

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 427

Taksonomia 27

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



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronach internetowych
www.pracenaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

Publikacja udostępniona na licencji Creative Commons
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2016

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp	9
Beata Bal-Domańska: Propozycja procedury oceny zrównoważonego rozwoju w układzie <i>presja – stan – reakcja</i> w ujęciu przestrzennym / Proposal of the assessment of poviats sustainable development in the pressure – state – response system in spatial terms.....	11
Tomasz Bartłomowicz: Pomiar preferencji konsumentów z wykorzystaniem metody <i>Analytic Hierarchy Process</i> / Analytic Hierarchy Process as a method of measurement of consumers’ preferences.....	20
Maciej Beręsewicz, Marcin Szymkowiak: Analiza skupień wybranych lokalnych rynków nieruchomości w Polsce z wykorzystaniem internetowych źródeł danych / Cluster analysis of selected local real estate markets in Poland based on Internet data sources.....	30
Beata Bieszk-Stolorz: Wybrane modele przeciętnego efektu oddziaływania w analizie procesu wychodzenia z bezrobocia / Chosen average treatment effect models in the analysis of unemployment exit process.....	40
Justyna Brzezińska: Modele IRT i modele Rascha w badaniach testowych / IRT and Rasch models in test measurement.....	49
Mariola Chrzanowska, Nina Drejerska: Geograficznie ważona regresja jako narzędzie analizy poziomu rozwoju społeczno-gospodarczego na przykładzie regionów Unii Europejskiej / Geographically weighted regression as a tool of analysis of socio-economic development level of regions in the European Union.....	58
Sabina Denkowska: Zastosowanie analizy wrażliwości do oceny wpływu nieobserwowanej zmiennej w <i>Propensity Score Matching</i> / The application of sensitivity analysis in assessing the impact of an unobserved confounder in Propensity Score Matching.....	66
Adam Depta: Zastosowanie analizy czynnikowej do wyodrębnienia aspektów zdrowia wpływających na jakość życia osób jaskających się / The application of factor analysis to the identification of the health aspects affecting the quality of life of stuttering people.....	76
Mariusz Doszyń, Sebastian Gnat: Taksonomiczno-ekonometryczna procedura wyceny nieruchomości dla różnych miar porządkowania / Taxonomic and econometric method of real estate valuation for various classification measures.....	84

Marta Dziechciarz-Duda, Anna Król: Segmentacja konsumentów smartfonów na podstawie preferencji wyrażonych / Segmentation of smartphones' consumers on the basis of stated preferences	94
Ewa Genge: Zmienne towarzyszące w ukrytym modelu Markowa – analiza oszczędności polskich gospodarstw domowych / Latent Markov model with covariates – Polish households' saving behaviour	103
Joanna Górna, Karolina Górna: Modelowanie wzrostu gospodarczego z wykorzystaniem narzędzi ekonometrii przestrzennej / Economic growth modelling with the application of spatial econometrics tools	112
Alicja Grześkowiak: Wielowymiarowa analiza kompetencji zawodowych według grup wieku ludności / Multivariate analysis of professional competencies with respect to the age groups of the population	122
Agnieszka Kozera, Feliks Wysocki: Problem ustalania współrzędnych obiektów modelowych w metodach porządkowania liniowego obiektów / The problem of determining the coordinates of model objects in object linear ordering methods	131
Mariusz Kubus: Lokalna ocena mocy dyskryminacyjnej zmiennych / Local evaluation of a discrimination power of the variables.....	143
Paweł Lula, Katarzyna Wójcik, Janusz Tuchowski: Analiza wydźwięku polskojęzycznych opinii konsumenckich ukierunkowanych na cechy produktu / Feature-based sentiment analysis of opinions in Polish.....	153
Aleksandra Łuczak, Agnieszka Kozera, Feliks Wysocki: Ocena sytuacji finansowej jednostek samorządu terytorialnego z wykorzystaniem rozmytych metod klasyfikacji i programu R / Assessment of financial condition of local government units with the use of fuzzy classification methods and program R	165
Dorota Rozmus: Badanie stabilności taksonomicznej czynnikowej metody odległości probabilistycznej / Stability of the factor probability distance clustering method	176
Adam Sagan, Aneta Rybicka, Justyna Brzezińska: <i>Conjoint analysis</i> oparta na modelach IRT w zagadnieniu optymalizacji produktów bankowych / An IRT-approach for conjoint analysis for banking products preferences.....	184
Michał Stachura: O szacowaniu centrum populacji określonego obszaru na przykładzie Polski / On estimating centre of population of a given territory. Poland's case	195
Michał Stachura, Barbara Wodecka: Wybrane aspekty i zastosowania modeli zdarzeń ekstremalnych / Selected facets and application of models of extremal events	205
Iwona Staniec, Jan Żółtowski: Wykorzystanie analizy log-liniowej do wyboru czynników determinujących współpracę w przedsiębiorczości	

technologicznej / Use of log-linear analysis for the selection determinants of cooperation in technological entrepreneurship.....	215
Marcin Szymkowiak, Wojciech Roszka: Potencjał gospodarczy gmin aglomeracji poznańskiej w ujęciu taksonomicznym / The economic potential of municipalities of the Poznań agglomeration in the light of taxonomy analysis.....	224
Lucyna Wojcieszka: Zastosowanie modeli klas ukrytych w badaniu opinii respondentów na temat roli państwa w gospodarce / Implementation of latent class models in the respondents' survey on the role of the country in economy.....	234

Wstęp

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania”, zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

Krzysztof Jajuga, Marek Walesiak

Lucyna Wojcieszka

Uniwersytet Ekonomiczny we Wrocławiu
e-mail: lucyna.wojcieszka@ue.wroc.pl

**ZASTOSOWANIE MODELI KLAS UKRYTYCH
W BADANIU OPINII RESPONDENTÓW
NA TEMAT ROLI PAŃSTWA W GOSPODARCE**

**IMPLEMENTATION OF LATENT CLASS MODELS
IN THE RESPONDENTS' SURVEY
ON THE ROLE OF THE COUNTRY IN ECONOMY**

DOI: 10.15611/pn.2016.427.24

Streszczenie: Celem artykułu jest przedstawienie wyników badania opinii respondentów na temat roli państwa w gospodarce na przykładzie Polski poprzez: wskazanie koncepcji dominujących (socjalizmu, interwencjonizmu, liberalizmu, keynesizmu), segmentację respondentów oraz ocenę istotności wpływu cech respondentów na przynależność do poszczególnych klas. Analizą objęto 991 respondentów, zgromadzono 11 892 obserwacji 12 zmiennych obserwowanych mierzonych na skali dychotomicznej. W analizie wykorzystano modele klas ukrytych oraz modele regresji klas ukrytych, a ich estymację przeprowadzono za pomocą pakietu poLCA oraz programu R. Na tej podstawie sformułowano wnioski: w świetle przyjętych kryteriów informacyjnych najlepiej dopasowanym modelem do danych empirycznych jest podział na 4 klasy: 1 – 31%, 2 – 26%, 3 – 6%, 4 – 37%; w klasie 1 i 2 dominuje przekonanie, że państwo powinno pełnić rolę państwa opiekuńczego m.in. poprzez rozbudowę systemu socjalnego (socjalizm) i tworzenie nowych przedsiębiorstw państwowych (keynesizm); respondenci z klasy 3 opowiadają się za gospodarką liberalną; a w klasie 4 dominują respondenci „niezdecydowani”; żadna ze zmiennych towarzyszących nie wywiera istotnego wpływu na przynależność respondentów do poszczególnych klas.

Słowa kluczowe: preferencje wyrażone, rola państwa w gospodarce, modele klas ukrytych program R.

Summary: The aim of the article is to present the respondents' opinion survey on the role of the country in the economy with regard to Poland by: indicating the dominant concepts (socialism, interventionism, liberalism and Keynesianism), respondents' segmentation, characteristics influence on belonging to classes. Gathering of empirical data on the basis of the conducted survey was subcontracted to ResearchOnline company. Subjective scope covered 991 respondents, while the objective one – 11 892 observations of 12 variables of the observed set measured on dichotomous scale. Models of latent class models and models of latent class regression were used. The estimation was conducted by means of poLCA package and the R project. The conclusions were given: in the light of the informational

criteria the most adequate model is a division of the surveyed group into 4 classes according to their contribution: 1–31%, 2–26%, 3–6%, 4–37%; in the class 1, 2 dominates the conviction that the state should play the role of the welfare state by means of increasing the minimum wage and extension of the social system (socialism) as well as creating new national companies (Keynesianism); class 3 respondents support the idea of liberal economy; class 4 is dominated by “undecided” respondents; none of the covariates significantly influences on the respondents’ belonging to particular classes.

Keywords: stated preferences, the role of the government in economy, latent class models, R program.

1. Wstęp

Nad zasadnością i zakresem roli państwa w gospodarce zastanawiali się przedstawiciele wszystkich dotychczasowych nurtów ekonomii, poczynając od A. Smitha po P. Samuelsona – twórcę teorii dóbr publicznych. Współcześnie problemem tym interesują się nie tylko ekonomiści, ale również działacze sfery politycznej i społecznej.

Dawne koncepcje roli państwa w gospodarce największych nurtów w ekonomii klasycznego liberalizmu i keynesowskiego interwencjonizmu nie tylko nie straciły aktualności, ale również doczekały się swoistych, współczesnych modyfikacji w sferze zarówno teoretycznej, jak i praktycznej. Najbardziej obserwowalnym potwierdzeniem tej tezy mogą być realia obecnej sceny politycznej i podział partyjny rządu pod względem orientacji politycznych. Obserwując aktywność czterech największych partii politycznych, z dużym przekonaniem należy stwierdzić, że aktualny spór nie dotyczy zasadności roli państwa w gospodarce, tylko zakresu i płaszczyzn owego wpływu. Z tego też powodu w niniejszym artykule podjęto próbę wskazania aktualnych tendencji w tym zakresie na podstawie opinii respondentów (preferencji wyrażonych) na temat czterech podstawowych koncepcji roli państwa w gospodarce [Bartkowiak 2008]: klasycznego liberalizmu – tzw. koncepcji państwa-minimum; łagodnego interwencjonizmu rozumianego bardziej jako protekcjonizm (w tym walutowy); państwa opiekuńczego (socjalizmu); keynesowskiego interwencjonizmu. W tym kontekście dokonano również segmentacji respondentów.

Zebrań danych empirycznych – opinii respondentów – przeprowadzono w okresie od 1 sierpnia do 10 września 2015 r. za pomocą elektronicznego kwestionariusza ankiety. Gromadzenie danych przeprowadziła firma marketingowej ResearchOnline, którą wybrano ze względu na posiadanie przez nią rozbudowanej bazy danych o populacji. Metodą doboru próby był dobór losowy warstwowy proporcjonalny. Specyfika badań przeprowadzanych na próbie internautów wiąże się zazwyczaj z problemem dotarcia do grupy osób starszych i mniej wykształconych, a w konsekwencji powstaje problem związany z reprezentatywnością badanej po-

populacji. Dlatego też zdecydowano się na wspomnianą metodę doboru próby. Populację reprezentowaną w bazie danych podzielono na warstwy według następujących trzech charakterystyk: płeć (kobieta, mężczyzna), wiek (do 21, 21–40, 41+), wykształcenie (niższe, średnie, wyższe). Z tak powstałych warstw wylosowano następnie niezależne próby. Zgromadzono w ten sposób 991 poprawnie wypełnionych kwestionariuszy, tym samym 11 892 obserwacji 12 zmiennych obserwowanych (12 pytań) mierzonych na skali dychotomicznej (odpowiedź „Tak” lub „Nie”). W analizie danych wykorzystano modele klas ukrytych i regresji klas ukrytych. Estymację parametrów przeprowadzono z wykorzystaniem pakietu *poLCA* [Linzer, Lewis 2015] oraz niekomercyjnego programu R.

2. Mikroekonometryczne modele zmiennych dyskretnych najczęściej stosowane w badaniach empirycznych

Pomiar i analiza danych na poziomie indywidualnym (mikrodanych) możliwe są dzięki budowie i zastosowaniu formalnych modeli mikroekonometrycznych, wyodrębnionych ze względu na typ zmiennych, a tym samym skal i metod pomiaru.

W badaniach marketingowych, w tym badaniach preferencji wyrażonych dotyczących produktu, usług oraz postaw i opinii, do najczęściej występujących typów zmiennych należą zmienne o charakterze jakościowym, czy inaczej dyskretnym. Oznacza to, że pierwotne mikrodane muszą podlegać dalszej transformacji – modelowaniu, aby możliwym stało się wyodrębnienie z nich istotnych dla badacza informacji na temat nie tylko badanego przedmiotu (obiektu), ale również podmiotu (respondenta). Najczęściej stosowane w badaniach empirycznych modele zmiennych dyskretnych dla danych jakościowych przedstawiono w tab. 1.

Tabela 1. Mikroekonometryczne modele zmiennych dyskretnych dla danych jakościowych najczęściej stosowane w badaniach empirycznych

Modele dwumianowe	Modele wielomianowe		Modele klas ukrytych
Liniowy model prawdopodobieństwa	kategorii uporządkowanych	kategorii nieuporządkowanych	model klas ukrytych (bez zmiennych towarzyszących)
Model logitowy	uporządkowany model logitowy i probitowy	wielomianowy model logitowy i probitowy	model regresji klas ukrytych ze zmiennymi objaśniającymi
Model probitowy		warunkowy model logitowy (McFadden)	
Model logarytmiczno-liniowy (tablice kontyngencji)	uogólniony model uporządkowany	zagnieżdżony model logitowy	model regresji klas ukrytych ze zmiennymi objaśniającymi i zmiennymi towarzyszącymi
Model komplementarny (log-log)	model danych sekwencyjnych	mieszany model logitowy	

Źródło: opracowanie własne na podstawie [Gruszczyński 2012, s. 20; Gatnar, Walesiak (red.) 2011, s. 113; Bąk 2013, s. 150].

Na szczególną uwagę zasługują modele klas ukrytych i regresji klas ukrytych, które znajdują zastosowanie w analizie zmiennych zarówno dychotomicznych, jak i politomicznych, które mogą być mierzone na skalach metrycznych i niemetrycznych.

Zastosowanie modeli klas ukrytych w badaniach empirycznych prezentują m.in. prace: [Agresti 2002; Agresti, Lang 1993; Bąk 2013; Cameron, Trivedi 2005; Dayton, Macready 1988; Formann 2003; Garrett, Zeger 2000; Goodman 1970; Gupta, Jain, Sawhney 1999; Hagenaars 1993; Hagenaars, McCutcheon 2002; Linzer, Lewis 2011; Nawojczyk 2002; Reboussin, Ip, Wolfson 2008; Vermunt 2010].

3. Modele klas ukrytych i ich dopasowanie do danych empirycznych

Modele klas ukrytych (bez zmiennych towarzyszących) i modele regresji klas ukrytych (ze zmiennymi objaśniającymi, towarzyszącymi) ze swojej definicji muszą zawierać przynajmniej jedną kategoriałną zmienną ukrytą (klasę, segment). Ponieważ jest ona kategorią sztuczną – abstrakcyjną, nie podlega bezpośredniej obserwacji (mierzalności), a jedynie wynika z informacji zgromadzonych na podstawie zmiennych obserwowanych [Everitt, Dunn 2001, s. 305; Bąk 2013, s. 151]. Zatem proces klasyfikacji (segmentacji, agregacji, grupowania) odbywa się na podstawie estymacji parametrów modelu – zmiennych obserwowanych – prawdopodobieństw przynależności obiektów do danej klasy oraz wartości prawdopodobieństw k -tej odpowiedzi na i -te pytanie pod warunkiem przynależności do c -tej klasy.

Ogólny model klas ukrytych można zapisać w postaci [Bąk 2013, s. 150]:

$$f(y|\phi) = \sum_{c=1}^C \pi_c f_c(y|\theta_c), \quad (1)$$

gdzie: f – funkcja rozkładu obserwacji (preferencji empirycznych); y – preferencje empiryczne; $\phi = (\pi; \Theta)$ – nieznane parametry modelu; π_c – nieznana wielkość c -tego segmentu, tzw. parametr mieszający, reprezentujący przynależność obserwacji do poszczególnych klas ukrytych i spełniający warunki: $0 \leq \pi_c \leq 1$; $\sum_{c=1}^C \pi_c = 1$; f_c – funkcja rozkładu obserwacji w c -tej klasie; θ_c – parametry szacowane dla c -tego segmentu; $c = 1, \dots, C$ – numer segmentu.

Do najczęściej stosowanych kryteriów oceny jakości dopasowania modelu do danych empirycznych należą: kryterium informacyjne Akaikego (AIC) oraz kryterium informacyjne Bayesa Schwarza (BIC). Standardowa procedura obejmuje sprawdzenie wartości kryteriów modelu najmniej złożonego – dwuklasowego, stopniowo zwiększając liczbę klas o jeden, aż do momentu uzyskania najlepszego dopasowania modelu do danych. Model jest tym lepiej dopasowany do danych empirycznych, im mniejsza jest wartość kryteriów informacyjnych.

4. Zastosowanie modeli klas ukrytych w badaniu opinii respondentów na temat roli państwa w gospodarce

Badanie opinii respondentów na temat roli państwa w gospodarce przeprowadzono za pomocą elektronicznego kwestionariusza ankiety w okresie od 1 sierpnia do 10 września 2015 r., a więc niespełna dwa miesiące przed wyborami samorządowymi w Polsce. Gromadzenie danych zlecono firmie o nazwie ResearchOnline. Badana próba liczyła 1030 respondentów, w analizie uwzględniono natomiast 991 poprawnie wypełnionych kwestionariuszy, tym samym 11 892 obserwacji.

Kwestionariusz ankiety składał się z dwóch części. W pierwszej części znalazło się 12 pytań zamkniętych mierzonych na skali dychotomicznej – odpowiedź „tak” lub „nie”, w drugiej – 9 pytań dotyczących metryczki respondentów.

Respondenci zostali poproszeni o zaznaczenie odpowiedzi („tak” lub „nie”) na pytanie: „Czy zgadzasz się z poniższym stwierdzeniem, że państwo powinno...”. W badaniu wykorzystano 12 zmiennych obserwowanych Y_1, \dots, Y_{12} :

1. Y_1 (pyt. 1) Zmierzać do ujednoczenia wysokości podatków dla wszystkich grup społecznych?
2. Y_2 (pyt. 2) Zmierzać do łagodzenia nierówności dochodowych poprzez podwyższenie wysokości płacy minimalnej?
3. Y_3 (pyt. 3) Rozbudować system socjalny w Polsce?
4. Y_4 (pyt. 4) Chronić krajowej produkcji i stymulować jej wzrost?
5. Y_5 (pyt. 5) Ograniczać import produktów rolnych poprzez wprowadzanie ceł, koncesji oraz restrykcji sanitarno-zdrowotnych i środowiskowych?
6. Y_6 (pyt. 6) Sterować rynkowym kursem waluty w kierunku zgodnym z polityką gospodarczą?
7. Y_7 (pyt. 7) Nie ingerować w działanie rynku, ponieważ jego najważniejszym elementem jest konkurencja, a jedynym regulatorem powinna być cena?
8. Y_8 (pyt. 8) Minimalnie ingerować w sferę gospodarczą?
9. Y_9 (pyt. 9) Ograniczyć się do utrzymania porządku, obrony narodowej i dostarczania dóbr publicznych?
10. Y_{10} (pyt. 10) Pełnić aktywną rolę w gospodarce jako czynnik stabilizujący i przyczyniający się do podniesienia ogólnego bogactwa narodu?
11. Y_{11} (pyt. 11) Tworzyć nowe przedsiębiorstwa państwowe?
12. Y_{12} (pyt. 12) Być regulatorem rynku?

Odpowiedź twierdząca na pytania 1–3 wskazywała na wybór koncepcji państwa opiekuńczego; 4–6 – „łagodnego” interwencjonizmu, rozumianego jako protekcjonizm; 7–9 – klasycznego liberalizmu, koncepcji tzw. państwa-minimum; 10–12 – keynesowskiego interwencjonizmu.

W analizie uwzględniono również sześć zmiennych towarzyszących X_1 – X_6 :

- a) X_1 wiek: do 21 lat, 21–30 lat, 31–40 lat, 41–50lat, 51–60 lat, powyżej 60 lat;

- b) X_2 stan cywilny: panna/kawaler, zamężna/zonaty, w związku partnerskim, w separacji, rozwiedziona/rozwiedziony, wdowa/wdowiec;
- c) X_3 miejsce zamieszkania: wieś, miasto do 20 tys. mieszkańców, 20–50 tys., 50–100 tys., 100–200 tys., powyżej 200 tys.;
- d) X_4 wykształcenie: niższe (podstawowe, zawodowe), średnie, wyższe;
- e) X_5 miesięczny dochód netto: do 1000 zł, 1001–2000 zł, 2001–3000 zł, 3001–4000 zł, powyżej 4000 zł;
- f) X_6 główne źródło dochodu: własna działalność gospodarcza, praca na umowę o pracę, praca na umowę o dzieło/zlecenie, praca dorywcza, renta/emerytura, zasiłek/pomoc społeczna, na utrzymaniu rodziny, inne.

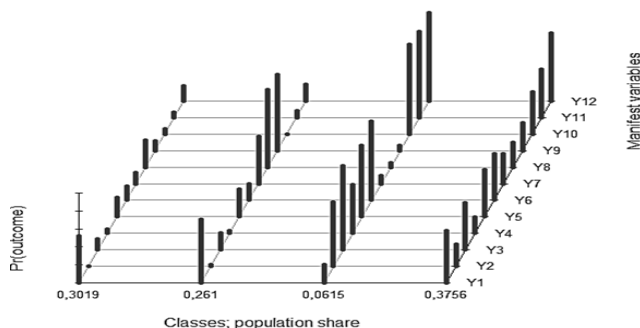
W analizie wykorzystano pakiet *poLCA* [Linzer, Lewis 2015] w programie R. W pierwszym kroku zbudowano model podstawowy (bez zmiennych towarzyszących), a następnie zwiększano liczbę klas o jeden. W tabeli 2 zestawiono otrzymane wartości kryteriów informacyjnych, na podstawie których należy stwierdzić, że najlepiej dopasowanym jest model cztero- i siedmioklasowy.

Tabela 2. Modele klas ukrytych bez zmiennych towarzyszących – wartości kryteriów informacyjnych

Liczba klas	AIC	BIC	G^2	X^2	Max log-lik
2	13239,06	13361,52	2048,462	7134,459	–6594,528
3	13080,88	13267,04	1864,291	6852,825	–6502,442
4	12976,46	13226,3*	1733,869	6826,463	–6437,231
5	12938,04	13251,56	1669,447	4966,239	–6405,02
6	12919,31	13296,51	1624,718	4260,576	–6382,655
7	12907,64*	13348,52	1587,044	4192,827	–6363,819

* Najniższe wartości.

Źródło: opracowanie własne na podstawie funkcji *poLCA.reorder* pakietu *poLCA* programu R.



Rys. 1. Model regresji klas ukrytych ze zmienną towarzyszącą wykształcenie X_4

Źródło: opracowanie własne.

Zazwyczaj w takich przypadkach wybiera się model najmniej złożony w celu uzyskania najbardziej wyraźnych wyników. W kolejnym kroku analizy szacowano już cztero-klasowe modele regresji klas ukrytych z różnymi zestawami zmiennych towarzyszących ($X1-X6$). W efekcie przyjętych kryteriów wybrano czteroklasowy model regresji klas ukrytych ze zmienną towarzyszącą wykształcenie ($X4$), który uznano za najlepiej dopasowany do danych empirycznych (rys. 1).

W ramach przeprowadzonej analizy dokonano również oceny istotności cech respondentów (zmiennych towarzyszących) na przynależność do poszczególnych klas, jak również analizę struktury badanej próby. Ze względu jednak na duże ograniczenie objętościowe artykułu nie zostaną zaprezentowane wszystkie uzyskane wyniki, a także odwzorowania graficzne skonstruowanych modeli.

5. Zakończenie

Przeprowadzone badanie opinii respondentów na temat roli państwa w gospodarce za pomocą analizy klas ukrytych umożliwiło obliczenie wartości parametrów prawdopodobieństwa odpowiedzi na poszczególne pytania, a także prawdopodobieństwa przynależności respondentów do klas, a w efekcie identyfikację wiodących tendencji w tym zakresie, segmentację respondentów oraz ocenę istotności wpływu wybranych cech respondentów na przynależność do klas (tab. 3).

Tabela 3. Model regresji klas ukrytych ze zmienną towarzyszącą wykształcenie $X4$ – wartości parametru prawdopodobieństwa odpowiedzi „tak”

Pytanie (Y)	Klasy/liczebność			
	1/303	2/262	3/55	4/371
Y1	0,4791	0,2889	0,8020	0,4165
Y2	0,9985*	0,9812*	0,2826	0,7557*
Y3	0,8842*	0,8116*	0,0637	0,4773
Y4	0,9588*	0,9751*	0,4605	0,8191*
Y5	0,7878	0,6934	0,2081	0,4709
Y6	0,8451	0,8214	0,1190	0,4848
Y7	0,8743	0,4754	0,9072*	0,6666
Y8	0,6976	0,1344	0,9498*	0,7244*
Y9	0,8948	0,1520	0,9452*	0,6914
Y10	0,9444*	1,0000*	0,0000	0,5315
Y11	0,9435*	0,9263*	0,0416	0,4610
Y12	0,8320*	0,8147*	0,0121	0,2407

* Najwyższe wartości w klasie.

Źródło: opracowanie własne na podstawie pakietu *poLCA* programu R.

Uzyskane wyniki pozwoliły sformułować następujące wnioski końcowe:

- w świetle kryteriów informacyjnych, liczba klas badanej zbiorowości wynosi 4 segmenty według udziałów: 1–31%, 2–26%, 3–6%, 4–37%;

- klasa 1 i 2 (57%) są do siebie najbardziej podobne pod względem opinii na temat roli państwa w gospodarce; dominuje przekonanie, że państwo powinno pełnić aktywną rolę w gospodarce – rolę państwa opiekuńczego poprzez takie działania jak: podwyższanie wysokości płacy minimalnej, rozbudowę systemu socjalnego, ochronę krajowej produkcji i tworzenie nowych przedsiębiorstw państwowych;
- respondenci z klasy 3 (6%) opowiadają się za gospodarką liberalną i uważają, że państwo nie powinno ingerować w sferę gospodarczą;
- respondenci z klasy 4 (37%) to grupa „niezdecydowanych”;
- prawdopodobieństwo przynależności respondentów do klasy 3 i 4 wzrasta wraz ze wzrostem poziomu wykształcenia;
- przy założonym poziomie istotności $\alpha = 0,05$ należy stwierdzić, że żadna ze zmiennych towarzyszących – charakterystyk respondentów nie wywiera istotnego wpływu na przynależność do klas.

Można zaryzykować stwierdzenie, że wyniki przeprowadzonego badania zostały potwierdzone przez rezultat wyborów parlamentarnych (z 25 października 2015), wedle którego wygrała partia polityczna opowiadająca się za rozbudową systemu socjalnego w Polsce.

Literatura

- Agresti A., 2002, *Categorical Data Analysis*, John Wiley & Sons, Hoboken, NJ.
- Agresti A., Lang J.B., 1993, *Quasi-symmetric latent class models with application to rater agreement*, *Biometrics*, vol. 49, no. 1, s. 131–139.
- Bartkowiak R., 2008, *Historia myśli ekonomicznej*, PWE, Warszawa.
- Bąk A., 2013, *Mikroekonometryczne metody badania preferencji konsumentów z wykorzystaniem programu R*, C.H. Beck, Warszawa.
- Cameron A.C., Trivedi P.K., 2005, *Microeconometrics. Methods and Applications*, Cambridge University Press, New York.
- Dayton C.M., Macready G.B., 1988, *Concomitant variable latent class models*, *Journal of the American Statistical Association*, vol. 83, no. 401, s. 173–178.
- Everitt B.S., Dunn G., 2001, *Applied Multivariate Data Analysis*, Arnold, London.
- Formann A.K., 2003, *Latent class model diagnosis from a frequentist point of view*, *Biometrics*, vol. 59, no. 1, s. 189–196.
- Garrett E.S., Zeger S.L., 2000, *Latent class model diagnosis*, *Biometrics*, vol. 56, no. 4, s. 1055–1067.
- Gatnar E., Walesiak M. (red.), 2011, *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*, C.H. Beck, Warszawa.
- Goodman L., 1970, *Multivariate analysis of qualitative data: Interactions among multiple classification*, *Journal of the American Statistical Association*, vol. 65, no. 329, s. 226–256.
- Gruszczyński M., 2012, *Mikroekonometria. Modele i metody analizy danych indywidualnych*, Wolters Kluwer Polska, Warszawa.
- Gupta S., Jain D.C., Sawhney M.S., 1999, *Modeling in the evolution of markets with indirect network externalities: An application to digital television*, *Marketing Science*, vol. 18, no. 3, Special Issue on Managerial Decision Making, s. 396–416.

- Hagenaars A.J., 1993, *Loglinear models with latent variables*, Sage, Newbury Park, CA.
- Hagenaars A.J., McCutcheon A.L., 2002, *Applied Latent Class Analysis*, Cambridge University Press, Cambridge.
- Linzer D., Lewis J., 2011, *poLCA: An R Package for polytomous variable latent class analysis*, Journal of Statistical Software, vol. 10, no. 42, s. 1–29.
- Linzer D., Lewis J., 2015, *Package poLCA – Polytomous Variable Latent Class Analysis*, <https://cran.r-project.org/> (8.09.2015).
- Nawojczyk M., 2002, *Przewodnik po statystyce dla socjologów*, SPSS, Kraków.
- Reboussin B.A., Ip E.H., Wolfson M., 2008, *Locally dependent latent class models with covariates: An application to under-age drinking in the USA*, Journal of the Royal Statistical Society. Series A Statistics in Society, vol. 171, no. 4, s. 877–897.
- Vermunt J.K., 2010, *Latent class modeling with covariates: Two improved three-step approaches*, Political Analysis, no. 18, no. 4, s. 450–469.