PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1096 — 2005
Ekonometria 15

Robert Kapłon

MODEL LOGITOWY Z KLASAMI UKRYTYMI

1. Wstęp

Dość często konsumenci są przedmiotem badań marketingowych. Jak wiadomo, różnią się pod wieloma względami. Pomimo to, konwencjonalne metody i procedury analizy danych marketingowych opierają się na założeniu, że badana próba jest homogeniczna. W konsekwencji otrzymuje się estymatory parametrów modelu, które są identyczne dla każdego konsumenta. Przykładowo, w prostym modelu pozwalającym uchwycić, jak zmiana ceny wpłynie na popyt, można otrzymać uśrednione wyniki, mimo że wcześniejsze badania wyraźnie wyłaniają dwie skrajne grupy konsumentów, z których jedna jest bardzo wrażliwa na zmianę ceny, druga natomiast – prawie wcale. Z innego przykładu [Dillon, Kumar 1994] można wyciągnąć wniosek, że nieuwzględnienie heterogeniczności jest przyczyną bardzo słabego dopasowania modelu do danych. W przykładzie tym przedstawiono dane, dotyczące ilości nabywanych słodyczy w ciągu 7 dni. Posłużono się rozkładem Poissona. Jedyny estymowany parametr to średnia. Naturalnie, model był źle dopasowany do danych empirycznych, dlatego powtórzono procedurę z uwzględnieniem heterogeniczności. Wyodrębniono siedem segmentów i otrzymano bardzo dobre rezultaty.

W świetle powyższych wywodów nie może dziwić, że uwzględnienie heterogeniczności przyczynia się do lepszego odzwierciedlenia struktury badanych. Dlatego dość często wykorzystuje się analizę klas ukrytych (lub mieszaninę rozkładów). W.A. Kamakura i G.J. Russell [1989], opierając się na mieszankach rozkładów, wykorzystali wielomianowy model logitowy do zbadania struktury cenowej – wyodrębnili jednocześnie segmenty. W analizie klas ukrytych A.K. Formann [1992] skupia się na regresji logistycznej.

W niniejszej pracy przedstawiono model logitowy, którego podstawą jest idea wielomianowego modelu logitowego. Takie ujęcie zagadnienia pozwoliło zwrócić uwagę na strategię decyzyjną konsumenta i jednocześnie porównać przedstawiony

model z zaproponowanym przez A.K. Formanna. Pokazano również, że w zależności od przyjętych założeń model opiera się na mieszankach rozkładów lub analizie klas ukrytych.

2. Sformułowanie modelu logitowego z klasami ukrytymi

Niech $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iJ})^{\mathrm{T}}$ będzie wektorem losowym, natomiast jego realizacją $\mathbf{y}_i^* = (y_{i1}^*, y_{i2}^*, \ldots, y_{iJ}^*)^{\mathrm{T}}$, przy czym y_{ij}^* przyjmuje wartość 1, gdy konsument i ($i = 1, 2, \ldots, N$) wybierze markę j ($j = 1, 2, \ldots, J$), oraz zero w przeciwnym wypadku. Ponadto przyjmuje się, że badana zbiorowość (konsumenci) dzielona jest na wzajemnie rozłączne, bezpośrednio nieobserwowalne klasy (segmenty). Z podziałem tym związana jest wprowadzona zmienna ukryta $\theta = (q_1, q_2, \ldots, q_S)^{\mathrm{T}}$, ($s = 1, 2, \ldots, S$) (por. [Kapłon 2002]). Zgodnie z istotą modeli zmiennych latentnych Y_{i1}, Y_{i2}, ..., Y_{iJ} ma warunkowy rozkład ze względu na q_1, q_2, ..., q_S (por. [Everitt 1987]). Zmienna ukryta jest dyskretna, dlatego niech prawdopodobieństwo tego, że losowo wybrany konsument znajdzie się w klasie C_s, wynosi $h(q_s)$, i oczywiście musi zachodzić: $\sum_s h(\theta_s) = 1$.

W kontekście tak postawionego zagadnienia, niech wektor losowy $\mathbf{Y}_i$ ma wielowymiarowy, warunkowy rozkład Bernoulliego:

$$\mathbf{Y}_i | \theta_s \sim B(1, p_{is}),\ p_{is} = \left(P_i^s(1), \ldots, P_i^s(j), \ldots, P_i^s(J)\right), \tag{1}$$

biorący się z założenia o lokalnej niezależności (zob. [Kapłon 2002]). Prawdopodobieństwo tego, że marka j zostanie wybrana przez konsumenta i należącego do segmentu C_s, oznaczono przez $P_i^s(j)$. Ze względu na wzór (1), warunkowe prawdopodobieństwo wyraża się w postaci:

$$\Pr(\mathbf{Y}_i | \theta_s; \mathbf{X}_i, \psi_s) = \prod_{j=1}^{J} \left[P_i^s(j)\right]^{y_{ij}^*} \left[1 - P_i^s(j)\right]^{1 - y_{ij}^*}, \tag{2}$$

gdzie: ψ_s jest wektorem nieznanych parametrów w klasie C_s natomiast $\mathbf{X}_i$ reprezentuje macierz, która zostanie zdefiniowana później. Dyskretna natura zmiennej latentnej pozwala wyrazić prawdopodobieństwo bezwarunkowe wzorem:

$$\Pr\left(\mathbf{Y}_i = \mathbf{y}_i^* \middle| \mathbf{X}_i, \Psi\right) = \prod_{s=1}^{S} h(\theta_s) \Pr\left(\mathbf{Y}_i = \mathbf{y}_i^* \middle| \theta_s; \mathbf{X}_i, \psi_s\right). \tag{3}$$

Aby zdefiniować prawdopodobieństwa wyboru marek, niech zgodnie z ideą wielomianowego modelu logitowego (por. [McFadden 1974]), użyteczność, jaką ma dla konsumenta i marka j w segmencie C_s, będzie wyrażona w stochastyczny sposób, tzn.:

$$U_{ij|s} = \beta_s^{\mathrm{T}} \mathbf{x}_{ij} + \varepsilon_{ij|s}, \tag{4}$$

gdzie: $\beta_s^{\mathrm{T}} = (\beta_{1s}, \beta_{2s}, \ldots, \beta_{As})$ reprezentuje wektor parametrów w klasie C_s, ocena konsumenta i atrybutu a ($a = 1, 2, \ldots, A$) marki j jest równa x_{ija}, co oczywiście pozwala zapisać $\mathbf{x}_{ij} = \left(x_{ij1}, \ldots, x_{ija}, \ldots, x_{ijA}\right)^{\mathrm{T}}$ i zdefiniować macierz $\mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iJ}]$, natomiast $\varepsilon_{ij|s}$ jest tą częścią użyteczności w segmencie C_s, której nie można bezpośrednio zaobserwować; w tym wypadku zakłada się rozkład ekstremalnych wartości typu I o funkcji gęstości:

$$f\left(\varepsilon_{ij|s}\right) = e^{-\varepsilon_{ij|s}} \exp\left(-e^{-\varepsilon_{ij|s}}\right), \tag{5}$$

przy czym

$$\forall_{\substack{i \neq n \\ j \neq m}} \operatorname{cov}\left(\varepsilon_{ij|s}, \varepsilon_{nm|s}\right) = 0.$$

Wypada w tym miejscu nawiązać do wzoru (3) i przyjętej notacji. Okazuje się bowiem, że przyjęcie macierzy $\mathbf{X}_i$ różnej dla każdego konsumenta nie pozwala nawiązywać do analizy klas ukrytych lecz mieszanek rozkładów. Tak też należy traktować prawdopodobieństwo (3). Aby z kolei wykorzystać analizę klas ukrytych, trzeba przyjąć ograniczenie na wspominaną macierz. Chociaż dalsze analizy będą prowadzone z uwzględnieniem macierzy różnych dla każdego konsumenta, to jednak w przykładzie empirycznym wybrano wariant z klasami ukrytymi.

W dalszej kolejności należy rozstrzygnąć, kiedy marka zostanie wybrana. Ze względu na istotę analizy klas ukrytych – i związaną z tym możliwość wyboru kilku marek – nie można wykorzystać strategii, która jest podstawą wielomianowego modelu logitowego (*MNL*). Dzieje się tak dlatego, że w modelu tym (*MNL*) definiuje się prawdopodobieństwo wyboru jednej marki, a tym samym przyjmuje się, że marka o największej użyteczności będzie wybrana. Z tego też względu dla prezentowanego modelu można przyjąć strategię, zgodnie z którą konsumenci wybiorą marki o użyteczności przekraczającej pewien poziom Δ_s. Przyjmujemy jego losowy charakter, zatem niech w pewnym segmencie C_s $\Delta_s \equiv \delta_s + \varepsilon_s$, przy czym δ_s i ε_s odzwierciedlają odpowiednio deterministyczną i losową naturę przyjętego poziomu. Wtedy interesujące nas prawdopodobieństwo można zapisać w postaci:

$$P_i^s(j) = \Pr\left(U_{ij|s} > \Delta_s\right) = \Pr\left(\varepsilon_s < \beta_s^{\mathrm{T}}\mathbf{x}_{ij} - \delta_s + \varepsilon_{ij|s}\right) =$$
$$= \int_{-\infty}^{\infty} f\left(\varepsilon_{ij|s}\right) F\left(\beta_s^{\mathrm{T}}\mathbf{x}_{ij} - \delta_s + \varepsilon_{ij|s}\right) d\varepsilon_{ij|s}, \tag{6}$$

gdzie $F(\cdot)$ jest dystrybuantą funkcji gęstości zdefiniowanej wzorem (5). Po pewnych przekształceniach prawdopodobieństwo (6) można zapisać w zwięzłej formie:

$$P_i^s(j) = \left[1 + \exp\left(\delta_s - \beta_s^{\mathrm{T}}\mathbf{x}_{ij}\right)\right]^{-1}. \tag{7}$$

Tak zdefiniowane prawdopodobieństwo ma postać regresji logistycznej. Czy wobec tego istniała uzasadniona potrzeba szczegółowego wyprowadzenia wzoru (7), skoro jego postać jest ogólnie znana? Nie dość wspomnieć, że już A.K. Formann [1992] proponuje model regresji logistycznej z klasami ukrytymi. W odróżnieniu jednak od wspomnianego modelu i znanej regresji logistycznej zaakcentowano fakt, że konsument, wybierając jakąś markę, posługuje się strategią decyzyjną. Przyjęto więc, że zostaną wybrane te produkty, których użyteczność przekroczy pewien poziom. Przenosząc ideę tej strategii na model Formanna, okazuje się, że tym poziomem jest wartość równa zeru[1]. Widać więc, że prawdopodobieństwo (7) zostało zdefiniowane w bardziej ogólny sposób. Jak się później okaże – w części empirycznej pracy – nie w każdym segmencie można przyjąć za ten poziom wartość zero.

Dla tak sformułowanego modelu poszukujemy wektora Ψ nieznanych parametrów, maksymalizując poniższą funkcję wiarygodności:

$$L\left(\Psi|\mathbf{Y}\right) = \prod_{i=1}^{N} \Pr\left(\mathbf{Y}_i = \mathbf{y}_i^* \middle| \mathbf{X}_i, \Psi\right). \tag{8}$$

3. Procedura poszukiwania estymatorów największej wiarygodności

W praktyce wykorzystanie funkcji (8) nastręcza wielu problemów. Można ich jednak uniknąć, jeśli wykorzystuje się algorytm **EM** zaproponowany przez A.P. Dempstera, N.M. Lairda i D.B. Rubina [1977]. Szczegółowe omówienie można znaleźć w monografii G.J. McLachlana, T. Krishnan [1997].

Traktując macierz zmiennych latentnych $\mathbf{Z} = \left[z_{is}\right]_{N \times S}$ jako „brakujące" dane, definiujemy funkcję wiarygodności dla kompletnego zbioru, czyli ($\mathbf{Y}$, $\mathbf{Z}$). W tym celu przyjmujemy, że

[1] Inną cechą odróżniającą model Formanna od przedstawianego w pracy jest reparametryzacja prawdopodobieństw przynależności do klasy tego pierwszego.

$$\mathbf{X}_i z_{is} = \begin{cases} 1, & \text{gdy konsument } i \in C_s, \\ 0, & \text{w przeciwnym wypadku,} \end{cases} \tag{9}$$

o wielomianowym rozkładzie postaci:

$$\mathbf{z}_i | h(\theta) \sim M(1, h(\theta)). \tag{10}$$

Wtedy logarytm funkcji wiarygodności dla kompletnego zbioru danych ma postać:

$$\log L_c(\Psi|\mathbf{Y}, \mathbf{Z}) = \sum_{i=1}^{N}\sum_{s=1}^{S} z_{is} \sum_{j=1}^{J} \left\{ y_{ij}^* \log P_i^s(j) + \left(1 - y_{ij}^*\right) \log\left[1 - P_i^s(j)\right] \right\} + \\ + \sum_{i=1}^{N}\sum_{s=1}^{S} z_{is} \log h(\theta_s). \tag{11}$$

Do powyższej funkcji wiarygodności stosuje się wspomniany algorytm **EM**, na który składają się dwa kroki:

Krok E: W iteracji $(k + 1)$ obliczyć $Q\left(\Psi|\Psi^{(k)}\right)$, przyjmując za nieznany wektor parametrów Ψ początkowe wartości równe $\Psi^{(k)}$, gdzie

$$Q\left(\Psi|\Psi^{(k)}\right) = E_{\Psi^{(k)}}\left[\log L_c\left(\Psi|\mathbf{Y}, \mathbf{Z}\right)|\mathbf{Y}, \Psi^{(k)}\right)\right]. \tag{12}$$

Ponieważ

$$E_{\Psi^{(k)}}\left[Z_{is}|\mathbf{Y}\right] = K_{is}^{(k)}, \tag{13}$$

gdzie

$$K_{is}^{(k)} = \frac{h(\theta_s)\Pr\left(\mathbf{Y}_i = \mathbf{y}_i^* \middle| \theta_s; \mathbf{X}_i, \psi_s^{(k)}\right)}{\sum_s h(\theta_s)\Pr\left(\mathbf{Y}_i = \mathbf{y}_i^* \middle| \theta_s; \mathbf{X}_i, \psi_s^{(k)}\right)}, \tag{14}$$

dlatego funkcja (12) przekształca się do następującej postać:

$$Q\left(\Psi|\Psi^{(k)}\right) = \sum_{i=1}^{N}\sum_{s=1}^{S} K_{is}^{(k)} \sum_{j=1}^{J} \left\{ y_{ij}^* \log P_i^s(j) + \left(1 - y_{ij}^*\right) \log\left[1 - P_i^s(j)\right] \right\} + \\ + \sum_{i=1}^{N}\sum_{s=1}^{S} K_{is}^{(k)} \log h(\theta_s). \tag{15}$$

Krok M: Należy maksymalizować wartość oczekiwaną, która została obliczona w poprzednim kroku **E**, czyli

$$\Psi^{(k+1)} = \arg\max_{\Psi} Q\left(\Psi|\Psi^{(k)}\right). \tag{16}$$

W wypadku estymatorów prawdopodobieństwa przynależności do klas możliwe jest podanie ich postaci analitycznej. Odwołując się do (16), należy obliczyć odpowiednie pochodne cząstkowe. Przyrównanie ich do zera pozwala otrzymać następującą wartość:

$$\hat{h}(\theta_s) = \frac{1}{N}\sum_{i=1}^{N} K_{is}^{(k)}. \tag{17}$$

Aby oszacować pozostałe parametry, tzn. $\{\psi_s\}_{s=1}^{S}$, zgodnie z (16), można wykorzystać algorytm Newtona-Raphsona. W metodzie tej wymaga się podania gradientu i macierzy hesianu [Everitt 1987]. Dla klarowności dalszych zapisów przyjęto indeksowanie atrybutów od zera i wtedy $\forall_{i,j} x_{ij0} = -1$, a tym samym $\beta_{0s} = \delta_s$. Na podstawie (15) obliczono gradient[2]

$$\mathbf{g}^{(k)} = \frac{\partial Q\left(\Psi|\Psi^{(k)}\right)}{\partial \beta} = \left[\lambda_{01} \quad \lambda_{11} \quad \ldots \quad \lambda_{a1} \quad \ldots \quad \lambda_{A1} \quad \ldots \quad \lambda_{as} \quad \ldots \quad \lambda_{AS}\right]^{\mathrm{T}}, \tag{18}$$

gdzie

$$\lambda_{as}^{(k)} = \sum_i \sum_j K_{is}^{(k)} x_{ija}\left[y_{ij}^{*} - P_i^s(j)\right],$$

oraz macierz hesianu

$$\mathbf{H}^{(k)} = \frac{\partial^2 Q\left(\Psi|\Psi^{(k)}\right)}{\partial \beta \partial \beta^{\mathrm{T}}} = \mathrm{diag}\left[\mathbf{H}_1^{(k)} \ldots \mathbf{H}_s^{(k)} \ldots \mathbf{H}_S^{(k)}\right], \tag{19}$$

w której każdy element diagonalny jest kwadratową macierzą wymiaru $A + 1$, przy czym

[2] Pominięto tutaj kroki algorytmu Newtona-Raphsona. Wyróżniony krok k odnosi się tylko do algorytmu **EM**.

$$\mathbf{H}_s^{(k)} = \begin{bmatrix} h_{00}^{(k)} & \dots & h_{a0}^{(k)} & \dots & h_{A0}^{(k)} \\ \vdots & & \vdots & \vdots & \vdots \\ h_{0a}^{(k)} & \dots & h_{aa}^{(k)} & \dots & h_{Aa}^{(k)} \\ \vdots & & \vdots & & \vdots \\ h_{0A}^{(k)} & \dots & h_{aA}^{(k)} & \dots & h_{AA}^{(k)} \end{bmatrix}; \tag{20}$$

$$h_{ab}^{(k)} = -\sum_{i=1}^{N}\sum_{j=1}^{J} K_{is}^{(k)} x_{ija} x_{ijb} P_i^s(j)\left[1 - P_i^s(j)\right], \; a, b = 0, 1, \dots, A.$$

Po oszacowaniu nieznanych parametrów $\{\psi_s\}_{s=1}^{S}$ powraca się do kroku **E**, a później znowu do kroku **M**. Kroki te powtarza się naprzemiennie, dopóki wartość funkcji wiarygodności nie zmieni się o małą, arbitralnie ustaloną wartość.

4. Przykład empiryczny

Na potrzeby niniejszego artykułu przeprowadzono badania ankietowe, dotyczące ośmiu marek kawy: Exclusive Mild, Tchibo Family, Grand Mocca, Jacobs Merido, Jacobs Krönung, Maxwell House, Nescafe Classic oraz Nescafe Gold. W wyniku redukcji wymiarowości przestrzeni percepcji drogą analizy czynnikowej wyodrębniono cztery zmienne:

X_1 – walory smakowe I (gorycz, smak, kwaskowatość),

X_2 – wizerunek (cena, prestiż),

X_3 – walory smakowe II (aromat, moc, posmak),

X_4 – dostępność.

Po ocenie atrybutów wymienionych kaw konsumenci mieli powiedzieć, które marki zakupiliby. Zgodnie z koncepcją modelu, mogli zdecydować się na kilka produktów.

Najpierw należy ustalić liczbę klas. Ponieważ modele o różnych liczbach klas nie są względem siebie hierarchiczne, dlatego test statystyczny oparty na podwojonym logarytmie ilorazu funkcji wiarygodności nie może być użyty [Titterington, Smith, Markov 1985]. W takim wypadku odwołuje się często do kryteriów informacyjnych [Kapłon 2002]. Oszacowano więc parametry dla czterech modeli, tzn. z jedną klasą, dwoma itd., zestawiając odpowiednie wartości w tab. 1. Do estymacji parametrów napisano odpowiedni kod programu w środowisku MATLAB.

Wszystkie modele oprócz pierwszego (1 klasa) należy uznać za dobrze opisujące badane zjawisko. Przemawiają za tym niskie, w porównaniu ze stopniami swo-

Tabela 1. Logarytm funkcji wiarygodności oraz kryteria informacyjne dla modeli z różną liczbą klas

Liczba klas	Log$L$	AIC	AIC*	CAIC	BIC	BIC*
1	−1089,0139	2188,0279	−128,7519	2210,1965	2205,1965	−987,1056
2	−992,8889	2007,7778	−308,9253	2056,5487	2045,5487	−1146,7534
3	−980,0206	1994,0411	−322,6619	2069,4144	2052,4144	−1139,8877
4	−975,4287	1996,8574	−319,8457	2098,8330	2075,8330	−1116,4692

Źródło: opracowanie własne.

body, wartości statystyk testowych, tj.: chi-kwadrat, G^2, Cressie-Read[3]. W wypadku odrzuconego modelu są one wysokie i wynoszą odpowiednio 575,33, 371,32 i 582,28 przy 251 stopniach swobody. Dlatego należy się oprzeć się na wynikach z tab. 1 i wybrać ten spośród pozostałych trzech, dla których wartości kryteriów są najniższe. Okazuje się (zacienione pola reprezentują najniższe wartości), że najlepszym kandydatem jest model z dwoma klasami ukrytymi. Dalsza analiza będzie dotyczyła tego modelu.

Pojawia się tu jednak naturalne pytanie, czy model z trzema klasami – wobec najniższych kryteriów AIC – nie jest też dobrym kandydatem. Za jego odrzuceniem przemawiają dwa argumenty. Pierwszy odnosi się bezpośrednio do „preferowania" przez AIC modeli nazbyt rozbudowanych [Kass, Raftery 1995]. Dlatego H. Bozdogan [1987] zaproponował kryterium CAIC, które nakłada większą karę na modele o zbyt dużej liczbie parametrów. Ponieważ kryteria BIC i BIC* również preferują modele mniej złożone [Kass, Raftery 1995], dlatego rozsądne wydaje się być poleganie na BIC, BIC* oraz CAIC. Oczywiście, model z dwoma klasami, posiada tam wartości najmniejsze. Drugi argument utwierdza w przekonaniu o zrezygnowaniu z jednej klasy, gdyż prawdopodobieństwo przynależności do pierwszej z trzech klas jest niskie: $h(\theta_1) = 0{,}12$.

Przyjąwszy model z dwoma klasami, w tab. 2 prezentujemy wyniki estymacji. Warto odnotować wartości statystyk testowych. Statystyka chi-kwadrat wyniosła 188,12 ($p < 0{,}997$), $G^2 = 179{,}07$ ($p < 0{,}999$), natomiast Cressie-Read równa jest 195,6 ($p < 0{,}991$). W nawiasie podano, z uwzględnieniem 245 stopni swobody,

Tabela 2. Wartości szacowanych parametrów – estymatory parametrów modelu w odpowiedniej klasie

Klasa	β_0	β_1	β_2	β_3	β_4	Prawdopodobieństwa przynależności do klasy
Pierwsza	2,3742	−5,1636	−1,7879	−0,4096	2,7920	0,3703
Druga	0.4277	−2,7044	3,1308	−0,1079	0,0002	0,6297

Źródło: opracowanie własne.

[3] Wspomniane statystyki są często wykorzystywanymi miernikami dopasowania modeli do danych. Szczegółowe informacje można znaleźć w wielu pracach, np. [Dayton 1998]. Wartości owych statystyk zostaną później podane dla najlepszego modelu.

tzw. *p-value*. Uznaje się więc model za bardzo dobrze odzwierciedlający strukturę danych.

W segmencie pierwszym, mniej licznym (0,3703), użyteczność produktu powinna być znacznie wyższa od użyteczności w segmencie drugim. O tym informuje wartość progu krytycznego równa 2,3742. Pokazuje to również, o czym warto wspomnieć w kontekście wcześniejszych rozważań, że ustalanie progu krytycznego na poziomie zero może przyczynić się do znacznych błędów.

W rozważanej klasie dostępność (2,7920) jest ważnym atrybutem decydującym o wyborze. Inaczej jest w drugim segmencie, w którym atrybut ten nie ma większego znaczenia (0,0002). Różnica między segmentami pojawia się też na poziomie wizerunku marki. Marki o większym prestiżu oraz wyższej cenie cieszą się większym powodzeniem w bardziej licznym segmencie. Konsumenci należący do klasy pierwszej zainteresowani są kawami o niskiej cenie (–1,7879). Wymagają również, aby cechowały je bardzo niska gorycz i kwaskowatość (–5,1636). Podobne preferencje mają badani należący do klasy drugiej, przy czym poziom atrybutów nie jest aż tak niski (–2,7044). W największym stopniu zgoda między segmentami dotyczy atrybutu wyrażającego walory smakowe II. Choć kawy słabsze, mniej aromatyczne o słabym posmaku mają większą szansę na wybór, to jednak wpływ tej zmiennej na zakup jest znikomy. Może to być przesłanką do wyeliminowania tej zmiennej ze zbioru zmiennych rozważanych[4].

W konkluzji stwierdza się, że analiza klas ukrytych wnosi wiele istotnych informacji do modelu logitowego. Pozwala rezygnować z kłopotliwego założenia dotyczącego jednorodności obserwacji. W konsekwencji jej wykorzystania otrzymuje się segmenty (klasy ukryte), w których preferencje są podobne. Pozwala to porównywać szacowane parametry między klasami. Jednak najważniejszą rzeczą jest to, jak pokazuje przykład empiryczny, że dla tak skonstruowanej strategii decyzyjnej jest możliwe uznanie modelu logitowego za dobrze opisujący badane zjawisko (dla dwóch klas). Gdyby jednak przyjąć jedną klasę, a więc podejście najczęściej stosowane, wtedy taki model należałoby odrzucić. Przemawiają za tym zbyt wysokie wartości statystyk testowych.

[4] Aby tym rozważaniom nadać wymiar statystyczny, należałoby wykorzystać statystykę *t*-Studenta, w celu zbadania istotności szacowanych parametrów. Zrezygnowano z tego, gdyż celem niniejszego artykułu były metodologiczne rozważania. Przykład natomiast stanowił ich uzupełnienie.

Literatura

Bozdogan H. (1987), *Model Selection and Akaike's Information Criterion: The General Theory and its Analytical Extensions*. „Psychometrika", vol. 52, nr 3, s. 345-370.

Dayton C.M. (1998), *Latent Class Scaling Analysis*, Sage University Papers Series on Quantitative Applications in the Social Sciences, Sage Thousand Oaks, CA.

Dempster A.P., Laird N.M., Rubin D.B. (1977), *Maximum Likelihood from Incomplete Data via the EM Algorithm*, „Journal of the Royal Statistical Society", ser. B, nr 1(39), s. 1-22.

Dillon W., Kumar A. (1994), *Advanced Methods of Marketing Research*, [w:] R.P. Bagozzi (red.), *Latent Structure and other Mixture Models in Marketing: An Integrative Survey and Overview*, Blackwell, Oxford, s. 295-351.

Everitt B.S. (1984), *An Introduction to Latent Variable Models*, Chapman and Hall.

Everitt B.S. (1987), *Introduction to Optimization Methods and their Application in Statistics*, Chapman and Hall.

Formann A.K. (1992), *Linear Logistic Latent Class Analysis for Polytomous Data*. „Journal of the American Statistical Association", vol. 87, s. 476-486.

Kamakura W.A., Russell G.J. (1989), *A Probabilistic Choice Model for Market Segmentation and Elasticity Structure*, „Journal of Marketing Research", vol. 26, s. 379-390.

Kaplon R. (2002), *Analiza danych dyskretnych za pomocą metody LCA*, Taksonomia 9, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 942, AE, Wrocław.

Kass R.E., Raftery A.E. (1995), *Bayes Factors*, „Journal of the American Statistical Association", vol. 90, s. 773-793.

McFadden D. (1974), *Conditional Logit Analysis of Qualitative Choice Behavior*, [w:] P. Zarembka (red.), *Frontiers in Econometrics*, Academic Press, New York, s. 105-142.

McLachlan G.J., Krishnan T. (1997), *The EM Algorithm and Extensions*, John Wiley, New York.

Titterington D.M., Smith A.F.M., Markov U.E. (1985), *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons.

LOGIT MODEL WITH LATENT CLASS ANALYSIS

Summary

Using a single logit model for an entire population is potentially dangerous in masking the differential effects due to consumer heterogeneity. Therefore, as a remedy a logit model are combined with a widely known statistical method as latent class analysis. Also, a more generic model based on finite mixture distribution is discussed. In particular author focusing on consumer decision strategy shows that logistic latent class analysis are a special case. The maximum likelihood function is optimized by means of an EM algorithm and Newton-Raphson in the M-step. The practical applicability of logit latent class analysis is demonstrated by real data.

Dr inż. Robert Kaplon jest pracownikiem Instytutu Organizacji i Zarządzania Politechniki Wrocławskiej.