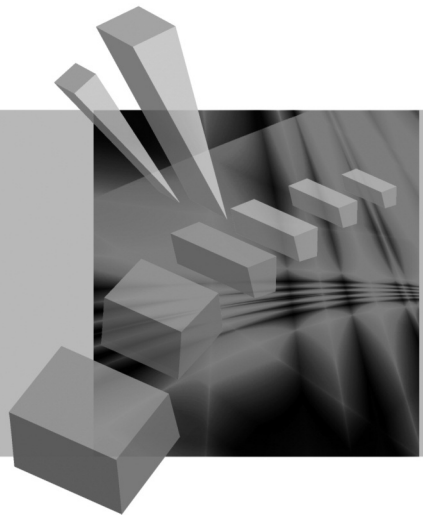
PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu
RESEARCH PAPERS
of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pociecha,
Mirosław Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities http://cejsh.icm.edu.pl
oraz w The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon 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

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 242
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS 2012

Taksonomia 19. Klasyfikacja i analiza danych – teoria i zastosowania ISSN 1899-3192

Andrzej Bąk
Uniwersytet Ekonomiczny we Wrocławiu

MODELE KATEGORII NIEUPORZĄDKOWANYCH W BADANIACH PREFERENCJI

Streszczenie: Wśród mikroekonometrycznych modeli kategorii nieuporządkowanych wyróżnia się najczęściej wielomianowy model logitowy, warunkowy model logitowy i mieszany model logitowy. Podstawę rozróżnienia tych typów modeli stanowi głównie charakter zmiennych objaśniających uwzględnionych w modelu. Rozróżnienie to nie jest jednak jednoznacznie interpretowane. Celem artykułu jest wskazanie podstawowych różnic między typami modeli logitowych oraz prezentacja przykładów estymacji różnych typów tych modeli dla różnych typów danych z wykorzystaniem programu R.

Słowa kluczowe: preferencje, modele kategorii nieuporządkowanych, program R.

1. Wstęp

W badaniach preferencji zmierza się do identyfikacji czynników, którymi kierują się konsumenci, wybierając określone produkty lub usługi. Czynniki te są związane zarówno z cechami konsumentów, jak i z charakterystykami produktów lub usług. Empiryczne badania preferencji opierają się na teorii addytywnego jednoczesnego pomiaru łącznego [Coombs, Dawes, Tversky 1977, s. 50 i nast.] oraz teorii użyteczności losowej [Coombs, Dawes, Tversky 1977, s. 214 i nast.], które powstały na gruncie psychologii matematycznej i psychometrii. Teoria użyteczności losowej stanowi podstawę teoretyczną probabilistycznych modeli wyborów dyskretnych wykorzystywanych w badaniach preferencji, do których należą modele kategorii nieuporządkowanych [Greene 2008, s. 840-847].

Wśród mikroekonometrycznych modeli kategorii nieuporządkowanych wyróżnia się najczęściej wielomianowy model logitowy, warunkowy model logitowy [McFadden 1974] i mieszany model logitowy [Winkelmann, Boes 2006]. Podstawę rozróżnienia tych typów modeli stanowi głównie charakter zmiennych objaśniających. Rozróżnienie to nie jest jednak jednoznacznie interpretowane. Celem artykułu jest wskazanie podstawowych różnic między typami modeli logitowych oraz prezentacja przykładów estymacji różnych typów tych modeli dla różnych typów danych w badaniach preferencji z wykorzystaniem programu R. Przede wszystkim w artykule przedstawiono następujące zagadnienia: charakterystykę wybranych typów modeli

logitowych, interpretację zmiennych objaśniających w różnych typach tych modeli, organizację danych wykorzystywanych w estymacji parametrów różnych typów modeli, funkcje dostępne w programie R, które mogą znaleźć zastosowanie w estymacji modeli logitowych, funkcje napisane w języku programowania R, które zostały wykorzystane w obliczeniach.

2. Modele logitowe kategorii nieuporządkowanych[1]

W badaniach preferencji wykorzystujących metody wyborów dyskretnych szczególnie ważną rolę odgrywają wielomianowe i warunkowe modele logitowe oraz ich połączenie w postaci tzw. mieszanych modeli logitowych [Cameron, Trivedi 2009, s. 500], nazywanych również hybrydowymi modelami logitowymi [Winkelmann, Boes 2006, s. 154]. Modele te mieszczą się w grupie wielomianowych modeli kategorii nieuporządkowanych.

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

$$P_{ki} = \frac{\exp(\mathbf{x}_k^T \boldsymbol{\beta}_i)}{\sum_{l=1}^{n} \exp(\mathbf{x}_k^T \boldsymbol{\beta}_l)}, \text{ przy czym } \boldsymbol{\beta}_n = \mathbf{0}, \tag{1}$$

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

$\mathbf{x}_k^T$ – wektor reprezentujący *k*-ty wiersz macierzy $\mathbf{X}$ (zmiennych objaśniających);

$\boldsymbol{\beta}_i$ – wektor szacowanych parametrów związany z *i*-tą kategorią zmiennej objaśnianej.

Oszacowane wartości prawdopodobieństw w modelu (1) sumują się do 1 w obrębie każdej konfiguracji zmiennych objaśniających. Rozkład prawdopodobieństw spełniających ten warunek można jednak uzyskać przy różnych wartościach parametrów $\boldsymbol{\beta}_i$, a zatem model pozostaje zdefiniowany niejednoznacznie. W celu rozwiązania tego problemu przyjmuje się pewne ograniczenia dotyczące wektora parametrów nazywane normalizacją, zakładając np. $\boldsymbol{\beta}_n = \mathbf{0}$. Oznacza to jednocześnie, że jedna

[1] Punkt opracowany na podstawie pracy [Bąk 2010].

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

Macierz $\mathbf{X}$ w modelu (1) zawiera charakterystyki respondentów, których preferencje dotyczące produktów lub usług są przedmiotem badań. Charakterystyki respondentów są stałe względem tych produktów lub usług.

Warunkowy model logitowy został zaproponowany przez McFaddena [1974]) jako uogólnienie wielomianowego modelu logitowego. Podstawowym kryterium rozróżniania tych modeli jest charakter zmiennych objaśniających, tzn. macierzy $\mathbf{X}$ w równaniu (1). Jeżeli zmienne objaśniające charakteryzują konsumentów, to na ogół wykorzystuje się wielomianowy model logitowy. Jeśli natomiast zmienne objaśniające opisują obiekty będące przedmiotem wyboru (produkty lub usługi), to z reguły stosuje się warunkowy model logitowy.

W warunkowym modelu logitowym prawdopodobieństwo wyboru *i*-tego profilu ze zbioru liczącego *n* elementów jest szacowane na podstawie zależności [So, Kuhfeld 1995, s. 7; Long 1997, s. 178 i nast.; Powers, Xie 2008, s. 256 i nast.; Cameron, Trivedi 2009, s. 500; Gruszczyński 2010, s. 172-173]:

$$P_{ki} = \frac{\exp(\mathbf{z}_{ki}^T \boldsymbol{\alpha})}{\sum_{l=1}^{n} \exp(\mathbf{z}_{kl}^T \boldsymbol{\alpha})}, \tag{2}$$

gdzie: $\mathbf{z}_{ki}^T$ – *k*-ty wektor macierzy $\mathbf{Z}$ (zmiennych objaśniających) opisujący *i*-tą opcję;

$\boldsymbol{\alpha}$ – wektor parametrów (wartość α_j jest związana z *j*-tą zmienną objaśniającą).

Macierz $\mathbf{Z}$ w modelu (2) zawiera charakterystyki produktów lub usług, względem których badane są preferencje respondentów. Wartości zmiennych objaśniających opisujących produkty lub usługi są specyficzne w przekroju opcji wyboru oferowanych respondentom (np. w badaniu ankietowym).

Mieszany model logitowy (3) jest połączeniem modeli (1) i (2), a więc uwzględnia charakterystyki zarówno respondentów, jak i opcji wyboru (produktów lub usług) [So, Kuhfeld 1995; Powers, Xie 2008, s. 258 i nast.; Cameron, Trivedi 2009, s. 500]:

$$P_{ki} = \frac{\exp(\mathbf{x}_k^T \boldsymbol{\beta}_i + \mathbf{z}_{ki}^T \boldsymbol{\alpha})}{\sum_{l=1}^{n} \exp(\mathbf{x}_k^T \boldsymbol{\beta}_l + \mathbf{z}_{kl}^T \boldsymbol{\alpha})}. \tag{3}$$

3. Estymacja parametrów wielomianowych modeli logitowych w programie R

Szacowanie parametrów wielomianowych, warunkowych i mieszanych modeli logitowych w programie R można przeprowadzić z wykorzystaniem funkcji `optim()` z pakietu `stats`. Funkcję `optim` można wykorzystać do maksymalizacji funkcji największej wiarygodności, która umożliwia znalezienie najlepszego dopasowania modelu logitowego danych empirycznych (zob. [Jackman 2007]). Kryterium tego dopasowania jest wartość funkcji wiarygodności. Funkcja `optim()` korzysta z iteracyjnych algorytmów optymalizacji: sympleksu (Neldera-Meada), zmiennej metryki (*quasi*-Newtona, Broydena-Fletchera-Goldfarba-Shannona), gradientów sprzężonych (Fletchera-Reevesa), *quasi*-Newtona z ograniczeniami (algorytm L-BFGS-B), sieci neuronowych (SNN – *Simulated Neural Network*).

Wybrane argumenty funkcji `optim()` są następujące [R Development Core Team 2011]:

`optim(par, fn, gr=NULL, method=c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN"), control=list())`

`par` wartości początkowe szacowanych parametrów,
`fn` funkcja, której wartość jest optymalizowana,
`gr` gradient (wektor pochodnych cząstkowych),
`method` algorytm optymalizacji,
`control` parametry kontrolne.

Funkcja `optim()` domyślnie służy do minimalizacji funkcji największej wiarygodności, ale parametr kontrolny `control=list(fnscale=-1)` umożliwia zmianę kierunku optymalizacji w celu maksymalizacji wartości tej funkcji.

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

```
# fnw(a,x,y) - funkcja największej wiarygodności dla modelu logitowego
# na podstawie: S. Jackman [2007]
# a - wartości startowe parametrów modelu
# x - wartości zmiennych objaśniających
# y - zmienna objaśniana o wartościach TRUE/FALSE (wybór/brak wyboru)
# wywołanie - jako parametr funkcji optim{stats}
fnw<-function(a,x,y) {
   mu<-x%*%a   #składnik systematyczny modelu - użyteczność opcji
   eta<-exp(mu)                    #licznik
   suma<-tapply(eta,s,sum)         #mianownik
   pr<-eta[y]/suma                 #prawdopodobieństwa wyboru opcji
   lnw<-sum(log(pr))               #suma logarytmów prawdopodobieństw
   return(lnw)                     #logarytm fnw }.
```

4. Wykorzystanie funkcji `optim()` w estymacji modeli logitowych

W przykładach szacowania modeli logitowych (wielomianowego, warunkowego i mieszanego) wykorzystano dane (zbiór o nazwie `travel`) z pracy [So, Kuhfeld 1995] opisujące wybór środka podróży w zależności od czasu podróży (`travtime`) – jest to zmienna specyficzna dla opcji wyboru o poziomach `autotime`, `plantime` i `trantime`. Do wyboru jest jedna z trzech opcji: samochód (`auto`), samolot (`plane`) lub przewóz publiczny autobusem lub pociągiem (`transit`). Zawartość zbioru `travel` zapisana w pliku `travel.csv` jest następująca:

```
   autotime plantime trantime age chosen
1      10.0      4.5     10.5  32      2
2       5.5      4.0      7.5  13      1
3       4.5      6.0      5.5  41      3
4       3.5      2.0      5.0  41      3
5       1.5      4.5      4.0  47      1
6      10.5      3.0     10.5  24      2
7       7.0      3.0      9.0  27      1
8       9.0      3.5      9.0  21      2
9       4.0      5.0      5.5  23      1
10     22.0      4.5     22.5  30      2
11      7.5      5.5     10.0  58      2
12     11.5      3.5     11.5  36      3
13      3.5      4.5      4.5  43      1
14     12.0      3.0     11.0  33      2
15     18.0      5.5     20.0  30      2
16     23.0      5.5     21.5  28      2
17      4.0      3.0      4.5  44      2
18      5.0      2.5      7.0  37      3
19      3.5      2.0      7.0  45      1
20     12.5      3.5     15.5  35      2
21      1.5      4.0      2.0  22      1.
```

W strukturze tych danych każdy wiersz przedstawia zbiór (sytuację wyboru), z którego respondent wybrał jedną z trzech opcji (profilów). Zmienną specyficzną dla respondentów jest wiek (`age`), liczba respondentów wynosi 21, natomiast wybrany środek podróży reprezentuje zmienna `chosen` (o wartościach `1 - auto`, `2 - plane`, `3 - transit`).

W **warunkowym modelu logitowym** zmienne objaśniające charakteryzują przedmiot wyboru (profile tworzące zbiór, z którego respondent dokonuje wyboru), a więc są specyficzne dla opcji wyboru. Oszacowanie takiego modelu wymaga prze-

kształcenia danych z pliku `travel.csv` do postaci zapisanej w pliku `travel_2.csv` (10 pierwszych wierszy)[2]:

```
   subject option choice travtime age
    [1,]       1      1       0    10.0  32
    [2,]       1      2       1     4.5  32
    [3,]       1      3       0    10.5  32
    [4,]       2      1       1     5.5  13
    [5,]       2      2       0     4.0  13
    [6,]       2      3       0     7.5  13
    [7,]       3      1       0     4.5  41
    [8,]       3      2       0     6.0  41
    [9,]       3      3       1     5.5  41
   [10,]       4      1       0     3.5  41.
```

W strukturze tych danych każdy wiersz przedstawia jedną opcję wyboru (jeden profil), zmienna `subject` reprezentuje numer respondenta, zmienna `option` numer opcji wyboru (`1 - auto`, `2 - plane`, `3 - transit`) , zmienna objaśniana `choice` wskazuje wybraną opcję (`1` – wybrana opcja, `0` – niewybrana opcja), zmienna `travtime` jest specyficzna dla opcji wyboru i reprezentuje czas podróży, a zmienna `age` jest specyficzna dla respondentów i reprezentuje ich wiek (wartość tej zmiennej powtarza się trzykrotnie, ponieważ są trzy opcje wyboru). Liczba wierszy w tym zbiorze danych wynosi 63 (21 respondentów, 3 opcje wyboru).

Skrypt 1 wykorzystuje funkcję największej wiarygodności `fnw()` do estymacji warunkowego modelu logitowego za pomocą funkcji `optim()`.

Skrypt 1.

```
source("fnw.r")
dane<-read.csv2("travel_2.csv", header=TRUE)
head(dane,6)
attach(dane)
s<-subject   #identyfikator respondenta
p<-option    #identyfikator opcji
w<-choice==1 #identyfikator wyboru TRUE/FALSE
y<-p[w]      #numery wybranych opcji
X<-as.matrix(cbind(travtime))   #macierz danych
k<-dim(X)[2] #liczba zmiennych objaśniających
a<-rep(0,k)  #wartości startowe parametrów modelu
nazwy<-colnames(X)
clm<-optim(par=a,fn=fnw,x=X,y=w,control=list(trace=TRUE,fnscale=-1),
method="BFGS",hessian=TRUE)
se<-sqrt(diag(solve(-clm$hessian)))   #standardowe błędy parametrów
B<-clm$par                            #parametry
```

[2] Ze względu na ograniczoną objętość artykułu nie zamieszczono procedur i funkcji wykonujących przekształcenia danych.

```
wyniki<-cbind(B,se)
wyniki<-cbind(wyniki,wyniki[,1]/wyniki[,2],exp(wyniki[,1]))
dimnames(wyniki)<-list(nazwy,c("B","se","Z","exp(B)"))
print(signif(wyniki,4)).
```

W wyniku wykonania skryptu 1 otrzymuje się oszacowanie parametru (B) dla zmiennej specyficznej dla opcji wyboru (travtime):

```
                B     se      Z exp(B)
travtime -0.2655 0.1021 -2.599 0.7668.
```

W **wielomianowym modelu logitowym** zmienne objaśniające charakteryzują respondentów, a więc są specyficzne dla podmiotu dokonującego wyboru. Oszacowanie takiego modelu za pomocą funkcji optim() wymaga przekształcenia danych z pliku travel_2.csv do postaci zapisanej w pliku travel_3.csv (10 pierwszych wierszy):

```
      subject option choice travtime auto plane ageauto ageplane
 [1,]       1      1      0     10.0    1     0      32         0
 [2,]       1      2      1      4.5    0     1       0        32
 [3,]       1      3      0     10.5    0     0       0         0
 [4,]       2      1      1      5.5    1     0      13         0
 [5,]       2      2      0      4.0    0     1       0        13
 [6,]       2      3      0      7.5    0     0       0         0
 [7,]       3      1      0      4.5    1     0      41         0
 [8,]       3      2      0      6.0    0     1       0        41
 [9,]       3      3      1      5.5    0     0       0         0
[10,]       4      1      0      3.5    1     0      41         0.
```

W strukturze tych danych każdy wiersz przedstawia jedną opcję wyboru. Trzy wiersze tworzą zbiór, z którego respondent wybrał jedną opcję (profil). Opcje wyboru są w tym zbiorze reprezentowane przez zmienne zero-jedynkowe (auto, plane), które zostały pomnożone przez zmienną specyficzną dla respondentów (age). W wyniku takiej interakcji powstały zmienne ageauto i ageplane, które reprezentują zmienną specyficzną dla respondentów age i są jednocześnie specyficzne dla opcji wyboru. Profilem odniesienia jest opcja trzecia (transit), pominięta w tym zbiorze.

Skrypt 2 wykorzystuje funkcję największej wiarygodności fnw() do estymacji wielomianowego modelu logitowego za pomocą funkcji optim().

Skrypt 2.

```
source("fnw.r")
dane<-read.csv2("travel_3.csv", header=TRUE)
head(dane,6)
attach(dane)
s<-subject     #identyfikator respondenta
p<-option      #identyfikator opcji
```

```
w<-choice==1   #identyfikator wyboru TRUE/FALSE
y<-p[w]        #numery wybranych opcji
X<-as.matrix(cbind(auto,plane,ageauto,ageplane))   #macierz danych
k<-dim(X)[2]   #liczba zmiennych objaśniających
a<-rep(0,k)    #wartości startowe parametrów modelu
nazwy<-colnames(X)
clm<-optim(par=a,fn=fnw,x=X,y=w,control=list(trace=TRUE,fnscale=-1),
method="BFGS",hessian=TRUE)
se<-sqrt(diag(solve(-clm$hessian)))   #standardowe błędy parametrów
B<-clm$par                            #parametry
wyniki<-cbind(B,se)
wyniki<-cbind(wyniki,wyniki[,1]/wyniki[,2],exp(wyniki[,1]))
dimnames(wyniki)<-list(nazwy,c("B","se","Z","exp(B)")).
print(signif(wyniki,4)).
```

W wyniku wykonania skryptu 2 otrzymuje się oszacowania parametrów (`B`) dla zmiennej specyficznej dla respondentów (`age`):

```
                B      se      Z  exp(B)
auto      3.04900 2.42700  1.256 21.0800
plane     2.72400 2.29400  1.187 15.2400
ageauto  -0.07106 0.06518 -1.090  0.9314
ageplane -0.05007 0.05961 -0.840  0.9512.
```

W tym modelu są dwa wyrazy wolne (`auto` i `plane`) i dwa parametry dla zmiennej `age` (`ageauto` i `ageplane`), które reprezentują odpowiednio efekt wpływu na prawdopodobieństwo wyboru samochodu w odniesieniu do przewozu publicznego i samolotu w odniesieniu do przewozu publicznego.

W **mieszanym (hybrydowym) modelu logitowym** występują zmienne objaśniające charakteryzujące zarówno respondentów, jak i opcje wyboru. Struktura zbioru danych jest taka sama jak w przypadku wielomianowego modelu logitowego (`travel_3.csv`).

Skrypt 3 wykorzystuje funkcję największej wiarygodności `fnw()` do estymacji mieszanego modelu logitowego za pomocą funkcji `optim()`.

Skrypt 3.

```
source("fnw.r")
dane<-read.csv2("travel_3.csv", header=TRUE)
head(dane,6)
attach(dane)
s<-subject   #identyfikator respondenta
p<-option    #identyfikator opcji
w<-choice==1 #identyfikator wyboru TRUE/FALSE
y<-p[w]      #numery wybranych opcji
X<-as.matrix(cbind(auto,plane,ageauto,ageplane,travtime))   #macierz
danych
k<-dim(X)[2] #liczba zmiennych objaśniających
a<-rep(0,k)  #wartości startowe parametrów modelu
nazwy<-colnames(X)
```

```
clm<-optim(par=a,fn=fnw,x=X,y=w,control=list(trace=TRUE,fnscale=-1),
method="BFGS",hessian=TRUE)
se<-sqrt(diag(solve(-clm$hessian)))  #standardowe błędy parametrów
B<-clm$par                           #parametry
wyniki<-cbind(B,se)                  #parametry i błędy
wyniki<-cbind(wyniki,wyniki[,1]/wyniki[,2],exp(wyniki[,1]))
dimnames(wyniki)<-list(nazwy,c("B","se","Z","exp(B)")).
print(signif(wyniki,4)).
```

W wyniku wykonania skryptu 3 otrzymuje się oszacowania parametrów (`B`) dla zmiennych specyficznych dla respondentów i opcji wyboru:

```
                B      se       Z   exp(B)
auto      2.50100 2.39600  1.0440 12.20000
plane    -2.78000 3.53000 -0.7873  0.06206
ageauto  -0.07828 0.06333 -1.2360  0.92470
ageplane  0.01696 0.07442  0.2279  1.01700
travtime -0.60850 0.27130 -2.2430  0.54420.
```

Oszacowane parametry mieszanego modelu można interpretować w kategoriach prawdopodobieństwa [Gruszczyński 2002]. Jeżeli czas podróży wzrasta (ujemna wartość parametru przy zmiennej `travtime`), to zmniejsza się prawdopodobieństwo wyboru takiej opcji. Wraz z rosnącym wiekiem respondentów będzie malało prawdopodobieństwo wyboru samochodu (ujemna wartość parametru przy zmiennej `ageauto`) w odniesieniu do prawdopodobieństwa wyboru przewozu publicznego, natomiast będzie wzrastać prawdopodobieństwo wyboru samolotu (dodatnia wartość parametru przy zmiennej `ageplane`) w odniesieniu do prawdopodobieństwa wyboru przewozu publicznego.

5. Podsumowanie

Metoda największej wiarygodności umożliwia estymację różnych typów modeli kategorii nieuporządkowanych. Oferowana w programie R funkcja `optim()` z pakietu `stats` wymaga opracowania funkcji największej wiarygodności dla modeli logitowych, którą można wykorzystać w takiej samej postaci do estymacji warunkowego, wielomianowego i mieszanego modelu logitowego.

W procedurach estymacji modeli kategorii nieuporządkowanych ważną rolę odgrywa struktura danych empirycznych. Szacowanie różnych typów tych modeli wymaga przekształcania danych do odpowiednich formatów. W tym celu warto kontynuować prace zmierzające do opracowania uniwersalnych procedur transformacji danych do pożądanej postaci w zależności od szacowanego modelu.

Literatura

Agresti A., *Categorical Data Analysis*, Second Edition, Wiley, New York 2002.

Bąk A., *Analiza danych o preferencjach z wykorzystaniem mikroekonometrycznych modeli kategorii nieuporządkowanych i programu R*, [w:] K. Jajuga, M. Walesiak, *Klasyfikacja i analiza danych – teoria i zastosowania*, Taksonomia 17, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 107, UE, Wrocław 2010.

Cameron A.C., Trivedi P.K., *Microeconometrics. Methods and Applications*, Cambridge University Press, New York 2009.

Coombs C.H., Dawes R.M., Tversky A., *Wprowadzenie do psychologii matematycznej*, PWN, Warszawa 1977.

Greene W.H., *Econometric Analysis*, 6th ed., Prentice Hall, Upper Saddle River, 2008.

Gruszczyński M. (red.), *Mikroekonometria. Modele i metody analizy danych indywidualnych*, Wolters Kluwer, Warszawa 2010.

Gruszczyński M., *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Oficyna Wydawnicza Szkoły Głównej Handlowej, Warszawa 2002.

Jackman S., *Models for Unordered Outcomes*, Political Science 150C/350C, 2007, http://jackman.stanford.edu/classes/350C/07/unordered.pdf (14.10.2011).

Long J.S., *Regression Models for Categorical and Limited Dependent Variables*, SAGE Publications, Thousand Oaks-London-New Delhi 1997.

McFadden D., *Conditional Logit Analysis of Qualitative Choice Behavior*, [w:] P. Zarembka (red.), *Frontiers in Econometrics*, Academic Press, New York-San Francisco-London 1974.

Powers D.A., Xie Y., *Statistical Methods for Categorical Data Analysis*, 2nd ed., Emerald, Bingley 2008.

R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, 2011, http://cran.r-project.org/.

So Y., Kuhfeld W.F., *Multinomial Logit Models*, 1995, http://www.sascommunity.org/ sugi/SUGI95/, 14.10.2011.

Winkelmann R., Boes S., *Analysis of Microdata*, Springer-Verlag, Berlin, Heidelberg 2006.

MODELS FOR UNORDERED CATEGORIES IN PREFERENCE ANALYSIS

Summary: Among microeconometric models for unordered categories multinomial logit model, conditional logit model and mixed logit model are most frequently mentioned. The character of the independent variables included in the model is mainly the basis for distinguishing among these types of models. This distinction is not clearly interpreted. The main aim of this article is to identify the fundamental differences among the types of logit models and to present the examples of estimation of various types of this models for various types of data using R program.

Keywords: preferences, models for unordered categories, R program.