

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 328

Taksonomia 23

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

Redaktorzy naukowci

Krzysztof Jajuga, Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

www.ibuk.pl, www.ebscohost.com,

w Dolnośląskiej Bibliotece Cyfrowej www.dbc.wroc.pl,

The Central and Eastern European Online Library www.ceeol.com,

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa

www.wydawnictwo.ue.wroc.pl

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

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2014

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

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Wstęp	11
Małgorzata Rószkiewicz , Wykorzystanie metaanalizy w budowaniu modelu pomiarowego w przypadku braku niezmienniczości zasad pomiaru na przykładzie pomiaru zadowolenia z życia.....	13
Elżbieta Sobczak , Harmonijność inteligentnego rozwoju regionów Unii Europejskiej	21
Ewa Roszkowska, Renata Karwowska , Analiza porównawcza województw Polski ze względu na poziom zrównoważonego rozwoju w roku 2010.....	30
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel , Analiza porównawcza wybranych filtrów w analizie synchronizacji cyklu koniunkturalnego.....	41
Marcin Salamaga , Próba konstrukcji tablic „wymierania scenicznego” spektakli operowych na przykładzie Metropolitan Opera.....	51
Iwona Foryś , Wykorzystanie analizy dyskryminacyjnej do typowania rynków podobnych w procesie wyceny nieruchomości niemieszkalnych	59
Jerzy Korzeniewski , Selekcja zmiennych w klasyfikacji – propozycja algorytmu	69
Sabina Denkowska , Testowanie wielokrotne przy weryfikacji wieloczynnikowych modeli proporcjonalnego hazardu Coxa.....	76
Ewa Chodakowska , Teoria równań strukturalnych w klasyfikacji zmiennych jawnych i ukrytych według charakteru ich wzajemnych oddziaływań	85
Iwona Konarzewska , Model PCA dla rynku akcji – studium przypadku	94
Katarzyna Wójcik, Janusz Tuchowski , Dobór optymalnego zestawu słów istotnych w opiniach konsumentów na potrzeby ich automatycznej analizy	106
Aleksandra Łuczak , Zastosowanie metody AHP-LP do oceny ważności determinant rozwoju społeczno-gospodarczego w jednostkach administracyjnych	116
Aleksandra Witkowska, Marek Witkowski , Klasyfikacja pozycyjna banków spółdzielczych według stanu ich kondycji finansowej w ujęciu dynamicznym	126
Adam Depta , Zastosowanie analizy korespondencji do oceny jakości życia ludności na podstawie kwestionariusza SF-36v2	135
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej , Indukcja reguł dla danych niekompletnych i niezbalansowanych: modele klasyfikatorów i próba ich zastosowania do predykcji ryzyka operacyjnego w torakochirurgii	146

Małgorzata Misztal , Wybrane metody oceny jakości klasyfikatorów – przegląd i przykłady zastosowań.....	156
Anna M. Olszewska , Wykorzystanie wybranych metod taksonomicznych do oceny potencjału innowacyjnego województw	167
Iwona Bąk , Porównanie jakości grupowań powiatów województwa zachodniopomorskiego pod względem atrakcyjności turystycznej.....	177
Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn , Segmentacja gospodarstw domowych według wydatków na turystykę zorganizowaną.....	186
Agnieszka Wałęga , Podejście syntetyczne w analizie spójności ekonomicznej gospodarstw domowych.....	196
Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek , Zastosowanie analizy korespondencji do badania wpływu elektrowni wiatrowych na jakość życia ludności	205
Joanna Banaś, Krzysztof Małecki , Klasyfikacja punktów pomiarów ankietowych kierowców na granicy Szczecina z wykorzystaniem zmiennych symbolicznych.....	214
Aneta Becker , Wykorzystanie informacji granularnej w analizie wymagań rynku pracy.....	222
Katarzyna Cheba, Joanna Holub-Iwan , Wykorzystanie analizy korespondencji w segmentacji rynku usług medycznych.....	230
Adam Depta, Iwona Staniec , Identyfikacja czynników decydujących o jakości życia studentów łódzkich uczelni.....	238
Katarzyna Dębowska, Jarosław Kilon , Reguły asocjacyjne w analizie wyników badań metodą Delphi.....	247
Anna Domagała , O wykorzystaniu analizy głównych składowych w metodzie <i>Data Envelopment Analysis</i>	254
Alicja Grześkowiak , Analiza wykluczenia cyfrowego w Polsce w ujęciu indywidualnym i regionalnym.....	264
Anna M. Olszewska, Anna Gryko-Nikitin , Pomiar postrzegania jakości kształcenia uczelni wyższej na danych porządkowych z wykorzystaniem środowiska R.....	273
Karolina Paradysz , Hierarchiczna metoda grupowania powiatów jako podejście benchmarkowe w ocenie bezrobocia według BAEL-u w wybranych typach małych obszarów	282
Radosław Pietrzyk , Porównanie metod pomiaru efektywności zarządzania portfelami funduszy inwestycyjnych.....	290
Agnieszka Przedborska, Małgorzata Misztal , Wybrane metody statystyki wielowymiarowej w ocenie skuteczności terapeutycznej głębokiej stymulacji elektromagnetycznej u pacjentów z chorobą zwyrodnieniową stawów.....	299

Wojciech Roszka, Marcin Szymkowiak , Podejście kalibracyjne w statystycznej integracji danych	308
Iwona Skrodzka , Zastosowanie wybranych metod klasyfikacji do analizy kapitału ludzkiego krajów Unii Europejskiej	316
Agnieszka Stanimir , Wielowymiarowa analiza czynników sprzyjających włączeniu społecznemu	326
Dorota Strózik, Tomasz Strózik , Przestrzenne zróżnicowanie poziomu życia w województwie wielkopolskim.....	334
Izabela Szamrej-Baran , Identyfikacja przyczyn ubóstwa energetycznego w Polsce przy wykorzystaniu modelowania miękkiego.....	343
Janusz Tuchowski, Katarzyna Wójcik , Klasyfikacja obiektów w systemie Krajowych Ram Kwalifikacji opisanych za pomocą ontologii	353
Aleksandra Matuszewska-Janica , Grupowanie krajów Unii Europejskiej ze względu na poziom feminizacji sektorów gospodarczych	361
Monika Rozkrut, Dominik Rozkrut , Identyfikacja strategii innowacyjnych przedsiębiorstw usługowych w Polsce	369

Summaries

Małgorzata Rószkiewicz , The use of meta-analysis in building the measurement model in case of the absence of measurement invariance on the example of measuring of life satisfaction.....	20
Elżbieta Sobczak , Harmonious smart growth of European Union regions.....	29
Ewa Roszkowska, Renata Karwowska , The comparative analysis of Polish voivodeships with respect to sustainable development in 2010.....	40
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel , Comparative analysis of chosen filters in business cycles analysis	50
Marcin Salamaga , The attempt of construction of the life tables for opera works on the example of the Metropolitan Opera	58
Iwona Foryś , Using discriminant analysis to select similar markets in non-residential property valuation process.....	68
Jerzy Korzeniewski , Variable selection in classification – algorithm proposal	75
Sabina Denkowska , Multiple testing in the verification process of multifactorial Cox proportional hazards models	84
Ewa Chodakowska , The theory of structural equations modelling in the classification of observed variables and latent constructs according to the character of their relationship.....	93
Iwona Konarzewska , Modelling stock market by PCA factor model – case study	105

Katarzyna Wójcik, Janusz Tuchowski , Selection of the optimal set of relevant words in consumers opinions in the context of the opinion mining ..	115
Aleksandra Łuczak , Application of AHP-LP to the evaluation of importance of determinants of socio-economic development in the administrative units	125
Aleksandra Witkowska, Marek Witkowski , A dynamic approach to the ranking of cooperative banks by their financial condition	134
Adam Depta , Application of correspondence analysis for the measurement of quality of life – questionnaire SF-36v2 based research	145
Marek Lubicz, Maciej Zięba, Konrad Pawelczyk, Adam Rzechonek, Marek Marciniak, Jerzy Kołodziej , Classification rules extraction for missing and imbalance data: models of classifiers and initial results in the rules-based thoracic surgery risk prediction.....	155
Małgorzata Misztal , Selected methods for assessing the performance of classifiers – an overview and examples of applications.....	166
Anna M. Olszewska , The application of selected quantitative methods to the evaluation of voivodeship innovation level potential.....	176
Iwona Bąk , The comparison of the quality of groupings of poviats of West Pomeranian Voivodeship in terms of tourism attractiveness	185
Agnieszka Kozera, Joanna Stanisławska, Romana Głowicka-Wołoszyn , Household segmentation with respect to the expenditure on organized tourism.....	195
Agnieszka Wałęga , Synthetic approach in the analysis of economic coherence of households	204
Joanna Banaś, Małgorzata Machowska-Szewczyk, Bożena Mroczek , Using the correspondence analysis to examine the impact of wind turbines on the quality of life.....	213
Joanna Banaś, Krzysztof Małecki , Classification of measurement survey points of drivers on the boundary of Szczecin using symbolic variables...	221
Aneta Becker , The use granular information in the analysis of the requirements of the labor market.....	229
Katarzyna Cheba, Joanna Hołub-Iwan , The application of the correspondence analysis of patients segmentation on the medical service market	237
Adam Depta, Iwona Staniec , Identification of the factors that determine the quality of students life at universities in Lodz.....	246
Katarzyna Dębkowska, Jarosław Kilon , Association rules in the analysis of research results the Delphi method	253
Anna Domagała , About using Principal Component Analysis in Data Envelopment Analysis	263
Alicja Grześkowiak , Analysis of the digital divide in Poland at the individual and regional level	272

Anna M. Olszewska, Anna Gryko-Nikitin , Assessment of perception of quality of teaching at an institution of higher learning based on the ordinal data with the utilization of R environment.....	281
Karolina Paradysz , The hierarchical method of grouping poviats as a benchmark approach in the assessment of unemployment by BAEL in selected types of small areas	289
Radosław Pietrzyk , Comparison of methods of measuring the performance of investment funds portfolios.....	298
Agnieszka Przedborska, Małgorzata Misztal , Selected multivariate statistical analysis methods in the evaluation of efficacy of deep electromagnetic stimulation in patients with degenerative joint disease	307
Wojciech Roszka, Marcin Szymkowiak , A calibration approach in statistical data integration	315
Iwona Skrodzka , Application of some methods of classification to the analysis of human capital in the European Union.....	325
Agnieszka Stanimir , Multivariate analysis of social inclusion factors.....	333
Dorota Strózik, Tomasz Strózik , Spatial differentiation of the standard of living in Great Poland Voivodeship	342
Izabela Szamrej-Baran , Identification of fuel poverty causes in Poland using soft modelling	352
Janusz Tuchowski, Katarzyna Wójcik , Classification of objects in the National Classification Framework described by the ontology.....	360
Aleksandra Matuszewska-Janica , Clustering of European Union states taking into consideration the levels of feminization of economic sectors..	368
Monika Rozkrut, Dominik Rozkrut , Identification of service sector innovation strategies in Poland.....	379

Alicja Grześkowiak

Uniwersytet Ekonomiczny we Wrocławiu

ANALIZA WYKLUCZENIA CYFROWEGO W POLSCE W UJĘCIU INDYWIDUALNYM I REGIONALNYM¹

Streszczenie: W pracy podjęto tematykę pierwotnego i wtórnego wykluczenia cyfrowego w Polsce. Za pomocą map natężeń przedstawiono zróżnicowanie podziału cyfrowego w ujęciu regionalnym w latach 2008-2012. Uwarunkowania wtórnego wykluczenia cyfrowego na poziomie indywidualnym, takie jak płeć, wiek, wykształcenie oraz miejsce zamieszkania, przeanalizowano, stosując model regresji logistycznej z porządkową zmienną zależną.

Słowa kluczowe: wykluczenie cyfrowe, porządkowa regresja logistyczna, wizualizacja macierzy.

1. Wstęp

Zgodnie z definicją stosowaną przez OECD podział cyfrowy (wykluczenie cyfrowe) to różnice w dostępie do nowoczesnych technologii oraz w ich wykorzystaniu w różnych celach pomiędzy jednostkami, gospodarstwami domowymi, przedsiębiorcami i obszarami geograficznymi [OECD 2001, s. 5]. Zasadnicze konsekwencje wykluczenia cyfrowego to pogłębianie podziałów społecznych i przyczynianie się do wykluczenia społecznego przede wszystkim wskutek zmniejszenia szans na rynku pracy, ale także poprzez ograniczanie kontaktów i więzi społecznych [Batorski 2009, s. 229-231]. Niski stopień wykorzystania technologii ICT stanowi barierę rozwoju konkurencyjności regionów. Tymczasem według danych Eurostatu w Polsce odsetek osób, które nigdy nie korzystały z Internetu, wyniósł w 2013 r. aż 32% (wobec średniej unijnej 21%). Jak wskazuje Wojniak, można wyróżnić dwa poziomy wykluczenia cyfrowego – pierwszy (pierwotny) uwarunkowany dostępem do nowoczesnych technologii oraz drugi (wtórny) związany z faktycznym wykorzystaniem narzędzi informatycznych [Wojniak 2013, s. 2]. Wiele analiz poświęcono pierwszemu poziomowi wykluczenia cyfrowego, ale wydaje się, że wraz z upowszechnianiem dostępu do komputerów i Internetu będzie wzrastało znaczenie poziomu drugiego. Problematyka zróżnicowania kompetencji

¹ Praca naukowa sfinansowana ze środków Narodowego Centrum Nauki w ramach projektu badawczego 2012/05/B/HS4/02499.

informatycznych Polaków była poruszana m.in. w pracach Batorskiego [2009, 2011] i Czerniawskiej [2012].

W prezentowanej pracy zostały zrealizowane dwa cele analityczne odnoszące się do pierwotnego i wtórnego wykluczenia cyfrowego:

- przedstawienie za pomocą techniki wizualizacji typu *Heatmap* stopnia i dynamiki wykluczenia cyfrowego pierwszego poziomu w ujęciu regionalnym,
- zbadanie uwarunkowań wykluczenia cyfrowego drugiego poziomu na podstawie danych indywidualnych w dwóch obszarach (obsługi komputera oraz wykorzystania Internetu) za pomocą regresji logistycznej z porządkową zmienną zależną.

2. Źródła danych i metody badawcze

W analizie wykorzystano dwa źródła informacji. Dane dotyczące regionów z lat 2008-2012 pochodzą z cyklu badań prowadzonych przez GUS pn. „Wykorzystanie technologii informacyjno-(tele)komunikacyjnych w przedsiębiorstwach i gospodarstwach domowych”. Dane indywidualne zostały zaczerpnięte z Diagnozy Społecznej² z 2011 r.

Zastosowane metody badawcze różnią się w zależności od poziomu analizy. W celu zobrazowania regionalnego zróżnicowania zjawiska wykluczenia cyfrowego na poziomie dostępu oraz użytkowania narzędzi ICT zastosowano jedną z technik wizualizacji macierzy – mapy natężeń (mapy ciepła – *heatmaps*). Tego rodzaju wizualizacja reprezentuje wartości zmiennej ujęte w dwuwymiarowej macierzy za pomocą zróżnicowanej kolorystyki [Wilkinson 2005, s. 143]. Wydaje się, że wobec rosnącej ilości informacji warto wykorzystywać tę metodę w naukach społecznych, ponieważ umożliwia syntetyczne przedstawienie zachodzących prawidłowości i może zastępować w celach interpretacyjnych słabo czytelne obszerne tabele danych.

W celu identyfikacji czynników warunkujących wtórne wykluczenie cyfrowe zastosowano model regresji logistycznej z porządkową zmienną zależną (*ordinal logistic regression*). Taki charakter zmiennej wynikał ze stopniowania zakresu umiejętności i kompetencji informatycznych w teorii wykluczenia cyfrowego poziomu drugiego.

Porządkowy charakter zmiennej zależnej Y można odzwierciedlić w konstrukcji modelu poprzez wyrażenie szansy na przyjęcie przez nią co najwyżej określonej wartości g w odniesieniu do zdarzenia przeciwnego [zob. Kleinbaum, Klein 2010, s. 469; Bartholomew i in. 2008, s. 246]), co po zastosowaniu przekształcenia logitowego można zapisać jako³:

² Rada Monitoringu Społecznego. Diagnoza społeczna: zintegrowana baza danych. www.diagnoza.com (27.07.2013).

³ Możliwe są również alternatywne formy definiowania modelu [Kleinbaum i Klein 2010, s. 466-469], jednakże ze względu na zastosowanie programu SPSS do estymacji parametrów przedstawiono sformułowanie zgodne z procedurami zawartymi w tym pakiecie statystycznym.

$$\ln \frac{P(Y \leq g)}{P(Y > g)} = \alpha_g - \sum_{i=1}^k \beta_i X_i, \quad (1)$$

gdzie: X_1, \dots, X_k – zmienne niezależne,

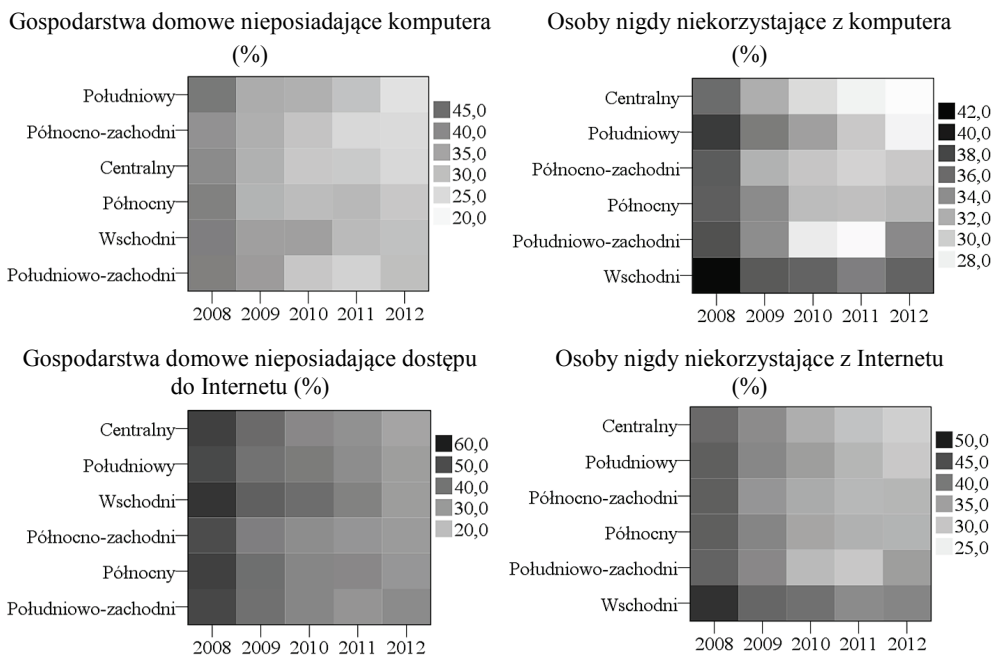
β_1, \dots, β_k – parametry przy zmiennych niezależnych,

α_g – wyraz wolny odpowiadający danej kategorii zmiennej zależnej.

Modelowanie zmiennej Y przyjmującej m wartości wymagałoby oszacowania $m - 1$ równań, ale w przedstawianej procedurze zakłada się, że zmienne niezależne w podobny sposób wpływają na kształtowanie się zmiennej zależnej w odniesieniu do kolejnych jej kategorii, stąd wektor parametrów β jest jednakowy dla wszystkich równań, natomiast różne są oszacowania wyrazów wolnych [Kopczewska i in. 2009, s. 534].

3. Wykluczenie cyfrowe w ujęciu regionalnym

Dane dostępne w ramach statystyki publicznej umożliwiają ocenę zjawiska wykluczenia cyfrowego poziomu pierwszego w przekroju regionów (NUTS-1). Na rys. 1 zaprezentowano za pomocą map natężeń kształtowanie się czterech wskaźników wykluczenia cyfrowego w latach 2008-2012: odsetka gospodarstw domowych



Rys. 1. Wskaźniki wykluczenia cyfrowego w ujęciu regionów NUTS-1 w latach 2008-2012

Źródło: opracowanie własne na podstawie danych GUS.

nieposiadających komputera oraz nieposiadających dostępu do Internetu, odsetka osób w wieku 16-74 lata nigdy niekorzystających z komputera oraz nigdy niekorzystających z Internetu.

Graficzna prezentacja stopnia wykluczenia cyfrowego wyraźnie wskazuje na ograniczanie skali tego zjawiska w latach 2008-2012. Warto zauważyć, że w 2008 r. ponad połowa gospodarstw nie dysponowała dostępem do Internetu, a odsetek respondentów deklarujących, że nigdy nie korzysta z zasobów globalnej sieci, przekraczał 40% we wszystkich regionach. Cztery lata później w najlepszym pod tym względem regionie centralnym brak dostępu wystąpił w 26,6% gospodarstw, a nieużytkowanie Internetu wśród 28,8% ankietowanych osób.

Układ odcieni kolorów w macierzach pozwala stwierdzić, że wykluczenie cyfrowe poziomu pierwszego wykazuje zróżnicowanie terytorialne. W zakresie korzystania z komputera oraz Internetu wyraźnie zarysowuje się niekorzystna sytuacja w regionie wschodnim, a pozytywna w rejonach centralnym i południowym.

Ograniczanie stopnia i zasięgu wykluczenia cyfrowego poziomu pierwszego jest bardzo pozytywnym zjawiskiem. Należy jednak podkreślić, że przebiega ono nierównomiernie, co może przekładać się na zróżnicowanie konkurencyjności regionów.

4. Wykluczenie cyfrowe – ocena na podstawie danych indywidualnych

Jak wskazano, w Polsce następuje spadek pierwotnego podziału cyfrowego związanego z dostępem do nowoczesnych technologii. Na znaczeniu będzie więc zyskiwało wykluczenie cyfrowe drugiego poziomu związane z umiejętnościami i kompetencjami poszczególnych użytkowników. Analizę uwarunkowań obecnie istniejącego wtórnego podziału cyfrowego przeprowadzono na podstawie danych z Diagnozy Społecznej, w której badano kompetencje informatyczne w zakresie [Rada Monitoringu Społecznego 2011, s. 36-37]:

- korzystania z komputera (sześć umiejętności: kopiowanie lub przenoszenie pliku albo folderu; używanie polecenia kopiowania, wycinania i wklejania w celu powielenia lub przemieszczenia wybranych fragmentów dokumentu; wykorzystanie podstawowych funkcji matematycznych w arkuszu kalkulacyjnym; tworzenie elektronicznej prezentacji; instalowanie nowych urządzeń; pisanie programu komputerowego z użyciem języka programowania),
- użytkowania Internetu: (dwadzieścia dwie umiejętności: korzystanie z portali społecznościowych; obsługa rachunku bankowego; sprawdzanie i wysyłanie poczty elektronicznej; korzystanie z komunikatorów; uczestniczenie w czatach; uczestniczenie w grupach lub forach dyskusyjnych; telefonowanie; tworzenie lub modyfikowanie strony WWW; zbieranie materiałów potrzebnych do nauki lub pracy; uczestniczenie w kursach lub szkoleniach; szukanie pracy, wysyła-

nie ofert dotyczących zatrudnienia; kupowanie produktów i usług; uczestniczenie w aukcjach; granie w gry sieciowe; ściąganie darmowego oprogramowania; ściąganie darmowej muzyki, filmów; tworzenie i publikowanie własnych tekstów, np. bloga, grafiki, muzyki lub innej twórczości; uzyskiwanie informacji ze stron instytucji publicznych; pobieranie lub wypełnianie formularzy urzędowych; słuchanie muzyki lub radia; oglądanie telewizji, plików wideo; czytanie gazet lub książek).

Analiza obejmuje wyłącznie respondentów, którzy deklarowali, że są użytkownikami komputera lub Internetu. Wymienione powyżej dwie grupy umiejętności informatycznych rozpatrzono osobno, oznaczając odpowiadające im modele jako *Model I* oraz *Model II*. Kierując się liczbą deklarowanych przez respondentów umiejętności, utworzono dwie porządkowe zmienne zależne – oddzielnie dla umiejętności korzystania z komputera i zastosowań Internetu (tab.1).

W przypadku użytkowania Internetu nie wyróżniono kategorii oznaczonej 0, ponieważ odsetek osób, które deklarowały korzystanie z globalnej sieci, a nie wskazały żadnego z wymienionych zastosowań, był bardzo niski. Za osoby wykluczone cyfrowo w sensie wtórnym uznano respondentów korzystających z komputera, ale niewykonujących żadnej z podanych czynności. Kategoria nr 1 zawiera osoby wykluczone cyfrowo w umiarkowanym stopniu, czyli wykorzystujące narzędzia ICT w bardzo ograniczonym zakresie [por. *Jakość życia* 2013, s. 82; Czerniawska 2012, s. 16-17]. Podziału cyfrowego dopełniają kategorie 2 i 3, które w tym ujęciu można określić jako odpowiadające osobom o średnio zaawansowanych i zaawansowanych umiejętnościach.

Tabela 1. Utworzone kategorie zmiennych zależnych

Nr kategorii	Korzystanie z komputera (Model I)	Użytkowanie Internetu (Model II)
0	żadna z wymienionych czynności	-
1	1-2 czynności	do 3 czynności
2	3-4 czynności	4-11 czynności
3	5-6 czynności	12-22 czynności

Źródło: opracowanie własne.

Jako zmienne niezależne przyjęto następujące charakterystyki: wiek, płeć, miejsce zamieszkania (miasto, wieś) oraz poziom wykształcenia (podstawowe i niższe, zasadnicze zawodowe lub gimnazjum, średnie, wyższe i policealne).

Ocenę zaproponowanych modeli przeprowadzono na podstawie testu ilorazu wiarygodności porównującego model zawierający tylko wyraz wolny z modelem pełnym zawierającym wszystkie proponowane zmienne niezależne. W obu przypadkach wysokie wartości statystyki χ^2 prowadzą do odrzucenia hipotezy zerowej (przy *p-value* 0,000) stwierdzającej, że wszystkie parametry oprócz wyrazu wolne-

go, są równe zero, na rzecz zaproponowanych modeli. Rezultaty estymacji parametrów obu modeli wraz z oceną ich istotności za pomocą testu Walda zawarto w tabeli 2.

Tabela 2. Wyniki estymacji parametrów modeli regresji logistycznej, statystyka testu Walda oraz wartości *p-value*

Kategorie	Model I			Model II		
	oszacowanie parametru	statystyka Walda	<i>p-value</i>	oszacowanie parametru	statystyka Walda	<i>p-value</i>
Mężczyzna	0,576	295,875	0,000	0,245	55,007	0,000
Kobieta	Ref.	–	–	Ref.	–	–
Wieś	–0,285	66,314	0,000	–0,366	110,203	0,000
Miasto	Ref.	–	–	Ref.	–	–
Podstawowe i niższe	–1,765	364,977	0,000	–1,516	315,076	0,000
Zas. zawodowe/gimn.	–1,512	1048,545	0,000	–1,214	708,845	0,000
Średnie	–0,822	416,221	0,000	–0,580	204,359	0,000
Wyższe i policealne	Ref.	–	–	Ref.	–	–
Wiek	–0,058	2081,153	0,000	–0,059	2254,173	0,000
Stała – kategoria 0	–4,146	3511,090	0,000	–4,126	3553,882	0,000
Stała – kategoria 1	–2,514	1588,255	0,000	–1,858	952,712	0,000
Stała – kategoria 2	–0,595	94,425	0,000	–	–	–

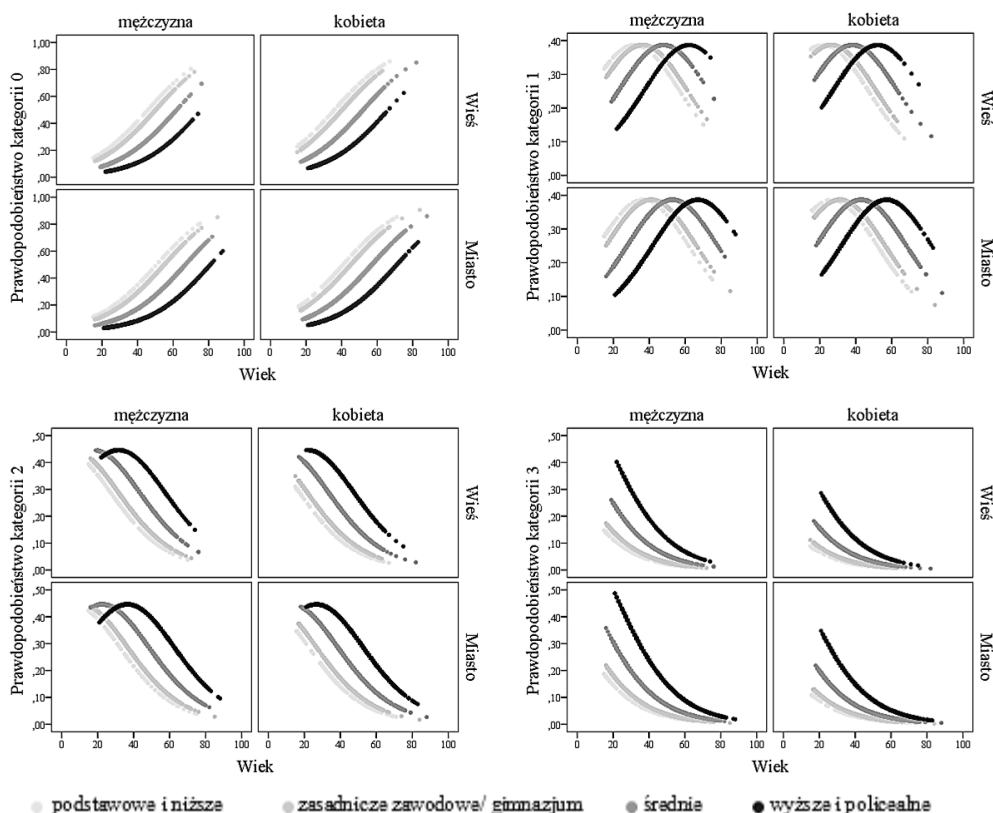
Źródło: obliczenia własne na podstawie danych z Diagnozy Społecznej.

Wszystkie parametry modeli są statystycznie istotne, co oznacza, że zaproponowane zmienne niezależne mają wpływ na stopień wtórnego wykluczenia cyfrowego. Wyższe prawdopodobieństwo tego zjawiska (zarówno pod względem wykorzystania komputera, jak i użytkowania Internetu) występuje w przypadku mieszkańców wsi, kobiet, osób o niższym poziomie wykształcenia i starszych wiekiem. Warto zauważyć, że najwyższe wartości współczynników dla obu modeli otrzymano przy zmiennych wyrażających poziom wykształcenia, co wskazuje na bardzo silną zależność wykluczenia cyfrowego od procesów edukacyjnych. Współczynniki przy zmiennej reprezentującej miejsce zamieszkania świadczą o większym prawdopodobieństwie niekorzystania z Internetu niż z komputera na obszarach wiejskich.

Otrzymane modele umożliwiają obliczenie prawdopodobieństw przyjęcia przez zmienną poszczególnych kategorii przy określonych poziomach zmiennych niezależnych. W celach interpretacyjnych użyteczne jest graficzne przedstawienie zależności prawdopodobieństwa od wartości rozpatrywanych cech.

Na rys. 2 zobrazowano kształtowanie się prawdopodobieństwa przynależności do poszczególnych kategorii użytkowników komputera.

Na szczególną uwagę w kontekście wykluczenia cyfrowego poziomu drugiego zasługują kategorie oznaczone "0" oraz "1". Szanse na zaistnienie wykluczenia cyfrowego wtórnego (kategoria "0") wzrastają bardzo mocno wraz z wiekiem re-

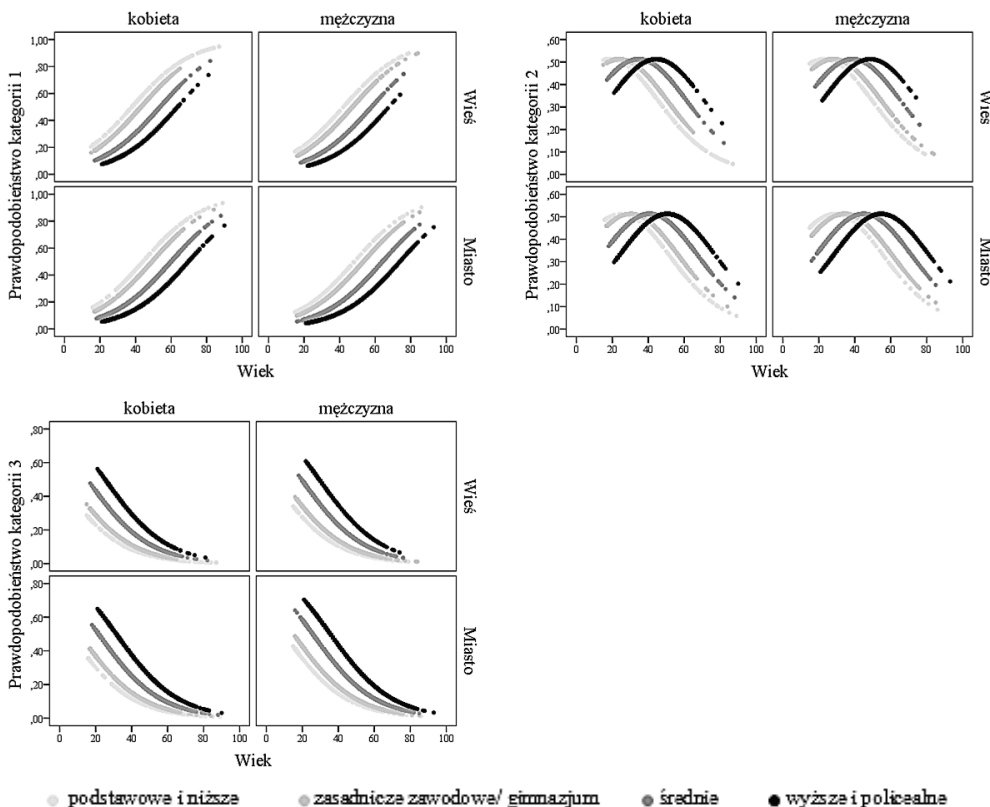


Rys. 2. Oszacowane prawdopodobieństwa przynależności do poszczególnych kategorii użytkowników komputera w zależności od wieku, płci, miejsca zamieszkania i poziomu wykształcenia

Źródło: opracowanie własne na podstawie danych z Diagnozy Społecznej.

spondentów. Warto zauważyć, że punkty charakteryzujące osoby o niskim poziomie wykształcenia (podstawowe i niższe, zasadnicze zawodowe lub gimnazjum) układają się w bezpośredniej bliskości, natomiast punkty odnoszące się do osób z wykształceniem średnim, wyższym i policealnym są od nich bardziej oddalone. Wskazuje to na zdecydowanie większe prawdopodobieństwo wtórnego wykluczenia cyfrowego osób słabo wykształconych. Szanse na umiarkowany stopień wykluczenia (kategoria "1") również zwiększają się z wiekiem, przy czym ich przebieg przypomina fragment paraboli – po osiągnięciu określonego wieku prawdopodobieństwo umiejętności wykonania jednej lub dwóch czynności z zakresu obsługi komputera maleje na rzecz całkowitego wykluczenia wtórnego.

Kształtowanie się prawdopodobieństwa wystąpienia wykluczenia cyfrowego w odniesieniu do użytkownika Internetu zaprezentowano na rys. 3. Umiarkowane wykluczenie cyfrowe poziomu drugiego jest reprezentowane przez kategorię "1" (deklarowanie korzystania z co najwyżej trzech funkcjonalności internetowych).



Rys. 3. Oszacowane prawdopodobieństwa przynależności do poszczególnych kategorii użytkowników Internetu w zależności od wieku, płci, miejsca zamieszkania i poziomu wykształcenia

Źródło: opracowanie własne na podstawie danych z Diagnozy Społecznej.

Szanse na wystąpienie umiarkowanego wykluczenia cyfrowego poziomu drugiego w zakresie korzystania z Internetu zwiększają się znacząco wraz z wiekiem w przypadku osób z wykształceniem podstawowym lub niższym. Widoczne są powiązania z miejscem zamieszkania i płcią.

5. Podsumowanie

W Polsce występują obie formy wykluczenia cyfrowego. Pozytywnym zjawiskiem jest spadek stopnia wykluczenia cyfrowego pierwszego poziomu związanego z dostępem do nowoczesnych technologii. Jednocześnie należy zwrócić uwagę na różnice terytorialne w zakresie dostępu i deklarowanego wykorzystania komputerów i Internetu. Analizy na podstawie danych z Diagnozy Społecznej wskazują na istnienie zjawiska wtórnego podziału cyfrowego, na który wpływ mają takie czyn-

niki, jak wiek, płeć, miejsce zamieszkania oraz poziom wykształcenia. Wydaje się, że przyczyny tego stanu rzeczy mogą mieć zróżnicowany charakter. Przewaga ośrodków miejskich nad terenami wiejskimi wynika zapewne z uwarunkowań infrastrukturalnych. Źródłem zależności stopnia wykluczenia od wieku mogą być: brak warunków oraz niechęć osób starszych do uczenia się nowych rzeczy. Z kolei osoby o wyższym poziomie wykształcenia miały większą szansę zetknąć się z potrzebą wykorzystywania nowych technologii w życiu zawodowym lub w trakcie procesów edukacyjnych. Występowanie zróżnicowania wykluczenia cyfrowego w regionach oraz względem wybranych cech społeczno-demograficznych powinno być uwzględniane przy kształtowaniu polityki regionalnej oraz działaniach mających na celu polepszanie jakości kapitału ludzkiego, szczególnie w momencie podejmowania decyzji odnośnie do wykorzystania środków unijnych w nowej perspektywie finansowej na lata 2014-2020.

Literatura

- Batorski D. (2009), *Wykluczenie cyfrowe w Polsce*, [w:] *Spoleczeństwo informacyjne*, D. Grodzka (red.), „Studia BAS” 2009, nr 3(19), s. 223-249
- Batorski D. (2011), *Korzystanie z technologii informacyjno-komunikacyjnych*, [w:] J. Czapiński i in. (red.), *Social Diagnosis 2011 Objective and Subjective Quality of Life in Poland (Diagnoza Społeczna 2011, Warunki i Jakość Życia Polaków) Full Report*, Contemporary Economics 5.3.
- Bartholomew D., Steele F., Galbraith J., Moustaki I. (2008), *Analysis of multivariate social science data*, CRC Press, A. Chapman & Hall Book, Boca Raton.
- Czerniawska D. (2012), *Wykluczenie cyfrowe. Strukturalne uwarunkowania korzystania z Internetu w Polsce i województwie mazowieckim*, MGG Conferences, Warszawa.
- Jakość życia* (2013), *Jakość życia. Kapitał społeczny, ubóstwo i wykluczenie społeczne w Polsce*, GUS, Warszawa.
- Kleinbaum D.G., Klein M. (2010), *Logistic regression: a self-learning text*, Springer, New York.
- Kopczewska K., Kopczewski T., Wójcik P. (2009), *Metody ilościowe w R: aplikacje ekonomiczne i finansowe*, CeDeWu. PL Wydawnictwa Fachowe, Warszawa.
- Rada Monitoringu Społecznego (2011), *Diagnoza społeczna, kwestionariusze 2011*. www.diagnoza.com (27.07.2013).
- OECD (2001), *Understanding the digital divide*, Paris.
- Wilkinson L. (2005), *The Grammar of Graphics*, Springer.
- Wojniak J. (2013), *Od podziału do nierówności – nowy wymiar cyfrowego wykluczenia*. Aequalitas, vol. 2, nr 1 (2).

ANALYSIS OF THE DIGITAL DIVIDE IN POLAND AT THE INDIVIDUAL AND REGIONAL LEVEL

Summary: The first- and second-level digital divide in Poland is considered in this paper. Regional differences in digital divide in 2008-2012 are presented by the means of heatmaps. Determinants of the second-level digital divide at the individual level, such as gender, age, education and place of residence, were analyzed by the application of ordinal logistic regression model.

Keywords: digital divide, ordinal logistic regression, matrix visualization.