

**Michał Purczyński**

Kompania Piwowarska w Poznaniu

Uniwersytet Ekonomiczny w Poznaniu

## ZASTOSOWANIE MIESZANEGO MODELU LOGITOWEGO DO BADANIA PREFERENCJI KONSUMENTÓW PIWA

### Streszczenie

Celem artykułu jest pokazanie przykładu zastosowania mieszanego modelu logitowego (inaczej nazywanego modelem z parametrami losowymi) do badania preferencji konsumentów. Model ten jest bardzo elastyczny i pozwala na przybliżenie dowolnego modelu użyteczności losowej. Jego praktyczne wykorzystanie pokazano na przykładzie analizy konsumentów rynku piwowarskiego w Polsce. Do estymacji parametrów modelu wykorzystano programu SAS. Pokazano, w jakim stopniu wiek, płeć, sytuacja materialna respondenta wpływają na postrzeganą przez niego użyteczność piwa i na wybory dokonywane przez konsumentów.

**Słowa kluczowe:** rynek piwowarski, mieszany model logitowy, badanie preferencji konsumentów, SAS.

### Wprowadzenie

W analizie preferencji konsumentów zmienna objaśniana opisuje wybory dokonane przez respondentów, będących zwykle konsumentami danego produktu. Zmienne objaśniające można podzielić na dwie podgrupy: zmienne charakterystyczne (specyficzne) dla wybieranej opcji (w tym przypadku produktu), na przykład cena, rodzaj opakowania, marka oraz zmienne specyficzne dla jed-

nostki wybierającej (np. wiek, wykształcenie, płeć). Do estymacji tego typu danych wykorzystywane są różnego rodzaju modele logitowe, będące już od wielu lat podstawowymi ekonometrycznymi „narzędziami” służącymi od analizy danych dyskretnych. Począwszy od najprostszego, dwumianowego modelu logitowego, w którym są tylko dwie możliwości wyboru, po modele wielomianowe, np. wielomianowy model logitowy (*Multinomial Logit* – MNL), warunkowy model logitowy (*Conditional Logit* – CL), pozwalające na wybór spośród wielu dostępnych opcji [1, s. 233–235; 2, s. 114–125; 3, s. 51; 9, s. 160–189].

Niestety, pomimo wielu takich zalet jak prostota estymacji i interpretacji, dostępność oprogramowania komputerowego, modele te nie są wolne od mankamentów. Przede wszystkim mocno krytykowane jest przyjmowane założenie o nieskorelowaniu składników losowych dla różnych opcji wyboru [5, s. 135; 9, s. 180], które w praktyce gospodarczej występuje bardzo często [9, s. 181]<sup>1</sup>. W takiej sytuacji wykorzystanie najpopularniejszych modeli logitowych jest nieuzasadnione i prowadzi do błędnej specyfikacji modelu. Otrzymane za ich pomocą estymatory tracą swoje pożądane własności – zgodność, asymptotyczną efektywność i asymptotyczną normalność [9, s. 182]. Jednym z modeli, który pozwala na szacowanie danych ze skorelowanymi składnikami losowymi, jest logitowy model mieszany (*mixed logit*), często nazywany modelem logitowym o zmiennych parametrach (*random parameters logit*). Kolejną zaletą mieszanego modelu logitowego jest fakt, że pozwala on na wykorzystanie jako zmiennych objaśniających cech charakteryzujących respondentów i opcje wyboru. Modele logitowe warunkowy i wielomianowy są natomiast używane wówczas, gdy zmienne objaśniające opisują odpowiednio respondentów lub opcje wyboru [1, s. 234].

Celem artykułu jest wykorzystanie mieszanego modelu logitowego do analizy preferencji konsumentów piwa. Plan artykułu jest następujący: w punkcie pierwszym scharakteryzowano typy modeli dekompozycyjnych, ze szczególnym uwzględnieniem zjawiska heterogeniczności konsumentów piwa. Punkt

---

<sup>1</sup> Modele wielomianowy i warunkowy przyjmują jedno z dość powszechnie krytykowanych założeń – o niezależności od nieistotnych możliwości (*independence of irrelevant alternatives* – IIA). Według tej zasady, iloraz prawdopodobieństw wyboru profilów (dostępnych opcji) powinien być stały niezależnie od tego, ile innych profilów zostanie dodane do zbioru [1, s. 120]. Oznacza to, że gdy jeden z profilów zostanie usunięty ze zbioru dostępnych możliwości, to prawdopodobieństwo zostanie podzielone między pozostałe profile proporcjonalnie do ich prawdopodobieństw wyboru [9, s. 181]. Niepożądane skutki tego założenia można zilustrować, znanym w literaturze przedmiotu, przykładem czerwonego i niebieskiego autobusu [7, s. 198–272].

drugi w całości poświęcono mieszanemu modelowi logitowemu, po czym w punkcie trzecim omówiono metody jego estymacji w programie SAS. Praktyczne zastosowanie mieszanych modeli logitowych do badania preferencji konsumentów piwa pokazano w czwartym punkcie.

## 1. Dekompozycyjne metody pomiaru preferencji konsumentów

Celem dekompozycyjnych metod pomiaru preferencji jest podzielenie użyteczności całkowitej produktu na użyteczności cząstkowe, odnoszące się do poszczególnych atrybutów, takich jak cena, opakowanie, marka itd. Wyselekcjonowane w badaniu atrybuty są zmiennymi niezależnymi, a zmienną zależną jest użyteczność całkowita produktu [13, s. 147]. W publikacjach dotyczących *conjoint analysis* używa się następującej terminologii: atrybutami lub czynnikami nazywa się zmienne objaśniające, poziomy są realizacjami poszczególnych zmiennych objaśniających, a profile to różne warianty atrybutów i ich poziomów.

W modelu dekompozycyjnym możliwe jest przedstawienie zależności odwzorowującej użyteczność całkowitą danego profilu w zależności od jego atrybutów za pomocą funkcji [15, s. 284]:

$$W_{ij} = f_i(X, \beta, \varepsilon_{ij}) \quad (1)$$

gdzie:

- $W_{ij}$  – użyteczność  $j$ -ego profilu dla  $i$ -tego respondenta,
- $f_i$  – funkcja preferencji  $i$ -tego respondenta,
- $X$  – macierz obserwacji zawierająca realizacje zmiennych objaśniających opisujących profile (poziomy atrybutów lub realizacje zmiennych sztucznych),
- $\beta$  – macierz parametrów (użyteczności cząstkowych),
- $\varepsilon_{ij}$  – składnik losowy modelu.

Estymacja parametrów modeli dekompozycyjnych może przebiegać na różnym poziomie agregacji danych. Wyróżnia się następujące poziomy: indywidualny, segmentowy i zagregowany [2, s. 129].

Najbardziej szczegółowe informacje otrzymuje się przez szacowanie użyteczności cząstkowych na **poziomie indywidualnym**. W takim przypadku do każdego respondenta przypisana jest jego własna funkcja użyteczności [2, s. 129]. Metoda ta jest intuicyjnie zrozumiała – każdy respondent dokonuje wyborów według, własnych, charakterystycznych tylko dla niego preferencji. Konsekwencją tak mocnego „rozdrobnienia” funkcji użyteczności jest bardzo dobre dopasowanie modelu do danych empirycznych. Szacowanie funkcji użyteczności dla na przykład 1000 konsumentów, w którym każdy konsument oceniał na przykład 300 profili, prowadzi do szalenie skomplikowanych obliczeniowo modeli, do których należy wykorzystać odpowiednie oprogramowanie i sprzęt komputerowy<sup>2</sup>.

Przeciwieństwem modeli indywidualnych są **modele zagregowane (homogeniczne)**. Oceny respondentów są tutaj agregowane na przekroju całej badanej próby, czyli do ogółu danych dopasowywana jest jedna funkcja użyteczności. Z oczywistych względów dopasowanie tego typu modelu do danych empirycznych jest zdecydowanie gorsze i nie pozwala na tak precyzyjne analizy, jak w przypadku modeli indywidualnych. W modelach zagregowanych nie jest możliwe szacowanie interakcji krzyżowych między atrybutami, na przykład między ceną i marką [12, s. 3]. Także dokładność prognostyczna modeli zagregowanych, zbadana na podstawie symulacji, nie jest tak dobra jak modeli indywidualnych [12, s. 4–5].

**Modele segmentowe** są rozwiązaniem pośrednim między modelami indywidualnymi i zagregowanymi. W tego typu modelach, badaną populację dzieli się na jednorodne, według określonej reguły, podgrupy, dla których szacuje się funkcje użyteczności. Metod postępowania z tego typu modelami jest wiele:

- a) procedury dwufazowe, gdzie w pierwszym kroku dzieli się badaną zbiorowość na podgrupy za pomocą na przykład formalnych metod klasyfikacji lub subiektywnych ocen *a priori*, a następnie szacuje się dla nich parametry funkcji użyteczności;

---

<sup>2</sup> Do jednej z najbardziej znanych na świecie firm zajmujących się badaniem preferencji konsumentów i analizą danych statystycznych należy Sawtooth Software. Przedsiębiorstwo to, mające swą siedzibę w Sequim w stanie Washington, USA, rozwija od strony naukowej i dostarcza oprogramowanie komputerowe przede wszystkim do analizy *conjoint*. Według American Marketing Association, Sawtooth Software było w 2005 r. na czwartym miejscu wśród firm produkujących oprogramowanie do analizy danych rynkowych (zaraz po SPSS, Microsoft Excel, and SAS System).

- b) procedury jednofazowe, w których za pomocą jednej funkcji szacuje się parametry funkcji użyteczności i grupuje zbiorowość w podgrupy, na przykład modele klas ukrytych, modele z parametrami losowymi.

Z modelami tymi wiąże się pojęcie heterogeniczność. Za pomocą modeli indywidualnych można zaobserwować wpływ czynników takich, jak płeć, wykształcenie, dochód, miejsce zamieszkania itp. na podejmowane przez respondenta decyzje. Ponadto za pomocą modeli indywidualnych można oszacować interakcje między tymi czynnikami, na przykład, jakie piwa preferują bogate kobiety z miasta i jaki czynnik charakterystyczny dla piwa, na przykład cena, opakowania czy marka, wpływa na ich decyzję i w jakim stopniu. W modelach zagregowanych ta niejednorodność badanej podgrupy jest „włączona” do składnika losowego i powiększa niedopasowanie modelu. Ignorowanie wyraźnych różnic indywidualnych może prowadzić do błędnego oszacowania parametrów i „przypisania” części zmienności niewłaściwej zmiennej objaśnianej [9, s. 26]. Model segmentowy jest w tym momencie rozwiązaniem kompromisowym – unika się szacowania ogromnej liczby parametrów, jednak nie rezygnuje całkowicie z uwzględniania zmienności jednostek.

## 2. Logitowy model mieszany

Jednym z modeli, który pozwala na badanie preferencji konsumentów i ich segmentację, jest logitowy model mieszany, inaczej nazywany modelem z parametrami losowymi. Jest on bardzo elastyczny i pozwala na przybliżenie dowolnego modelu użyteczności losowej [8, s. 1]. Ponadto eliminuje następujące ograniczenia standardowego modelu logitowego:

- a) pozwala, aby wybory były opisane za pomocą zmiennej losowej o określonym rozkładzie;
- b) umożliwia występowanie w badaniu substytucyjnych względem siebie profiliów<sup>3</sup>.

W modelu tym przyjmuje się założenie, że wektor parametrów ma określony rozkład prawdopodobieństwa, zwykle jest to rozkład normalny, aczkolwiek dopuszczalne są też i inne rozkłady. Kolejną przewagą mieszanego modelu logitowego nad modelami wielomianowym i warunkowym jest fakt, że pozwala

---

<sup>3</sup> Zob. przypis 1.

on na uwzględnienia informacji o preferencjach pochodzących spoza próby, dzięki czemu możliwe jest modelowanie heterogeniczności konsumentów [1, s. 142].

Niech  $q=1, \dots, Q$  oznacza respondenta, który dokonuje wyboru  $t=1, \dots, T$  spośród dostępnych opcji  $i=1, \dots, I$ . Założono, że respondent wybierze tę opcję, która z jego punktu widzenia ma największą użyteczność. Wartość użyteczności dla respondenta  $q$  przedstawia równanie (2):

$$U_{itq} = \beta'_q x_{itq} + \varepsilon_{itq} \quad (2)$$

gdzie:

- $i$  – numer opcji,
- $q$  – numer respondenta,
- $t$  – numer sytuacji wyboru,
- $x_{itq}$  – macierz zmiennych objaśniających, zawierająca zarówno cechy, opisujące opcje wyboru, na przykład dla piwa są to cena, opakowanie, marka, jak i cechy charakteryzujące respondenta, na przykład dochód, wykształcenie, miejsce zamieszkania [5, s. 135],
- $\varepsilon_{itq}$  – składnik losowy modelu.

W równaniu (2) założono, że składnik losowy jest sferyczny. Założenie to w przypadku badań preferencji konsumentów jest bardzo restrykcyjne i zwykle nie jest spełnione. Konsumenty tworzą mocno zróżnicowaną grupę, a ich preferencje zależą od wieku i zmieniają się wraz z nim, a także z dochodem i wykształceniem, co powoduje, że zjawisko autokorelacji i heteroskedastyczności składnika losowego jest normalne, aczkolwiek statystycznie niepożądane. Sposobem pozwalającym na oszacowanie modelu i polepszenie jego dopasowania przez uwzględnienie wszystkich dostępnych informacji jest podzielenie składnika losowego na dwie części<sup>4</sup>:  $\eta_{iq}$  i  $\varepsilon_{itq}$ . Składnik losowy  $\eta_{iq}$  ma rozkład prawdopodobieństwa, zależny od zmiennych egzogenicznych, takich jak wiek, dochody, wykształcenie [1, s. 235; 14, s. 668], natomiast  $\varepsilon_{itq}$  jest nieskorelowany i niezależny od charakterystyk respondentów i opcji wyboru [5, s. 135]. W tej sytuacji równanie (2) jest zapisane następująco (dla uproszczenia zrezygnowano z subskrypty  $t$ ):

<sup>4</sup> Dla uproszczenia notacji pominięto indeks  $t$  oznaczający numer sytuacji wyboru.

$$U_{iq} = \beta'_q x_{iq} + [\eta_{iq} + \varepsilon_{iq}] \quad (3)$$

### 3. Estymacja parametrów mieszanego modelu logitowego za pomocą programu SAS

Do estymacji mieszanego modelu logitowego wykorzystano program SAS. Procedury potrzebne do modelowania wyborów dyskretnych znajdują się w module STAT. Podstawową funkcją służącą do estymacji modeli wielomianowych jest funkcja CATMOD, jednak za jej pomocą nie można oszacować modeli warunkowego i mieszanego modelu logitowego. Do tego celu można użyć procedury PHREG stosowanej do estymacji modelu proporcjonalnego hazardu Coxa wykorzystywanego w analizie przeżycia<sup>5</sup> [14, s. 668–669]. Estymacja tego typu modelu odbywa się za pomocą metody częściowej wiarygodności *MCW*. W *MCW* funkcja wiarygodności jest iloczynem udziałów każdego z zaobserwowanych zdarzeń [4, s. 115].

Niech  $q = 1, \dots, Q$  oznacza respondenta, który dokonuje wyboru spośród  $i = 1, \dots, I$  dostępnych opcji. Zmienna *wybor*, służąca do identyfikacji, czy dana opcja została przyjęta lub odrzucona, ma wartość 1 w przypadku wybrania opcji przez respondenta i 2 w sytuacji odrzucenia. Dla jednego wyboru będzie  $I - 1$  wartości 2 i jedna wartość 1. Dalej przyjmijmy, że  $z_{qi}$  oznacza  $q$ -ty wektor macierzy  $Z$ , zawierającej zmienne charakterystyczne dla produktu (cena, marka),  $x_q$  – wektor reprezentujący  $q$ -ty wiersz macierzy  $X$ , zawierającej charakterystyki respondentów. Dla tak sformułowanego modelu funkcja częściowej wiarygodności ma postać [1, s. 235]:

$$L = \frac{\exp(\beta'_k x_q + \theta' z_{qk})}{\left[ \sum_{i=1}^I \exp(\beta'_i x_q + \theta' z_{qi}) \right]} \quad (4)$$

<sup>5</sup> Analiza przeżycia to zbiór metod statystycznych analizujących, jaki czas upłynie od początku badania do zajścia pewnego, interesującego badacza, zdarzenia. Metody te początkowo były rozwijane w naukach medycznych i biologicznych, gdzie np. badano długość życia osób, obecnie są coraz częściej wykorzystywane w naukach społecznych i ekonomicznych, np. do analizy skłonności odejścia pracownika z firmy, zaprzestania płacenia rachunków przez klienta, a także w inżynierii np. czas do awarii urządzenia [Wikipedia: [http://pl.wikipedia.org/wiki/Analiza\\_prze%C5%BCycia](http://pl.wikipedia.org/wiki/Analiza_prze%C5%BCycia)].

gdzie  $L$  – prawdopodobieństwo, z jakim jednostka  $q$  wybierze alternatywę  $k$ .

Przykładowy kod programu wykorzystujący procedurę PHREG do analizy preferencji konsumentów:

**program 1.**

```
proc phreg data = dane_ankietowe;
model wybor*wybor(2) = Z1 Z2 / ties=breslow;
strata pytanie;
run;
```

Pierwsza linia kodu wczytuje dane z odpowiednio przygotowanej bazy danych, w drugiej jest zdefiniowany model, gdzie  $Z1$ ,  $Z2$  oznaczają zmienne objaśniające, a w trzeciej linii kodu jest zmienna grupująca – *pytanie*. W przypadku badania wyborów konsumentów zmienną grupującą był numer pytania. Za pomocą tak sformułowanego zadania można estymować następujące modele: wielomianowy model logitowy, warunkowy model logitowy i mieszany model logitowy. Rozróżnienia między modelami dokonuje się, analizując rodzaj zmiennych objaśniających. Sposób wywołania procedury PHREG jest taki sam dla tych modeli.

#### 4. Zastosowanie mieszanego modelu logitowego do analizy preferencji konsumentów piwa

Praktyczne wykorzystanie mieszanego modelu logitowego zilustrowano przykładem z polskiego rynku piwowarskiego, wykorzystującym badanie typu *choice based conjoint* (CBC). Badanie to zostało zrealizowane w 2009 roku przez firmę badawczą MilwardBrown SMG/KRC na zlecenie Kompanii Piwowarskiej<sup>6</sup>. Celem analizy jest obliczenie, jak cena, opakowanie i marka wpływają na użyteczność piwa wśród badanych respondentów, oraz zbadanie, czy zależności te zmieniają się wraz z cechami charakteryzującymi respondenta, na

---

<sup>6</sup> Badanie przeprowadzono na reprezentatywnej próbie 1200 osób w wieku 18–60 lat, kupujących piwo co najmniej 2, 3 razy w miesiącu; miejscem badania był dom respondenta. Rozkład próby dla wieku, płci, wielkości miejsca zamieszkania był oparty na danych z niezależnego badania TGI – 36 tys. wywiadów rocznie. Ze względu na fakt, że Kompania Piwowarska – zleceniodawca badania – nie wyraziła zgody na opublikowanie pełnych danych, piwa oznaczono nazwami Marka 1, Marka 2 itp.



przykład z jego wiekiem, dochodem, płcią, wykształceniem, miejscem zamieszkania.

Badanie składało się z dwóch zasadniczych części: pierwsza część miała na celu określenie metryki respondenta oraz preferowanych przez niego sześciu marek piwa, część druga – zbadanie wrażliwości cenowej. Jeżeli, ankietowany nie znał lub nie potrafił sobie przypomnieć aż tylu marek, to prowadzący wspomagał go planszą z listą marek piw. Badanie było przeprowadzone za pomocą komputera, który symulował sklepową półkę z ustawionymi na niej piwami. Dzięki określeniu preferowanych marek piwa we właściwym badaniu zrezygnowano z prezentowania wszystkich dostępnych na rynku produktów (w przypadku tego badania rozważano 26 największych marek piwa według AC Nielsen).

Finalnie na ekranie komputera pojawiała się 6 marek piwa, każda z nich w dwóch opakowaniach: butelka i puszka 500 ml<sup>7</sup>, pod którymi znajdowały się przykładowe półkowe ceny. Zbadano 5 poziomów cenowych: -4%, 0%, +2%, +5%, +8%; za punkt odniesienia przyjęto przeciętną cenę rynkową produktu wyliczoną na podstawie danych AC Nielsen. Jeżeli cena żadnej z prezentowanych marek nie zadowalała respondenta, mógł on odmówić wyboru i przejść do następnej symulacji. Wielu respondentów korzystało z tej opcji wówczas, gdy ceny preferowanych przez nich piw wzrosły zbyt mocno, a ceny piw mniej przez nich lubianych nie były dostatecznie atrakcyjne.

Estymację parametrów mieszanego modelu logitowego rozpoczęto od przekształcenia struktury danych do postaci umożliwiającej wykorzystanie procedury PHREG. Przykładowy fragment danych przedstawiono w tabeli 1.

Analizę preferencji konsumentów piwa rozpoczęto od zbadania, jak postrzegana użyteczność piwa zmienia się w zależności od płci respondenta. Model został wywołany następującym programem:

```
program 2.
proc phreg data = dane_ankietowe;
model wybor * wybor(2) = mez_ce kob_ce opak marka1 marka2 marka3
marka4 ...
marka27 /ties=breslow;
strata pytanie;
run;
```

<sup>7</sup> Oczywiście pod warunkiem, że dana marka ma takie opakowanie.

Tabela 1. Fragment danych CBC przystosowany do estymacji za pomocą procedury PHREG programu SAS

Pytanie	Wybor	Marka1	Marka2	Marka3	***	Marka27	Rezygnacja	Opak.	Wiek1_ce	Wiek2_ce	Wiek3_ce	Wiek4_ce
1	2	0	0	0	***	0	0	2	0	0	0	3,20
1	2	0	0	0	***	0	1	0	0	0	0	0,00
1	1	1	0	0	***	0	0	2	0	0	0	3,22
1	2	0	1	0	***	0	0	1	0	0	0	3,49
1	2	0	1	0	***	0	0	2	0	0	0	3,60
1	2	0	0	0	***	0	0	1	0	0	0	2,59
1	2	0	0	0	***	0	0	2	0	0	0	2,99
1	2	0	0	0	***	0	0	1	0	0	0	3,05
1	2	0	0	0	***	0	0	2	0	0	0	3,77
1	2	0	0	0	***	0	0	1	0	0	0	3,00
1	2	0	0	0	***	0	0	2	0	0	0	3,40
1	2	1	0	0	***	0	0	1	0	0	0	3,25
1	2	0	0	0	***	0	0	1	0	0	0	2,67
2	1	1	0	0	***	0	0	2	0	0	0	3,22
2	2	0	1	0	***	0	0	1	0	0	0	3,49

Źródło: opracowanie własne.

W tabeli 2 przedstawiono fragment wyników kompilacji programu – zawarto tutaj podstawowe informacje na temat szacowanego modelu, na przykład podano informację o zbieżności algorytmu Newtona-Raphsona (N-R) oraz statystyki pokazujące jakość dopasowania modelu. Dopasowanie modelu jest mierzone za pomocą kryterium funkcji wiarygodności, kryterium informacyjnego Akaike i kryterium Schwarzera. Im mniejsza wartość tych kryteriów, tym dopasowanie modelu lepsze. Testem stosunku wiarygodności, testem punktowym i testem Walda weryfikuje się hipotezę, że wszystkie parametry są równe zero. Jak wiadomo, przyjęcie tej hipotezy oznacza, że zmienne włączone do modelu są statystycznie nieistotne.

Tabela 2. Wyniki kompilacji programu 2

Kryterium zbieżności (GCONV=1E-8) spełnione.				Zmienna	Ocena parametru	Błąd std.	Wsp. Chi-kwad.
Kryterium	Bez zmiennych towarzyszących	Ze zmiennymi towarzyszącymi					
-2 Log L	191 570	175 032		mez_ce	-0,60	0,03	426
AIC	191 570	175 090		kob_ce	-0,52	0,03	305
SBC	191 570	175 338		opak	-0,49	0,02	714
Testowanie globalnej hipotezy zerowej: BETA=0				marka1	1,27	0,09	214
Test	Chi-kwad.	Pr > chi kw.		marka2	1,22	0,09	169
Iloraz wiaryg.	16 538	<.0001		marka3	0,11	0,16	0
Ocena	16 750	<.0001		marka4	1,56	0,13	147
Wald	14 574	<.0001		***	***	***	***
				marka27	0,74	0,08	94

Źródło: opracowanie własne.

Patrząc na otrzymane rezultaty, można zauważyć, że zbieżność algorytmu N-R została osiągnięta, a zmienne włączone do modelu polepszają jego dopasowanie i są statystycznie istotne. W prawej części tabeli 2 podano oszacowania wartości parametrów modelu, ich błędy standardowe i statystyki testujące ich istotność. Ze względu na dużą liczbę szacowanych parametrów podano tylko wybrane wartości. Zdecydowana większość oszacowanych parametrów, charakteryzujących marki produktów jest statystycznie istotna, z bardzo dużą wartością statystyki chi-kwadrat, nieistotne są tylko dwie marki, oznaczone liczbami 3 i 14. Pozostałe zmienne, czyli cena zarówno w przypadku mężczyzn jak i kobiet oraz opakowanie, są statystycznie istotne. Interesujący i zgodny z intuicją jest fakt, że mężczyźni są bardziej wrażliwi na cenę piwa niż kobiety. Prawdopodobnie jest to spowodowane zdecydowanie większym spożyciem przez nich piwa, co powoduje, że udział tych wydatków w dochodzie mężczyzn jest większy niż w przypadku kobiet<sup>8</sup>. Producenci piwa wykorzystują tę zależność do ustalania cen produktów – zwykle piwa skierowane do kobiet są droższe od przeciętnych piw konsumowanych przez mężczyzn. Na przykład według danych AC Nielsen z 2010 roku średnia cena za butelkę 0,5 litra piwa Karmi i Redd's, pitych głównie przez kobiety, wynosiła 3,23 zł i 3,60 zł, a takich piw, jak Żywiec, Lech Premium, Tyskie, Żubr, kształtowała się odpowiednio na poziomie: 3,12 zł, 3,03 zł, 2,53 zł, 2,83 zł.

W tabeli 3 syntetycznie przedstawiono wyniki kolejnych analiz, w których badano, jak cena wpływa na postrzeganą użyteczność piwa w zależności od wieku konsumenta, wielkości tygodniowego spożycia i zamożności.

Ze względu na wiek konsumenta wyróżniono 4 przedziały – 18–24 lata, 25–34 lata, 35–44 lata, 45–65 lat. Niestety, autorzy nie zbadali grupy wiekowej powyżej 65. roku życia. Otrzymane wyniki są dosyć niespodziewane, mianowicie wiek w minimalny sposób wpływa na wrażliwość cenową konsumenta. Jedynie w grupie 45–65 lat, w której jest stosunkowo dużo emerytów i rencistów, wzrost ceny powoduje silniejszy spadek użyteczności, niż w przypadku osób poniżej 45. roku życia. W przedziale między 18 a 45 lat przeciętny wpływ ceny na użyteczność piwa jest praktycznie taki sam.

---

<sup>8</sup> Jak wiadomo, wraz ze wzrostem udziału wydatków rośnie wrażliwość cenowa konsumenta [10, s. 84–104].

Tabela 3. Oszacowania wybranych parametrów i ich błędy standardowe

Wpływ ceny na użyteczność piwa w zależności od:		Ocena	Błąd
		parametru	standardowy
wieku	18–24	–0,54	0,03
	25–34	–0,53	0,03
	35–44	–0,53	0,03
	45–65	–0,63	0,03
spożycia	< 1l	–0,61	0,03
	1l–2l	–0,50	0,03
	> 2l	–0,56	0,03
dochodu	pieniędzy nie starcza nawet na najpilniejsze potrzeby	–0,41	0,06
	musimy odmawiać sobie wielu rzeczy	–0,54	0,03
	na co dzień pieniędzy starcza, ale nie stać nas na większe wydatki	–0,60	0,03
	pieniędzy starcza na wszystkie wydatki, a część odkładamy	–0,53	0,03
	jesteśmy zamożni, nie musimy oszczędzać	0,21	0,10
płci	mężczyźni	–0,60	0,03
	kobiety	–0,52	0,03

Źródło: opracowanie własne.

W kolejnym kroku próbowano znaleźć zależność między deklarowaną wielkością spożycia a wpływem ceny na użyteczność piwa dla konsumenta. Teoretycznie wraz ze wzrostem spożycia piwa powinna rosnać wrażliwość cenowa konsumenta. W badaniu wyróżniono 4 podgrupy – osoby spożywające do litra piwa tygodniowo, osoby spożywające około dwóch litrów piwa tygodniowo oraz grupę osób pijącą powyżej dwóch litrów tygodniowo. Wyróżniono także grupę osób, które odmówiły dopowiedzi lub stwierdziły, że nie pamiętają, jakie spożywają ilości piwa.

Otrzymane wyniki są niejednoznaczne. O ile nie dziwi fakt, że osoby pijące powyżej dwóch litrów piwa tygodniowo są bardziej wrażliwe od osób pijących od litra do dwóch litrów, o tyle na pewno może dziwić uzyskanie bardzo dużego oszacowania parametru dla osób deklarujących spożycie poniżej litra tygodniowo. Prawdopodobnie osoby konsumujące niewielkie ilości piwa nie akceptują jego wysokich cen i bez problemu są w stanie z niego zrezygnować lub zastąpić innym napojem (ewentualnie alkoholem). Chcąc potwierdzić tę hipotezę, należałoby przeprowadzić bardziej szczegółowe badanie tej grupy konsumentów.

Z analizy wpływu ceny na użyteczność piwa w zależności od sytuacji ekonomicznej konsumenta wynika, że praktycznie tylko grupa zamożnych osób, niemuszających oszczędzać, istotnie różni się od pozostałych. Osoby te nie zwracają uwagi na cenę produktu, a nawet kupują te, które są zdecydowanie droższe od przeciętnych. Otrzymany wynik potwierdzają dane sprzedażowe firmy AC Nielsen, mianowicie segment piw Premium, czyli piw o wyższej jakości i cenie, jest w Polsce bardzo mały, ponieważ zależnie od metody liczenia stanowi maksymalnie kilka procent. Podobnie było w badaniu – odsetek osób deklarujących się jako zamożne także był niewielki. Pozostałe grupy osób podobnie oceniały wpływ ceny na postrzeganą użyteczność produktu.

## Podsumowanie

W artykule omówiono mieszany model logitowy, inaczej nazywany modelem z parametrami losowymi. Model ten jest bardzo elastyczny i pozwala na przybliżenie dowolnego modelu użyteczności losowej. Praktyczne wykorzystanie mieszanego modelu logitowego pokazano na przykładzie analizy konsumentów rynku piwowarskiego w Polsce. Do estymacji parametrów modelu wykorzystano program SAS. Pokazano, w jakim stopniu wiek, płeć, sytuacja materialna respondenta wpływają na postrzeganą przez niego użyteczność piwa. Otrzymane wyniki są w większości przypadków zgodne z intuicją – mężczyźni są bardziej wrażliwi na cenę piwa niż kobiety, natomiast osoby bardzo bogate nie przywiązują większej uwagi do ceny piwa (nawet można zauważyć dodatnią korelację między ceną a popytem na piwo w tej grupie respondentów). Wrażliwość na cenę rośnie wśród osób starszych, w wieku powyżej 45 lat. Nie znaleziono natomiast wyraźnej zależności między tygodniową wielkością spożycia a wrażliwością cenową. Być może rozważany model był zbyt ogólny i należałoby przeprowadzić bardziej szczegółowe badania.

## Literatura

1. Bąk A., *Analiza danych o preferencjach z wykorzystaniem modeli kategorii nieuporządkowanych i programu R*, Taksonomia 17, *Klasyfikacja i analiza danych – teo-*

- ria i zastosowania, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2010.
2. Bąk A., *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2004.
  3. Bąk A., *Modele wyborów dyskretnych i ich estymacja w programie R*, Taksonomia 16, *Klasyfikacja i analiza danych – teoria i zastosowania*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2009.
  4. Frątczak E., Babiker H., Gach-Ciepiela U., *Analiza historii zdarzeń – elementy teorii, wybrane przykłady zastosowań*, Wydawnictwo SGH, Warszawa 2005.
  5. Hensher D.A., Greene W.H., *The Mixed Logit Model: The State of Practice*, „Transportation” 2003, Vol. 30, No. 2.
  6. Lee E.T., Wang W.J., *Statistical Methods for Survival Data Analysis*, 3rd edit., John Wiley and Sons, New Jersey 2003.
  7. McFadden D.L., *Econometrics Models of Probabilistic Choice*, w: C.F. Manski, D. McFadden, *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge 1981.
  8. McFadden D., Train K., *Mixed MNL Model for Discrete Response*, „Journal of Applied Econometrics” 2000, Vol. 15 (5).
  9. *Mikroekonometria. Modele i metody analizy danych indywidualnych*, red. M. Gruszczyński, Wydawnictwo Oficyna, Warszawa 2010.
  10. Nagle T., *The Strategy and Tactics of Pricing*, Prentice-Hall, Englewood Cliffs 1987.
  11. Orme B., *Which Conjoint Method Should I Use?*, Research Paper Series, Sequim, Sawtooth Software, <http://www.sawtoothsoftware.com/download/techpap/whichmth.pdf>, dostęp: 12.02.2011.
  12. Orme K.B., Heft A.M., *Predicting Actual Sales with CBC: How Capturing Heterogeneity Improves Results*, Research Paper Series, Sequim, Sawtooth Software <http://www.sawtoothsoftware.com/download/techpap/predict.pdf>, dostęp 12.02.2011.
  13. Rószkiewicz M., *Narzędzia statystyczne w analizach marketingowych*, Wydawnictwo C.H. Beck, Warszawa 2002.
  14. So Y., Kuhfeld W.F., *Multinomial logit models*, SAS Technical Papers, SAS Institute, <http://support.sas.com/techsup/technote/mr2010g.pdf>, dostęp 12.02.2011.
  15. *Statystyczna analiza danych z wykorzystaniem programu R*, red. M. Walesiak, E. Gatnar, E., Wydawnictwo Naukowe PWN, Warszawa 2009.
  16. Train K., *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA 2002.

---

## APPLICATION OF MIXED LOGIT MODEL IN MEASURING BEER CONSUMERS PREFERENCES

### Summary

Mixed logit model, also called random parameters logit does not exhibit the restrictive “independence from irrelevant alternatives” property and explicitly accounts for correlations in unobserved utility over repeated choices by each customer. Mixed logit model is very flexible and allows for approximate any random utility model. A practical application of a mixed logit model is presented in the example of measuring beer consumers preferences. Model parameters were estimated using procedure PHREG in SAS program.

**Keywords:** mixed logit model, random parameters logit, beer industry, consumer preferences, SAS.

*Translated by Michał Purczyński*