

**Aneta Rybicka**

Uniwersytet Ekonomiczny we Wrocławiu

---

**MODELE KLAS UKRYTYCH  
W METODACH WYBORÓW DYSKRETNÝCH**

---

**Streszczenie:** W badaniu preferencji wyrażonych wykorzystujemy m.in. metody wyborów dyskretnych reprezentujących podejście dekompozycyjne. W związku z tym, że użyteczności cząstkowe oraz całkowite oszacowywane są na poziomie zagregowanym, niemożliwe jest bezpośrednie przeprowadzenie segmentacji konsumentów. W celu oszacowania, w metodach wyborów dyskretnych, użyteczności na poziomie segmentowym wykorzystujemy modele klas ukrytych. W artykule przedstawiono rodzaje modeli klas ukrytych, metodę estymacji parametrów, oprogramowanie komputerowe oraz przykłady zastosowań na podstawie literatury przedmiotu.

**Słowa kluczowe:** metody wyborów dyskretnych, segmentacja rynku, modele klas ukrytych.

## 1. Wstęp

W badaniach preferencji konsumentów wykorzystujemy dane historyczne oraz dane wyrażające intencje, deklaracje konsumentów. Opierając się na drugim typie danych, możemy przeprowadzić badania preferencji wyrażonych. W tego typu badaniach wykorzystujemy m.in. metody dekompozycyjne, tj. *conjoint analysis*, oraz metody wyborów dyskretnych.

Metody wyborów dyskretnych, w odróżnieniu od *conjoint analysis*, pozwalają oszacować użyteczności cząstkowe i całkowite na poziomie zagregowanym, w przekroju całej badanej grupy. Konsekwencją tego jest brak możliwości bezpośredniego przeprowadzenia segmentacji konsumentów. W celu oszacowania w metodach wyborów dyskretnych użyteczności na poziomie segmentowym wykorzystujemy modele klas ukrytych.

W artykule przedstawiono rodzaje modeli klas ukrytych, możliwość ich wykorzystania w segmentacji konsumentów metodami wyborów dyskretnych, metodę estymacji parametrów, oprogramowanie komputerowe wykorzystywane w badaniach tego typu oraz przykłady zastosowań na podstawie studiów literaturowych.

## 2. Charakterystyka metod wyborów dyskretnych

Metody wyborów dyskretnych to druga, obok *conjoint analysis*, grupa metod reprezentująca podejście dekompozycyjne.

Metody te charakteryzują się m.in. tym, że:

- odzwierciedlają rzeczywiste procesy zachodzące na rynku,
- pozwalają na rezygnację z wyboru (w sytuacji, gdy żaden z prezentowanych produktu nie satysfakcjonuje konsumenta),
- umożliwiają łączną ocenę wszystkich atrybutów jednocześnie,
- pozwalają na oszacowanie udziałów w rynku oraz prognozowanie udziałów nowo wprowadzanych produktów (wartości estymatorów parametrów modelu są obliczane na poziomie zagregowanym),
- wymagają mniejszej liczby rozstrzygnięć (niż *conjoint analysis*).

Dzięki metodom tym można odpowiedzieć m.in. na pytania [Hair i in. 1995, s. 562]:

- Jaka jest użyteczność każdego atrybutu?
- Jak ważny jest dla konsumenta każdy atrybut?
- Jak przedstawia się relatywny wkład każdego atrybutu i każdego poziomu w całkowity wybór profilu?
- Jakiego rodzaju zamiany (*trade-offs*) mogą być dokonane pomiędzy atrybutami?

Metody wyborów dyskretnych mają zastosowania w badaniach preferencji konsumentów w celu [Bąk 2004, s. 123]:

- określenia preferencji indywidualnych nabywców,
- oszacowania udziałów w rynku,
- segmentacji rynku nabywców.

W polskojęzycznej literaturze przedmiotu przedstawiano już badania preferencji oraz oszacowania udziałów w rynku. Natomiast brakuje przykładów prezentujących przeprowadzoną segmentację z wykorzystaniem metod wyborów dyskretnych. W tego typu badaniach wykorzystywane są modele klas ukrytych.

## 3. Modele klas ukrytych

W ostatnich latach opublikowano wiele prac związanych z analizą klas ukrytych oraz skończonymi modelami mieszanymi<sup>1</sup>. Modele takie, oprócz zmiennych obserwowalnych, zawiera jedną lub więcej zmiennych ukrytych, nieobserwowalnych, które reprezentują interesujące nas charakterystyki.

Ze względu na różne rozkłady zmiennych obserwowalnych i zmiennych ukrytych, możemy mówić o różnych modelach zmiennych ukrytych [Vermunt, Magdison 2003, s. 1].

---

<sup>1</sup> Przez pojęcie modele mieszane rozumie się modele, które zawierają zmienne o różnych rozkładach.

Tak jak podaje Bartholomew [Bartholomew, Knott 2002, s. 3], wyróżniamy cztery główne rodzaje modeli (zob. tab. 1):

- analiza czynnikowa (Factor Analysis – FA),
- analiza z ukrytymi charakterystykami (Latent Trait Analysis – LTA),
- analiza z ukrytymi profilami (Latent Profile Analysis – LPA),
- analiza z ukrytymi zmiennymi (Latent Class Analysis – LCA).

**Tabela 1.** Klasyfikacja modeli zmiennych ukrytych

Zmienna obserwowalna	Zmienna ukryta	
	ciągła	skokowa
Ciągła	Factor Analysis	Latent Profile Analysis
Skokowa	Latent Trait Analysis	Latent Class Analysis

Źródło: [Vermunt, Magdison 2003, s. 1].

Istnieją trzy główne obszary analizy z wykorzystaniem modeli klas ukrytych. Obejmują one: umieszczanie analizowanych przypadków w segmentach, redukcję zmiennych, konstrukcję skali oraz predykcje zmiennej zależnej [Magdison, Vermunt 2002, s. 2].

Można więc wyróżnić trzy główne rodzaje modeli klas ukrytych [Magdison, Vermunt 2002, s. 2]:

- modele klas ukrytych z wykorzystaniem segmentów (Latent Class Cluster Models),
- modele klas ukrytych z wykorzystaniem czynników (Latent Class Factor Models),
- modele klas ukrytych w regresji i modelach wyboru (Latent Class Regression and Choice Models).

Model klas ukrytych w regresji, znany również jako model segmentacji klas ukrytych, charakteryzuje się tym, że [Magdison, Vermunt 2002, s. 5]:

- jest wykorzystywany do predykcji zależnej zmiennej będącej funkcją predyktorów,
- zawiera zmienną ukrytą o  $R$ -kategoriach, z których każda reprezentuje homogeniczną populację (klasę, segment),
- dla każdego z ukrytych segmentów można wyestymować inny model regresji,
- klasyfikuje cechy w segmenty i symultanicznie szacuje dla każdego z nich modele regresji.

Zaletami tego podejścia są m.in. [Magdison, Vermunt 2002, s. 5-6]:

- osłabienie tradycyjnych założeń, mówiących, że każdy model dla wszystkich cech zakłada  $R = 1$ , co pozwala na oszacowanie osobnego modelu regresji dla każdego segmentu,
- diagnostyczne statystyki umożliwiają określenie wartości dla  $R$ ,
- w przypadku  $R > 1$ , model może zostać rozszerzony o dodatkowe zmienne objaśniające, by przeprowadzona analiza była dokładniejsza, a przyporządkowanie do segmentu było bardziej klarowne.

Typowe zastosowanie w marketingu tradycyjnego odpowiednika modelu klas ukrytych zawiera [Magdison, Vermunt 2002, s. 6]:

- studia, analizę satysfakcji klienta: identyfikacje poszczególnych determinant satysfakcji klienta, które są odpowiednie dla każdego segmentu,
- wspólne studia: identyfikację atrybutów produktów, które należą do różnych segmentów rynku,
- bardziej ogólnie: identyfikację ukrytych segmentów, które mogą wyjaśnić nie-obszerną heterogeniczność wśród danych.

W metodach wyborów dyskretnych modele klas ukrytych uwzględniają heterogeniczność preferencji konsumentów na poziomie segmentowym [Zwerina 1997, s. 75; Huber i in. 1999, s. 6]. W badaniach z wykorzystaniem modeli klas ukrytych zakłada się, że w badanej próbie istnieje skończona liczba grup konsumentów o podobnych preferencjach, natomiast między grupami występują istotne różnice (grupy te nie są znane *a priori*, tylko są „ukryte”, ponieważ nie jest znana ani przynależność poszczególnych konsumentów do określonych segmentów, ani liczba grup) [Bąk 2004, s. 134].

W statystyce wielowymiarowej modele klas ukrytych mieszczą się w grupie modeli rozkładów mieszanych [Domański, Pruska 2000, s. 30-36]. Mieszanki rozkładów są tworzone przez określoną liczbę rozkładów składowych. Udział każdego rozkładu składowego w mieszaninie jest określony przez tzw. parametr mieszający. Suma wartości parametrów mieszających równa się 1.

W badaniach, w których wykorzystujemy modele klas ukrytych w celu segmentacji, parametr mieszający interpretuje się jako rozmiar segmentu, natomiast celem estymacji modelu jest przede wszystkim oszacowanie liczby i wielkości poszczególnych segmentów.

Procedura konstrukcji i estymacji modelu klas ukrytych przedstawia się następująco [Bąk 2004, s. 134-135]:

- określenie rozkładu warunkowego preferencji danego respondenta (warunkiem jest przynależność danego respondenta do określonego segmentu);
- określenie rozkładu bezwarunkowego preferencji danego respondenta (jest to ważona suma rozkładów warunkowych, gdzie wagami są oszacowane prawdopodobieństwa przynależności poszczególnych respondentów do segmentów);
- sformułowanie funkcji największej wiarygodności (jest to iloczyn funkcji indywidualnych rozkładów preferencji przy założeniu, że są one niezależne), której argumentami są preferencje empiryczne oraz nieznanne parametry;
- estymacja modelu (oszacowanie parametrów i wielkości segmentów);
- obliczenie prawdopodobieństw *a posteriori* przynależności respondentów do segmentów.

#### 4. Estymacja parametrów i oprogramowania komputerowe

Algorytm E-M stosowany jest w programach komputerowych przeznaczonych do estymacji mieszanek rozkładów wielowymiarowych częściej niż inne algorytmy optymalizacyjne (np. Newtona-Raphsona), m.in. ze względu na „dobrą” zbieżność i „łatwość” zaprogramowania [Wedel, Kamakura 1998, s. 81].

Podstawową zaletę algorytmu E-M stanowi monotoniczna poprawa wartości funkcji wiarygodności w miarę zwiększania liczby iteracji. Procedura ta jest także na tyle uniwersalna, że może być stosowana do estymacji różnych modeli rozkładów mieszanych. W metodach dekompozycyjnych algorytm E-M może być wykorzystany do estymacji metrycznych modeli klas ukrytych (tradycyjne modele *conjoint analysis*, pomiar preferencji na skalach mocnych) i niemetrycznych modeli klas ukrytych (modele wyborów dyskretnych, pomiar preferencji na skalach słabych).

Ważnym zagadnieniem w procedurze estymacji modeli klas ukrytych jest ustalenie optymalnej liczby segmentów.

Do najpopularniejszych kryteriów wyboru liczby klas należą [Kasprzyk 2009, s. 292-294; Shen i in. 2006, s. 3-4]:

- AIC – Akaike (1974),
- AIC<sub>3</sub> – modyfikacja kryterium AIC przez Bozdogana (1994),
- CAIC – Constant AIC, modyfikacja kryterium informacyjnego zaproponowana przez Bozdogana (1992),
- BIC – bayesowskie kryterium informacyjne zaproponowane przez Schwarz (1978),
- ABIC – skorygowanie próby w kryterium informacyjnym BIC przez Socolow (1987),
- NEC – kryterium klasyfikacyjne, znormalizowane kryterium entropii zaproponowane przez Celeuxa i Soromenho (1996),
- ICL BIC – połączenie kryterium bayesowskiego BIC oraz klasyfikacyjnego przez Biernackiego, Celouxa i Govaerta (2000).

Za pomocą tych kryteriów wybierany jest ten model, dla którego dane kryterium przyjmuje wartość najmniejszą.

Procedury szacujące modele klas ukrytych są dostępne zarówno w uniwersalnych pakietach statystycznych (np. SPSS, SAS/STAT, R), jak i programach przeznaczonych wyłącznie do estymacji modeli klas ukrytych (np. GLIMMIX, The Latent Class Segmentation Module (Sawtooth Software), Latent GOLD®, Latent GOLD@Choice [<http://www.statisticalinnovations.com>]).

#### 5. Przykłady zastosowań

Jeden z pierwszych przykładów wykorzystania modeli klas ukrytych w metodach wyborów dyskretnych przedstawili w swoich pracach Swait [1994] oraz Bhat

[1997]. Inne przykłady zastosowań zawarte są m.in. w: [Shen i in. 2006; DeSarbo i in. 1995; Moore i in. 1996; Green, Hensher 2002; Pacifico 2009].

W literaturze przedmiotu przedstawione są porównania wyników badań wykorzystujących mieszany model logitowy i modele klas ukrytych (np. [Shen i in. 2006; Green, Hensher 2002]). W obu przypadkach wyniki porównania modeli klas ukrytych z mieszanym modelem logitowym (*mixed logit model*) dają podobne rezultaty, tzn. wskazują na lepsze wyniki w badaniach, w których wykorzystuje się modele klas ukrytych (niż mieszanym model logitowy). Jednakże wskazane są dalsze badania porównawcze tych metod.

## Literatura

- Bartholomew D.J., Knott M., *Latent Variable Models and Factor Analysis*, Arnold, London 2002.
- Bąk A., *Dekompozycyjne metody pomiaru preferencji w badaniach marketingowych*, Prace Naukowe Akademii Ekonomicznej nr 1013. Monografie i Opracowania nr 157, Wydawnictwo Akademii Ekonomicznej, Wrocław 2004.
- Bhat C., *An endogenous segmentation mode choice model with an application to intercity travel*, „Transportation Science” 1997, 31(1), s. 34-48.
- Biernacki C., Celeux G., Govaert G., *Assessing a mixture model for clustering with the integrated completed likelihood*, „IEEE Transactions on Pattern Analysis and Machine Intelligence” 2000, 22 (7), s. 719-725.
- Celeux G., Soromenho G., *An entropy criterion for assessing the number of clusters in a mixture model*, „Journal of Classification” 1996, 13, s. 195-212.
- DeSarbo W.S., Ramaswamy V., Cohen S.H., *Market segmentation with choice-based conjoint analysis*, „Marketing Letters” 1995, 6, s. 137-148.
- Domański C., Pruska K., *Nieklasyczne metody statystyczne*, PWE, Warszawa 2000.
- Green W.H., Hensher D.A., *A Latent Class Model for Discrete Choice Analysis: Contrast with Mixed Logit*, Institute of transport Studies, Working Paper, 2002, <http://pages.stern.nyu.edu/~wgreene/latent-class.pdf>.
- Hair J.F., Anderson R. E., Tatham R. L., Blach W.C., *Multivariate Data Analysis with Readings*, Prentice-Hall, Englewood Cliffs 1995.
- Huber J., Orme B.K., Miller R., *Dealing with Product Similarity in Conjoint Simulations*, 1999, [www.sawtoothsoftware.com/download/techpap/prodsim.pdf](http://www.sawtoothsoftware.com/download/techpap/prodsim.pdf).
- Kasprzyk I., *Wykorzystanie konfiguracyjnej analizy częstości w analizie klas ukrytych*, [w:] J. Dziechciarz (red.), *Projektowanie, ocena i wykorzystanie danych rynkowych*, Prace Naukowe Uniwersytetu Ekonomicznego nr 51, Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław 2009, s. 36-45.
- Magdison J., Vermunt J.K., *A Nontechnical Introduction to Latent Class Models*, Statistical Innovations White Paper #1,2002, [www.statisticalinnovations.com/articles/lcmmodels2.pdf](http://www.statisticalinnovations.com/articles/lcmmodels2.pdf).
- Moore W.L., Gray-Lee J., Louviere J., *A Cross-Validity Comparison of Conjoint Analysis and Choice Models at Different Levels of Aggregation*, Working Paper, University of Utah, November 1996.
- Pacifico D., *Modelling Unobserved Heterogeneity in Discrete Choice Models of Labour Supply*, MPRA Paper No. 19030, 2009, [http://mpra.ub.uni-muenchen.de/19030/2/MPRA\\_paper\\_19030.pdf](http://mpra.ub.uni-muenchen.de/19030/2/MPRA_paper_19030.pdf).
- Shen J., Sakata Y., Hashimoto Y., *A Comparison between Latent Class Model and Mixed Logit Model for Transport Mode Choice: Evidences from Two Datasets of Japan*, Discussion paper In

- Economics And Business, Discussion Paper 06-05, January 2006, <http://www2.econ.osaka-u.ac.jp/library/global/dp/0605.pdf>.
- Swait J., *A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data*, „Journal of Retailing and Consumer Services” 1994, 1(2), s. 77-89.
- Vermunt J.K., Magdison J., *Latent Variable*, *Encyclopedia of Social Science Research Methods*, Sage Publications, 2003, [www.statisticalinnovations.com/articles/Latvars.pdf](http://www.statisticalinnovations.com/articles/Latvars.pdf).
- Wedel M., Kamakura W.A., *Market Segmentation. Conceptual and Methodological Foundation*, Kluwer Academic Publishers, Boston – Dordrecht – London 1998.
- Zwerina K., *Discrete Choice Experiments in Marketing*, Physica-Verlag, Heidelberg – New York 1997.

**Źródło internetowe**

<http://www.statisticalinnovations.com>.

**LATENT CLASS MODELS IN DISCRETE CHOICE METHODS**

**Summary:** Revealed preference analysis can be carried out in various ways – discrete choice methods are one of them. Due to the fact that total and part-worth utilities are estimated at aggregated level the segmentation cannot be made. In order to estimate utilities at segment level we apply latent class models. The paper presents latent class models, computer software and known from literature examples of applications.