin Salamaga:</b> Verification J.H. Dunning's theory of economic development by economic sectors in some EU countries .....	329
<b>Justyna Wilk, Michał Bernard Pietrzak, Stanisław Matusik:</b> Socio-economic situation as a determinant of internal migration in Poland .....	342
<b>Hanna Gruchociak:</b> Delimitation of local labor markets in Poland on the basis of the employment-related population flows research.....	350
<b>Radosław Pietrzyk:</b> Selectivity and timing in Polish mutual funds performance measurement .....	361
<b>Sabina Denkowska:</b> Multiple testing procedures.....	369

**Andrzej Bąk, Tomasz Bartłomowicz**

Uniwersytet Ekonomiczny we Wrocławiu

---

## **MIKROEKONOMETRYCZNE MODELE WIELOMIANOWE I ICH ZASTOSOWANIE W ANALIZIE PREFERENCJI Z WYKORZYSTANIEM PROGRAMU R**

---

**Streszczenie:** Celem artykułu jest przedstawienie wybranych wielomianowych modeli logitowych wyborów dyskretnych i ich zastosowanie w pomiarze preferencji konsumentów. Podstawą do rozróżniania typów modeli wielomianowych jest głównie charakter zmiennych objaśniających w modelu. Ponieważ to rozróżnienie nie jest należycie interpretowane w literaturze przedmiotu, w artykule przedstawia się podstawowe różnice między wielomianowymi modelami logitowymi stosowanymi w analizie preferencji konsumentów. W estymacji modeli wykorzystano program, pakiety R i funkcje oprogramowane w języku R.

**Słowa kluczowe:** preferencje, wielomianowe model logitowe, program R.

### **1. Wstęp**

Jednym z najważniejszych elementów badań marketingowych, umożliwiającym wyjaśnienie przyczyn wyborów konsumenckich jest pomiar oraz analiza preferencji konsumentów. Podstawą teoretyczną tego typu badań są teorie mieszczące się w obrębie mikroekonomii, m.in. teoria addytywnego jednoczesnego pomiaru łącznego [Coombs, Dawes, Tversky 1977, s. 50] oraz teoria użyteczności losowej [Coombs, Dawes, Tversky 1977, s. 214], które umożliwiają kwantyfikację użyteczności, a w konsekwencji identyfikację czynników, którymi kierują się konsumenci, wybierając określone produkty lub usługi. Czynniki te mogą być związane zarówno z charakterystykami konsumentów, jak i atrybutami produktów lub usług. Oznacza to, iż empiryczne badania preferencji konsumentów opierają się na danych o jednostkowych obiektach badania, które w literaturze przedmiotu określa się mianem mikrodanych [Gruszczyński 2002].

Metody wykorzystywane do pomiaru preferencji umożliwiają „wydobycie” zawartych w mikrodanych informacji. Narzędziami mikroekonometrii są w tym przypadku metody i modele wyborów dyskretnych, do których należą modele kategorii nieuporządkowanych [Greene 2008, s. 840-847]. Najczęściej stosowane mikroeko-

nometryczne modele kategorii nieuporządkowanych to wielomianowy model logitowy, warunkowy model logitowy [McFadden, 1974] oraz mieszany model logitowy [Winkelmann, Boes 2006], których podstawą rozróżnienia jest głównie charakter zmiennych objaśniających modelu. Ponieważ w literaturze przedmiotu rozróżnienie to nie jest jednoznacznie interpretowane, w artykule przedstawiono podstawowe różnice między wyróżnionymi rodzajami modeli logitowych [Bąk 2012]. Ponadto celem artykułu jest wskazanie możliwości estymacji wielomianowych modeli logitowych kategorii nieuporządkowanych dla różnych zbiorów danych z wykorzystaniem programu, pakietów R oraz funkcji oprogramowanych w języku R.

## 2. Wielomianowe modele logitowe kategorii nieuporządkowanych

W przypadku kategorii nieuporządkowanych w badaniach marketingowych zastosowanie znajdują mikroekonometryczne modele wyborów dyskretnych w postaci wielomianowego modelu logitowego, warunkowego modelu logitowego oraz modelu będącego połączeniem wymienionych rodzajów modeli pod postacią tzw. mieszanego (hybrydowego) modelu logitowego [Cameron, Trivedi 2009, s. 500].

Wielomianowy model logitowy jest uogólnieniem modelu logitowego dla danych binarnych i może być stosowany, gdy zmienna objaśniana przyjmuje w sposób dyskretny wartości ze zbioru liczącego więcej niż dwie kategorie. Model wywodzi się z teorii użyteczności losowej oraz tzw. aksjomatu wyboru Luce'a (modelu stałej użyteczności) [Coombs, Dawes, Tversky 1977, s. 217; Bierlaire 1997]. Wielomianowy model logitowy można przedstawić w postaci [So, Kuhfeld 1995; Long 1997, s. 151; Powers, Xie 2008, s. 243; Cameron, Trivedi 2009, s. 500; Gruszczyński (red.) 2010, s. 161]:

$$P_{ki} = \frac{\exp(x_k^T \beta_i)}{\sum_{l=1}^n \exp(x_k^T \beta_l)}, \quad (1)$$

gdzie:  $P_{ki}$  – prawdopodobieństwo wyboru  $i$ -tej kategorii przy  $k$ -tym stanie zmiennych objaśniających opisujących konsumentów;

$x_k^T$  – wektor reprezentujący  $k$ -ty wiersz macierzy  $\mathbf{X}$  (wartości zmiennych objaśniających dla  $k$ -tego konsumenta);

$\beta_i$  – wektor parametrów związany z  $i$ -tą kategorią zmiennej objaśnianej.

Oszacowane wartości prawdopodobieństw w modelu (1) sumują się do jedności w obrębie każdej konfiguracji zmiennych objaśniających. Rozkład prawdopodobieństw można zatem uzyskać przy różnych wartościach parametrów  $\beta_i$ , co oznacza, iż model pozostaje zdefiniowany niejednoznacznie. W celu rozwiązania tego



problemu przyjmuje się pewne ograniczenia dotyczące wektora parametrów określone normalizacją, zakładając np.  $\beta_n = 0$ . Wówczas prawdopodobieństwo wyboru  $i$ -tej kategorii przy  $k$ -tym stanie zmiennych objaśniających określa zależność:

$$\beta_n = 0 \Rightarrow \exp(x_k^T \beta_i) = \exp(0) = 1, \text{ stąd } P_{ki} = \frac{1}{\sum_{l=1}^n \exp(x_k^T \beta_l)}. \quad (2)$$

Oznacza to, iż jedna z opcji wyboru (np. ostatnia) stanowi profil odniesienia, a pozostałe profile są różne od tej opcji [Agresti 2002, s. 268].

Jak wynika z zależności (1) oraz (2), macierz  $\mathbf{X}$  zawierająca charakterystyki konsumentów jest stała w przekroju profilów produktów lub usług. Konsekwencją tego jest szacowanie  $n - 1$  współczynników  $\beta_l$  dla każdej zmiennej objaśniającej, które prezentują efekt wpływu poszczególnych zmiennych na prawdopodobieństwo wyboru poszczególnych profilów produktów lub usług w relacji do (czyli bez) wspomnianego profilu odniesienia.

W przypadku warunkowego modelu logitowego zaproponowanego przez McFaddena [1974, s. 105-142] prawdopodobieństwo wyboru  $i$ -tego profilu ze zbioru liczącego  $n$  elementów jest szacowane na podstawie zależności [So, Kuhfeld 1995, s. 7; Long 1997, s. 178; Powers, Xie 2008, s. 256; Cameron, Trivedi 2009, s. 500; Gruszczyński (red.) 2010, s. 172-173]:

$$P_{ki} = \frac{\exp(z_{ki}^T \alpha)}{\sum_{l=1}^n \exp(z_{kl}^T \alpha)}, \quad (3)$$

gdzie:  $P_{ki}$  – prawdopodobieństwo wyboru  $i$ -tej kategorii przy  $k$ -tym stanie zmiennych objaśniających;

$z_{kl}^T$  –  $k$ -ty wektor macierzy  $\mathbf{Z}$  (zmiennych objaśniających opisujących  $i$ -tą opcję wybraną przez  $k$ -tego konsumenta);

$\alpha$  – wektor parametrów.

Jak wynika z zależności (3), zmienne objaśniające z macierzy  $\mathbf{Z}$  przyjmują różne wartości dla każdej z opcji wybieranych przez konsumentów. Jednocześnie w modelu warunkowym jest szacowany tylko jeden współczynnik  $\alpha$  dla każdej zmiennej objaśniającej z macierzy  $\mathbf{Z}$ . Wpływ zmiennych objaśniających na prawdopodobieństwa wyboru opcji wynika z różnic między wartościami tych zmiennych w przekroju opcji [Hoffman, Duncan 1988, s. 415-427].

Zarówno wielomianowy, jak i warunkowy model logitowy są wykorzystywane do analizy wyborów indywidualnych ze zbioru dostępnych alternatyw (kategorii, opcji, profilów), przy czym w modelu wielomianowym (1) ocenia się prawdopodobieństwo wyboru  $i$ -tej kategorii (przy  $k$ -tym stanie zmiennych objaśniających

charakteryzujących konsumentów), podczas gdy w modelu warunkowym (3) jest to prawdopodobieństwo wyboru  $i$ -tego profilu (ze zbioru liczącego  $n$  elementów). Wynika to z założenia – co stanowi jednocześnie główne rozróżnienie tych modeli – iż wielomianowy model logitowy skupia się na jednostce analizy (konsumentcie) i wykorzystuje w postaci zmiennych objaśniających indywidualne predyspozycje jednostki (charakterystyki konsumentów), w przeciwieństwie do warunkowego modelu logitowego, który skupia się na zestawie opcji dla poszczególnych jednostek (profilach produktów lub usług), a zmienne objaśniające stanowią w tym modelu charakterystyki tych opcji (atrzybuty produktów lub usług).

W mieszanym modelu logitowym (4), określanym w literaturze przedmiotu mianem hybrydowego, prawdopodobieństwo wyboru  $i$ -tej kategorii przy  $k$ -tym stanie zmiennych objaśniających określa zależność:

$$P_{ki} = \frac{\exp(x_k^T \beta_i + z_{ki}^T \alpha)}{\sum_{l=1}^n \exp(x_k^T \beta_l + z_{kl}^T \alpha)}. \quad (4)$$

Model mieszany uwzględnia obok jednostki analizy (konsumenta) zbiór alternatyw (opcji wyboru) dla poszczególnych jednostek. Oznacza to, iż macierze zmiennych objaśniających uwzględniają zarówno indywidualne predyspozycje jednostki (charakterystyki konsumentów), jak i charakterystyki alternatyw (atrzybuty produktów lub usług).

### 3. Estymacja parametrów wielomianowych modeli logitowych w programie R

W programie R w szacowaniu parametrów wielomianowych, warunkowych oraz mieszanych modeli logitowych wykorzystuje się funkcję `optim()` z pakietu `stats`. Choć funkcja ta z założenia jest narzędziem „ogólnego przeznaczenia”, funkcję `optim()` można zastosować do maksymalizacji funkcji największej wiarygodności<sup>1</sup>, która umożliwia znalezienie najlepszego dopasowania modelu logitowego do danych empirycznych [Jackman 2007]. Kryterium tego dopasowania jest wartość funkcji wiarygodności. Składnia oraz wybrane argumenty funkcji `optim()` są następujące [R Development Core Team 2011]:

```
optim(par, fn, gr=NULL, x, y, method=c("Nelder-Mead", "BFGS",
"CG", "L-BFGS-B", "SANN"), lower=-Inf, upper=Inf, con-
trol=list(), hessian=FALSE)
```

`par` – początkowe wartości parametrów do optymalizacji,

<sup>1</sup> Funkcja `optim()` korzysta z iteracyjnych algorytmów optymalizacji: sympleksu (Nelder-Meada), zmiennej metryki (quasi-Newtona, Broydena-Fletcher-Goldfarba-Shannona), gradientów sprzężonych (Fletcher-Reevesa), quasi-Newtona z ograniczeniami (algorytmu L-BFGS-B), sieci neuronowych (SNN – Simulated Neural Network).

`fn` – funkcja, której wartość jest optymalizowana,  
`x`, `y` – zmienne objaśniające i zmienna zależna,  
`gr` – gradient (wektor pochodnych cząstkowych),  
`lower`, `upper` – granice zmiennych,  
`control` – lista parametrów kontrolnych (gdy `control=list(fnscale=-1)`,  
funkcja największej wiarygodności jest maksymalizowana).

Jednocześnie funkcja największej wiarygodności `fnw()` przekazywana jako argument `fn` do funkcji `optim()` w celu oszacowania parametrów modeli logitowych ma postać:

```
#fnw(a,x,y) - log-likelihood function for logit model
#based on: S. Jackman (2007)
#a - initial values of parameters
#x - explanatory variables
#y - dependent variable with values TRUE/FALSE
#call as parameter fn in optim() function
fnw<-function(a,x,y) {
mu<-x%*%a #systematic component
eta<-exp(mu) #numerator
suma<-tapply(eta,s,sum) #denominator
pr<-eta[y]/suma #probabilities
lnw<-sum(log(pr)) #sum of the log probabilities
return(lnw) }.
```

#### 4. Przykład zastosowania wielomianowych modeli logitowych

W przykładzie ilustrującym sposób wykorzystania wybranych wielomianowych modeli logitowych identyfikacja i analiza preferencji respondentów dotyczy usług gastronomicznych. W badaniu wytypowano 4 atrybuty (wraz z odpowiadającymi im poziomami): cenę (do 10 zł, 10-20 zł, powyżej 20 zł), miejsce konsumpcji (bar, restaurację, stołówkę, punkt gastronomiczny), rodzaj konsumpcji (posiłek, deser, napój) oraz godzinę konsumpcji (poranną, popołudniową, wieczorną). Badanie przeprowadzono wśród mieszkańców Jeleniej Góry i okolic w roku 2010, co ostatecznie umożliwiło wykorzystanie danych<sup>2</sup> ze 136 prawidłowo wypełnionych kwestionariuszy ankietowych.

Liczba zmiennych wraz z liczbą poziomów ich realizacji oznacza, iż w przykładzie można było wykorzystać maksymalnie 108 różnych profilów usług gastronomicznych. Ostatecznie, co wynika z zastosowanej metody wyborów dyskretnych, na potrzeby badania wytypowano 3 zbiory danych po 6 profilów, w których na ostatniej pozycji znalazł się profil odniesienia. Wybrany zbiór ocenianych przez respondentów profilów prezentuje tab. 1.

---

<sup>2</sup> Dane zostały zebrane przez p. Martę Więclaw.

Tabela 1. Przykładowy zbiór profilów do wyboru

Cena	Miejsce konsumpcji	Rodzaj konsumpcji	Godzina konsumpcji	Wybieram
10-20 zł	bar	deser	poranna	1
do 10 zł	restauracja	napój	poranna	2
do 10 zł	stołówka	deser	popołudniowa	3
powyżej 20 zł	stołówka	napój	popołudniowa	4
10-20 zł	stołówka	posiłek	wieczorna	5
żaden z profilów				6

Źródło: opracowanie własne.

Respondenci za każdym razem wybierali 1 z 6 profilów. Zmienną specyficzną dla respondenta jest dochód ( $doch$ ), wybraną opcję wyboru określa cena ( $cena$ ), natomiast wybrany profil reprezentuje zmienna ( $wybór$ ).

W przypadku wielomianowego modelu logitowego w analizie wykorzystano pojedynczy zbiór profilów do wyboru, co oznacza, iż liczba obserwacji odpowiada liczbie respondentów  $136 (136 \cdot 1)$ , natomiast liczba wierszy w zbiorze danych to  $816$  wierszy ( $136 \cdot 6 \cdot 1$ ), których fragment przedstawia się następująco:

osoba	opcja	wybór	wyr1	wyr2	wyr3	wyr4	wyr5	doch1	doch2	doch3	doch4	doch5
1	1	1	0	1	0	0	0	4	0	0	0	0
2	1	2	1	0	1	0	0	0	4	0	0	0
3	1	3	0	0	0	1	0	0	0	4	0	0
4	1	4	0	0	0	0	1	0	0	0	4	0
5	1	5	0	0	0	0	0	1	0	0	0	4
6	1	6	0	0	0	0	0	0	0	0	0	0
7	2	1	0	1	0	0	0	3	0	0	0	0
8	2	2	0	0	1	0	0	0	3	0	0	0
9	2	3	1	0	0	1	0	0	0	3	0	0
10	2	4	0	0	0	0	1	0	0	0	3	0
11	2	5	0	0	0	0	0	1	0	0	0	3
12	2	6	0	0	0	0	0	0	0	0	0	0
13	3	1	0	1	0	0	0	3	0	0	0	0

W strukturze tych danych każdy wiersz przedstawia 1 opcję wyboru, przy czym pełny zbiór, z którego respondent wybrał jedną opcję (profil), tworzy 6 wierszy. Opcje wyboru są w zbiorze reprezentowane przez zmienną zero-jedynkową; profilem odniesienia jest opcja nr 6.

W wyniku zastosowania odpowiedniego skryptu języka R w wielomianowym modelu logitowym otrzymuje się oszacowania parametrów  $B$  dla charakterystycznej dla respondentów zmiennej specyficzej ( $dochód$ ):

	B	se	Z
wyr1	0.323600	1.5690	2.062e-01
wyr2	0.942500	0.7846	1.201e+00

```

wyr3 -0.735100 1.0990 -6.691e-01
wyr4 -0.693100 1.1540 -6.007e-01
wyr5 0.415800 0.8756 4.749e-01
doch1 -0.948700 0.6818 -1.391e+00
doch2 0.003285 0.2555 1.286e-02
doch3 0.095800 0.3520 2.722e-01
doch4 -0.000022 0.3758 -5.854e-05
doch5 -0.048350 0.2868 -1.686e-01

```

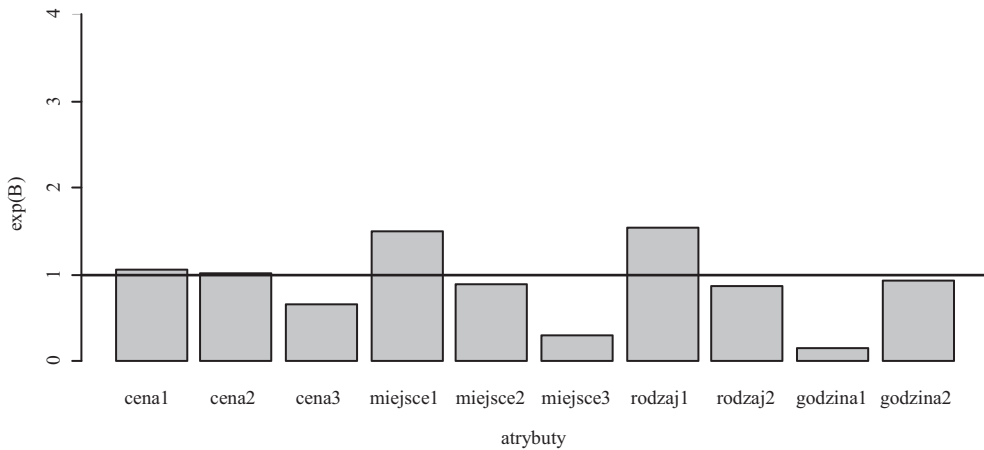
Uzyskany zestaw wyników (5 wyrazów wolnych (wyr3) oraz 5 parametrów dla zmiennej specyficznej (dochód) reprezentuje efekt wpływający na prawdopodobieństwo wyboru opcji (profilu) 1–5 w stosunku do profilu odniesienia oznaczonego numerem 6.

W przypadku warunkowego modelu logitowego w analizie wykorzystano wszystkie 3 zbiory profili do wyboru, co oznacza, iż liczba obserwacji odpowiada liczbie respondentów pomnożonej przez 3 ( $136 \cdot 3$ ), co daje 408 obserwacji, natomiast liczba wierszy w zbiorze danych to 2448 wierszy ( $136 \cdot 6 \cdot 3$ ). Fragment zbioru danych przedstawia się następująco:

wybór	cena1	cena2	cena3	miejs1	miejs2	miejs3	rodzaj1	rodzaj2	godz1	godz2
1	0	0	1	0	0	0	0	1	1	0
2	0	1	0	0	1	0	0	0	1	0
3	0	1	0	0	0	0	1	0	1	1
4	1	0	0	1	0	0	1	0	0	1
5	0	0	1	0	0	0	1	1	0	0
6	0	0	0	0	0	0	0	0	0	0

W wyniku zastosowania odpowiednich skryptów języka R w warunkowym modelu logitowym otrzymuje się oszacowania parametrów B dla charakterystycznych dla opcji wyboru wybranych wariantów zmiennych specyficznych (cena), (miejsce), (rodzaj), (godzina):

	B	se	Z	exp (B)
cena1	0.06252	0.2710	0.23070	1.0650
cena2	0.01121	0.2982	0.03759	1.0110
cena3	-0.42180	0.2426	-1.73900	0.6558
miejs1	0.41170	0.1849	2.22700	1.5090
miejs2	-0.12190	0.2679	-0.45510	0.8852
miejs3	-1.18500	0.2622	-4.51900	0.3058
rodzaj1	0.42890	0.2171	1.97600	1.5360
rodzaj2	-0.14410	0.1637	-0.87980	0.8658
godz1	-1.90900	0.3151	-6.05800	0.1482
godz2	-0.07132	0.1431	-0.49830	0.9312



**Rys. 1.** Wpływ zmiennych na wybór profilów

Źródło: opracowanie własne z wykorzystaniem programu R.

Interpretację wyników z wykorzystaniem ilorazów hazardu przedstawiono na rys. 1. Wartości ilorazów hazardu większe od 1 wskazują na stymulujący wpływ zmiennych na wybór profilów. Na podstawie wartości parametrów obliczono prawdopodobieństwa wyboru profilów, a uzyskane wartości wskazują, że najbardziej preferowany jest profil nr 14, następnie profile o numerze 11, 7 itd.:

Profil	Prawdopodobieństwo wyboru	
[1,]	14	0.155407827
[2,]	11	0.108690722
[3,]	7	0.108384655
[4,]	6	0.071208817
[5,]	12	0.071208817
[6,]	18	0.071208817
[7,]	17	0.067102958
[8,]	10	0.059358560
[9,]	8	0.056830329
[10,]	9	0.051394612
[11,]	15	0.043486641
[12,]	16	0.040435556
[13,]	5	0.033819156
[14,]	3	0.018691963
[15,]	2	0.016961742
[16,]	4	0.013300332
[17,]	1	0.009243355
[18,]	13	0.003265142

W przypadku mieszanego modelu logitowego w analizie wykorzystano 1 zbiór profili do wyboru, co oznacza, iż liczba obserwacji odpowiada liczbie 136 ( $136 \cdot 1$ ), natomiast liczba wierszy w zbiorze danych to 816 wierszy ( $136 \cdot 1 \cdot 3$ ). Fragment zbioru danych przedstawia się następująco:

	osoba	opcja	wybór	cena1	cena2	cena3	doch1	doch2	doch3	doch4	doch5
1	1	1	0	0	1	0	4	0	0	0	0
2	1	2	1	0	1	0	0	4	0	0	0
3	1	3	0	0	0	1	0	0	4	0	0
4	1	4	0	0	0	1	0	0	0	4	0
5	1	5	0	1	0	0	0	0	0	0	4
6	1	6	0	0	0	0	0	0	0	0	0
7	2	1	0	0	1	0	3	0	0	0	0
8	2	2	0	0	1	0	0	3	0	0	0
9	2	3	1	0	0	1	0	0	3	0	0
10	2	4	0	0	0	1	0	0	0	3	0
11	2	5	0	1	0	0	0	0	0	0	3
12	2	6	0	0	0	0	0	0	0	0	0
13	3	1	0	0	1	0	3	0	0	0	0

Jak można zauważyć, w modelu wykorzystano po jednej zmiennej charakteryzującej jednostkę analizy (respondenta) oraz jedną zmienną charakteryzującą zestaw alternatyw (opcji wyboru). Są to odpowiednio zmienne specyficzne (dochód) oraz (cena).

W wyniku zastosowania odpowiedniego skryptu programu R otrzymuje się oszacowania parametrów B:

	B	se	Z
cena1	0.416600	0.8777	0.47470
cena2	0.891000	0.7788	1.14400
cena3	-0.716500	0.9264	-0.77340
doch1	-1.195000	0.3953	-3.02300
doch2	0.019140	0.2537	0.07547
doch3	0.090220	0.3038	0.29700
doch4	0.007195	0.3100	0.02321
doch5	-0.048640	0.2876	-0.16910

Oszacowane parametry mieszanego modelu logitowego należy interpretować w kategoriach prawdopodobieństwa [Gruszczyński 2002]. W odniesieniu do zmiennej (cena) najbardziej prawdopodobny wybór to średnia cena (cena2), najmniej prawdopodobny wybór – wysoka cena, co wynika z ujemnej wartości parametru zmiennej (cena3). W odniesieniu do zmiennej (dochód) najbardziej prawdopodobnym wyborem jest opcja nr 3, natomiast najmniej prawdopodobnym wyborem – opcja nr 1.

## 5. Podsumowanie

W artykule wskazano na różnice między wybranymi modelami kategorii nieuporządkowanych: wielomianowym, warunkowym oraz mieszanym modelem logitowym. Modele te można traktować jako komplementarne w zastosowaniach do pomiaru ukrytych w mikrodanych preferencji. Przedstawione w artykule przykłady zastosowania wielomianowych modeli logitowych potwierdzają powyższe wnioski. Oznacza to możliwość prowadzenia na gruncie badań empirycznych bardziej szczegółowych analiz zachowań konsumentów z uwzględnieniem zarówno jednostki analizy – konsumentów (identyfikowanych przez jego charakterystyki), jak i przedmiotów analizy – produktów lub usług (opisywanych za pomocą ich atrybutów).

## Literatura

- Agresti A. (2002), *Categorical Data Analysis*, Second Edition, Wiley, New York.
- Bąk A. (2012), *Modele kategorii nieuporządkowanych w badaniach preferencji*, [w:] K. Jajuga, M. Walesiak (red.), *Klasyfikacja i analiza danych – teoria i zastosowania*, Taksonomia 17, Prace Naukowe UE we Wrocławiu nr 242, s. 86-95.
- Bierlaire M. (1997), *Discrete Choice Models*, <http://web.mit.edu/mbi/www/michel.html>. Cambridge, Massachusetts Institute of Technology.
- Cameron A.C., Trivedi P.K. (2009), *Microeconometrics. Methods and Applications*, Cambridge University Press, New York.
- Coombs C.H., Dawes R.M., Tversky A. (1977), *Wprowadzenie do psychologii matematycznej*, PWN, Warszawa.
- Greene W.H. (2008), *Econometric Analysis*, 6th ed. Prentice Hall, Upper Saddle River.
- Gruszczyński M. (2002), *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Oficyna Wydawnicza Szkoły Głównej Handlowej, Warszawa.
- Gruszczyński M. (red.) (2010), *Mikroekonometria. Modele i metody analizy danych indywidualnych*, Wolters Kluwer, Warszawa.
- Hoffman S.D., Duncan G.J. (1988), *Multinomial and Conditional Logit Discrete-Choice Models in Demography*, [w:] *Demography*, vol. 25, no. 3, Population Association of America, <http://www.jstor.org/stable/2061541>.
- Jackman S. (2007), *Models for Unordered Outcomes*, Political Science 150C/350C, <http://jackman.stanford.edu/classes/350C/07/unordered.pdf> (12.03.2012).
- Long J.S. (1997), *Regression Models for Categorical and Limited Dependent Variables*, SAGE Publications, Thousand Oaks – London – New Delhi.
- McFadden D. (1974), *Conditional Logit Analysis of Qualitative Choice Behavior*, [w:] P. Zarembka (red.), *Frontiers in Econometrics*, Academic Press, New York-San Francisco-London.
- Powers D.A., Xie Y. (2008), *Statistical Methods for Categorical Data Analysis*, 2nd ed. Emerald, Bingley.
- R Development Core Team (2011), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, <http://cran.r-project.org/>.
- So Y., Kuhfeld W.F. (1995), *Multinomial Logit Models*, <http://support.sas.com/techsup/technote/mr2010g.pdf> (12.03.2012).
- Winkelmann R., Boes S. (2006), *Analysis of Microdata*, Springer, Berlin.



## **MICROECONOMIC POLYNOMIAL MODELS AND THEIR APPLICATION IN THE ANALYSIS OF PREFERENCES USING R PROGRAM**

**Summary:** The main aim of this article is to present some polynomial logit models of discrete choice methods and their application in the measurement of consumer preferences. The basis for distinguishing between types of polynomial models is mainly the nature of explanatory variables in the model. Since this distinction is not adequately interpreted in the literature, this paper presents basic differences between the types of multinomial logit models used in the analysis of consumer preferences. To models estimation were used R packages and functions programmed in R program.

**Keywords:** preferences, multinomial logit models, R program.