

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 385

Taksonomia 25

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

Redaktorzy naukowci

Krzysztof Jajuga

Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2015

Redaktor Wydawnictwa: Aleksandra Śliwka

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 stronie internetowej Wydawnictwa
www.pracnaukowe.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 2015

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

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
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
Tomasz Bartłomowicz: Segmentacja konsumentów na podstawie preferencji wyrażonych uzyskanych metodą Maximum Difference Scaling	11
Barbara Batóg, Jacek Batóg, Andrzej Niemiec, Wanda Skoczylas, Piotr Waśniewski: Zastosowanie metod klasyfikacyjnych w identyfikacji kluczowych indyktorów osiągnięć w zarządzaniu wynikami przedsiębiorstw	20
Iwona Bąk: Wykorzystanie statystycznej analizy danych w badaniach turystyki transgranicznej na obszarach chronionych.....	28
Beata Bieszk-Stolorz: Ocena stopnia deprecjacji kapitału ludzkiego z wykorzystaniem nieliniowych modeli regresji.....	37
Mariola Chrzanowska, Nina Drejerska: Małe i średnie przedsiębiorstwa w strefie podmiejskiej Warszawy – określenie znaczenia lokalizacji z wykorzystaniem drzew klasyfikacyjnych.....	45
Adam Depta: Próba modelowania strukturalnego jakości życia osób jękaających się jako konstrukt ukrytego na podstawie kwestionariusza SF-36v2	53
Katarzyna Dębkowska: Wielowymiarowa analiza kondycji finansowej przedsiębiorstw sektora e-usług	63
Krzysztof Dmytrów, Mariusz Doszyń: Taksonomiczna procedura wspomagania kompletacji produktów w magazynie	71
Mariusz Doszyń, Sebastian Gnat: Propozycja procedury taksonomiczno-ekonometrycznej w indywidualnej wycenie nieruchomości.....	81
Marta Dziechciarz-Duda, Anna Król: Zastosowanie analizy <i>unfolding</i> i regresji hedonicznej do oceny preferencji konsumentów	90
Katarzyna Frodyma: Współzależność między poziomem rozwoju gospodarczego a udziałem energii ze źródeł odnawialnych w końcowym zużyciu w krajach Unii Europejskiej.....	99
Hanna Gruchociak: Porównanie struktury lokalnych rynków pracy wyznaczonych przy wykorzystaniu różnych metod w Polsce w latach 2006 i 2011 .	111
Alicja Grześkowiak, Agnieszka Stanimir: Postrzeganie środowiska pracy przez starszą i młodszą generację pracowników	120
Marta Hozer-Koćmiel, Christian Lis: Klasyfikacja krajów nadbałtyckich ze względu na czas prac wykonywanych w gospodarstwie domowym	129
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel: Zegar cyklu koniunkturalnego państw UE i USA w latach 1995-2013 w świetle badań synchronizacji.....	138
Aleksandra Łuczak: Wykorzystanie rozszerzonej interwałowej metody TOPSIS do porządkowania liniowego obiektów	147

Aleksandra Łuczak, Feliks Wysocki: Zintegrowane podejście do ustalania współczynników wagowych dla cech w zagadnieniach porządkowania linowego obiektów	156
Małgorzata Markowska, Danuta Strahl: Wykorzystanie klasyfikacji dynamicznej do identyfikacji wrażliwości na kryzys ekonomiczny unijnych regionów szczebla NUTS 2.....	166
Aleksandra Matuszewska-Janica, Marta Hozer-Koćmiel: Struktura zatrudnienia oraz wynagrodzenia kobiet i mężczyzn a przedmiotowa struktura gospodarcza w państwach UE.....	178
Anna M. Olszewska: Zastosowanie analizy korespondencji do badania związku pomiędzy zarządzaniem jakością a innowacyjnością przedsiębiorstw	187
Małgorzata Podogrodzka: Metoda aglomeracyjna w ocenie przestrzennego zróżnicowania starości demograficznej w Polsce	195
Ewa Roszkowska, Tomasz Wachowicz: Ocena ofert negocjacyjnych spoza dopuszczalnej przestrzeni negocjacyjnej.....	201
Ewa Roszkowska, Tomasz Wachowicz: Zastosowanie metody <i>unfolding</i> do wspomagania procesu negocjacji	210
Małgorzata Rószkiewicz: Próba diagnozy uwarunkowań poziomu wskaźnika braku odpowiedzi w środowisku polskich gospodarstw domowych.....	219
Marcin Salamaga: Próba identyfikacji muzycznych profili melomanów z wykorzystaniem drzew klasyfikacyjnych i regresyjnych	229
Agnieszka Sompolska-Rzechuła: Określenie czynników wpływających na prawdopodobieństwo poprawy poziomu rozwoju społecznego z wykorzystaniem modelu logitowego	239
Iwona Staniec: Wykorzystanie analizy czynnikowej w identyfikacji konstruktywów ukrytych determinujących ryzyko współpracy.....	248
Agnieszka Stanimir: Skłonność do zagranicznej mobilności młodszych i starszych osób	257
Mirosława Sztemberg-Lewandowska: Problemy decyzyjne w funkcjonalnej analizie głównych składowych.....	267
Tomasz Szubert: Demograficzno-społeczne determinanty określające subiektywny status jednostki w polskim społeczeństwie	276
Piotr Tarka: Własności 5- i 7-stopniowej skali Likerta w kontekście normalizacji zmiennych metodą Kaufmana i Rousseeuwa	286
Joanna Trzęsiok: Nielklasyczne metody regresji a problem odporności	296
Katarzyna Wawrzyniak: Ocena podobieństwa wyników uporządkowania województw uzyskanych różnymi metodami porządkowania	305
Katarzyna Wójcik, Janusz Tuchowski: Wykorzystanie metody opartej na wzorcach w automatycznej analizie opinii konsumenckich.....	314
Anna Zamojska: Zastosowanie analizy falkowej w ocenie efektywności funduszy inwestycyjnych	325

Summaries

Tomasz Bartłomowicz: Segmentation of consumers based on revealed preferences obtained with the Maximum Difference Scaling method	19
Barbara Batóg, Jacek Batóg, Andrzej Niemiec, Wanda Skoczylas, Piotr Waśniewski: Application of classification methods to identify the key performance indicators of performance management	27
Iwona Bąk: The application of statistical data analysis in the studies of cross-border tourism in protected areas.....	36
Beata Bieszk-Stolorz: Evaluating human capital depreciation by means of non-linear regression models.....	44
Mariola Chrzanowska, Nina Drejerska: Small and medium enterprises in the Warsaw suburban zone – determination of a localization’s role using classification trees	52
Adam Depta: An attempt of structural modelling of the quality of life of stuttering people as a latent construct, based on SF-36v2 questionnaire ...	62
Katarzyna Dębowska: Multidimensional analysis of financial condition of e-business services	70
Krzysztof Dmytrów, Mariusz Doszyń: Taxonomic procedure of supporting order-picking of products in a warehouse	80
Mariusz Doszyń, Sebastian Gnat: Taxonomic and econometric methods in individual real estate evaluation.....	89
Marta Dziechciarz-Duda, Anna Król: The application of unfolding analysis and hedonic regression in the investigation of consumers’ preferences	98
Katarzyna Frodyma: Interdependence between the level of economic development and the share of renewable energy in gross final energy consumption in the European Union.....	110
Hanna Gruchociak: Comparison of local labour markets structure designated using different methods in Poland in 2006 and 2011 years.....	119
Alicja Grześkowiak, Agnieszka Stanimir: Perception of working environment by older and younger generation of workers.....	128
Marta Hozer-Koćmiel, Christian Lis: Classification of the Baltic Sea Region countries due to the time of household work	137
Tadeusz Kufel, Magdalena Osińska, Marcin Błażejowski, Paweł Kufel: Business cycle clock for the EU and the USA in 1995-2013 in the light of synchronization research.....	146
Aleksandra Łuczak: The use of the extended interval TOPSIS methods for linear ordering of objects.....	155
Aleksandra Łuczak, Feliks Wysocki: Integrated approach for determining the weighting coefficients for features in issues of linear ordering of objects.....	165

Małgorzata Markowska, Danuta Strahl: The application of dynamic classification for the identification of vulnerability to economic crisis in the EU NUTS 2 regions	177
Aleksandra Matuszewska-Janica, Marta Hozer-Koćmiel: The structure of male and female employment and remuneration vs. the basic economy structure in the EU countries	186
Anna M. Olszewska: The application of the correspondence analysis for the study of the relations between quality management and innovation in the enterprises.....	194
Małgorzata Podogrodzka: Agglomeration method in the age and ageing in Poland by voivodships.....	200
Ewa Roszkowska, Tomasz Wachowicz: Scoring the negotiation offers from the outside of the feasible negotiation space	209
Ewa Roszkowska, Tomasz Wachowicz: Application of the unfolding analysis to negotiation support.....	218
Małgorzata Rószkiewicz: An attempt to diagnose the determinants of non-response rate in Polish households surveys	228
Marcin Salamaga: Attempt to identify music lovers profiles using classification and regression trees	238
Agnieszka Sompolska-Rzechuła: The definition of factors influencing the probability of improving the level of human development using the logit model.....	247
Iwona Staniec: The use of factor analysis to identify hidden constructs – determinants of the cooperation risk	256
Agnieszka Stanimir: Willingness to mobility abroad among younger and older persons	266
Mirosława Sztemberg-Lewandowska: Decision problems in functional principal components analysis.....	275
Tomasz Szubert: Socio-demographic factors determining subjective social status of an individual in Polish society	285
Piotr Tarka: Normalization methods of variables and measurement on 5 and 7 point Likert scale	295
Joanna Trzęsiok: Non-classical regression methods vs. robustness	304
Katarzyna Wawrzyniak: The evaluation of the similarity of the voivodships' orderings obtained by means of different methods.....	313
Katarzyna Wójcik, Janusz Tuchowski: Using pattern-based opinion mining.....	324
Anna Zamojska: Mutual funds performance measurement – wavelets analysis approach.....	333

Małgorzata Rószkiewicz

Szkoła Główna Handlowa w Warszawie

e-mail: mroszki@sgh.waw.pl

PRÓBA DIAGNOZY UWARUNKOWAŃ POZIOMU WSKAŹNIKA BRAKU ODPOWIEDZI W ŚRODOWISKU POLSKICH GOSPODARSTW DOMOWYCH

Streszczenie: W artykule podjęto próbę identyfikacji czynników mających wpływ na poziom wskaźnika braku odpowiedzi w środowisku polskich gospodarstw domowych w 2013 r. Podstawą analizy były wyniki historii kontaktów w badaniu na losowej próbie ponad 34 tys. gospodarstw domowych. Przeanalizowano oddzielnie uwarunkowania braku kontaktu oraz odmowy. W identyfikacji czynników prawdopodobieństwa braku odpowiedzi wykorzystano wielopoziomowy model regresji logistycznej wielomianowej. Wyniki ujawniły inne reguły uczestnictwa w badaniu w przypadku akceptacji kontaktu i inne dla realizacji wywiadu. Podejście wielopoziomowe ujawniło dodatkowo, że mniejsza skuteczność ankierterów w nawiązywaniu kontaktów z jednostkami badania przekładała się na ich mniejszą perswazyjność w nakłanianiu respondentów do współpracy.

Słowa kluczowe: wskaźnik braku odpowiedzi, błąd braku reakcji, błąd braku realizacji, wielopoziomowa regresja logistyczna wielomianowa.

DOI: 10.15611/pn.2015.385.24

1. Wstęp

Poziom wskaźnika braku odpowiedzi i będące tego konsekwencją obciążenia są wyznacznikiem jakości danych i jakości wyników prowadzonych analiz. Wyniki wielu badań terenowych wskazują na nasilanie się występowania braków odpowiedzi, a dodatkowo wraz z tą tendencją badacze odnotowują zmianę przyczyn ich występowania, co prowadzi do zmian natury obciążeń z tego tytułu [Martin, Matheson 1999; De Heer 1999; Steeh i in. 2001; Stoop 2005; Groves 2006]. W sytuacjach gdy wyniki pilotażu sygnalizują możliwość występowania wysokiego poziomu wskaźnika braków, rozważa się różne strategie mające przeciwdziałać samemu zjawisku, jak również jego konsekwencjom. Strategie te obejmują modyfikację metod doboru próby i metod dokonywania pomiarów oraz modyfikację

metod analizy danych. Szacowanie obciążeń wyników badania z tytułu braków odpowiedzi wymaga ustalenia, czy wynikają one z braku możliwości nawiązania kontaktu czy też z powodu odmowy udziału w badaniu, a także posiadania wiedzy o mechanizmach ich powstawania. Rozróżnienie braków odpowiedzi ze względu na przyczyny ich powstawania jest szczególnie ważne w badaniach porównawczych, gdyż ten sam poziom wskaźnika odpowiedzi może być kształtowany przy ich diametralnie odmiennej konfiguracji, co będzie prowadzić do błędów interpretacyjnych. Jednakże niezależnie od względów związanych z projektowaniem i organizacją badań terenowych problem braków danych jest również problemem badawczym, odnoszącym się do procesu, który ze względu na skalę występowania jest zjawiskiem społecznym i ma swoje uwarunkowania.

Dane dotyczące zaangażowania w proces badawczy polskich respondentów są rzadko, jeśli w ogóle, publikowane. Na ogół badacze ograniczają się do podania wskaźnika odpowiedzi, nie wyjaśniając charakteru braków odpowiedzi i ich uwarunkowań. Celem artykułu jest próba rozpoznania uwarunkowań występowania braków odpowiedzi w badaniach typu *face-to-face*, w środowisku polskich gospodarstw domowych. Odrębnie przeanalizowano uwarunkowania braku kontaktu oraz odmowy odpowiedzi, zakładając, że stanowią one dwa odrębne wymiary, na których są ulokowane jednostki badania [Lynn i in. 2002, s. 146].

2. Model badawczy

W literaturze tematu dominuje pogląd, że nie ma empirycznych dowodów na występowanie grup jednostek permanentnie nieuczestniczących w badaniach, czyli tzw. *hard-core*, które rezygnują zawsze i kontestują każdy proces badawczy, a jeśli takie istnieją, to są to bardzo nieliczne grupy i trudno rozważać je, przygotowując projekt badawczy [Goyder 1987, s. 187]. Na tej podstawie można przyjąć, że jednostki, które zbadano tylko i wyłącznie dzięki podjęciu większych wysiłków, są identyczne z tymi, które nie uczestniczyły w badaniu, a które tych dodatkowych wysiłków i środków wymagałyby. Oznacza to, że najtrudniejsi respondenci staliby się nieuczestniczącymi w badaniu, gdyby tych wysiłków nie było. Zaproponowana przez Lin i Schaeffer [1995] analogia między respondentami, którzy uczestniczyli w badaniu, ale z którymi nawiązanie kontaktu było trudne, oraz jednostkami w ogóle nieosiągalnymi w badaniu terenowym pozwala rozpoznawać uwarunkowania dostępności jednostek. Z kolei analogia między respondentami, którzy odmawiali udziału w badaniu przy kolejnych wizytach, ale ostatecznie wyrazili zgodę, oraz jednostkami, które definitywnie odmówiły udziału w badaniu, pozwala rozpoznawać uwarunkowania gotowości do współpracy. Podejście to wykorzystano w identyfikacji cech polskich gospodarstw domowych niedostępnych i niewspółpracujących.

Przedstawione w literaturze teorie wyjaśniające brak dostępności i brak gotowości do współpracy jednostek próby zakładają współgranie takich czynników, jak

cechy jednostek badania oraz cechy procesu badawczego i jego organizacji [Goyder 1987; Brehm 1993; Dillman 2000; Groves i in. 2000]. Rozszerza to zbiór potencjalnych uwarunkowań prawdopodobieństwa braku odpowiedzi, w którym obok charakterystyk jednostek badania występują cechy metodologii badania. Założenie to prowadzi do uznania, że model oddziaływań różnorodnych czynników na prawdopodobieństwo braku odpowiedzi ma strukturę wielopoziomową, gdzie na pierwszym poziomie występują cechy jednostek wybranych do badania, na drugim poziomie występują cechy organizacji procesu badawczego (np. ankieterów, którzy mogą powodować zmienność prawdopodobieństwa braków odpowiedzi między wiązkami jednostek badania), na poziomie trzecim zaś występują cechy metodologii realizowanego projektu (dobór próby i obszar badawczy ujęty w narzędziach badawczych), które mogą powodować zmienność prawdopodobieństwa braków odpowiedzi między projektami badawczymi.

3. Dane i metoda

Podstawą analizy były wyniki badania ankietowego zrealizowanego w 2013 r. na losowej próbie gospodarstw domowych w ramach projektu Uwarunkowania Decyzji Edukacyjnych (UDE), koordynowanego przez Instytut Badań Edukacyjnych, którego metodologię zaproponował zespół badaczy ze Szkoły Głównej Handlowej w Warszawie. Dysponując danymi jedynie z jednego projektu badawczego, należało ograniczyć rozważne podejście do dwóch poziomów, tj. poziomu badanych gospodarstw domowych oraz poziomu zespołu ankietarskiego realizującego badanie w terenie. W analizie uwarunkowań braków odpowiedzi wykorzystano zarejestrowaną historię kontaktów dla 122 831 wylosowanych adresów. Populację badaną stanowiły gospodarstwa domowe i ich członkowie, w których funkcjonują osoby w wieku 16-65 lat. Zrealizowano wywiady łącznie w 34753 gospodarstwach domowych.

Założono, że respondenci, którzy wymagali więcej niż jednej wizyty ankietera, by nawiązać kontakt i zrealizować wywiad, należą do grupy respondentów „trudnych”, którzy generowałiby braki odpowiedzi, gdyby ankieterzy nie ponieśli dodatkowych wysiłków związanych z kolejnymi wizytami. Wykorzystując analogie zaproponowane przez Lin i Schaeffer [1995], grupa ta stanowiła podstawę do wnioskowania o cechach jednostek badania, które generują błędy braku reakcji i braku realizacji. Zastosowane w analizie danych podejście wielopoziomowe pozwoliło dodatkowo uwzględnić wpływy nieobserwowanych cech ankietera na występowanie braku kontaktu oraz odmowy udziału w badaniu. Założono bowiem segmentację zarówno reakcji, jak i realizacji ze względu na przynależność jednostek badania do wiązek przynależnych poszczególnym ankieterom. Wybrane cechy gospodarstw domowych stanowiły zmienne niezależne pierwszego poziomu. Za zmienne niezależne drugiego poziomu przyjęto nieobserwowalne charakterystyki

ankieterów opisywane przez efekty losowe. Opisuje je zmienna losowa $U = [u^{(1)}, u^{(2)}]$ o dwuwymiarowym rozkładzie normalnym $N(0, \mathbf{\Omega})$, gdzie macierz $\mathbf{\Omega}$ wyznaczają wariancje $\sigma^{2(1)}$ i $\sigma^{2(2)}$ opisujące zmienność wewnątrzgrupową prawdopodobieństwa odpowiedzi odpowiednio dla $s=1$ (odmowa w pierwszym kontakcie) i $s = 2$ (brak kontaktu przy pierwszej wizycie) oraz kowariancja $\sigma^{(12)}$, która jest miarą zależności nieobserwowalnych wpływów ankierów na prawdopodobieństwo zarówno braku kontaktu, jak i odmowy udziału w badaniu. Wartość dodatnia tej kowariancji oznacza, iż ankierzy, którzy mają trudności w nawiązaniu kontaktu, osiągają również mniejszą perswazyjność względem jednostek badania, nakłaniając je do współpracy, wartość ujemna zaś, że ankierzy, którzy mają trudności w nawiązaniu kontaktu, osiągają bardzo dobre rezultaty w nakłanianiu respondentów do współpracy, czyli rekompensują porażki w nawiązywaniu kontaktów lepszą perswazyjnością udziału w badaniu. Wartość 0 oznacza zaś, że umiejętności ankiera w zakresie nawiązywania kontaktu i perswazyjności nie są ze sobą powiązane. Założono, że nie występują zależności między cechami badanych gospodarstw domowych oraz ukrytymi charakterystykami ankierów. Takie założenie jest oczywiste w przypadku prób adresowych. Występowanie takiej zależności oznaczałoby zaś, że schemat doboru próby uwzględniał dodatkowe informacje pozwalające łączyć cechy badanych jednostek z cechami zespołów ankierskich. Rozwiązanie takie jest możliwe, ale tu nie zostało rozważane.

Przyjęte założenia pozwalają na opisanie rezultatu pracy j -tego ankiera w i -tym gospodarstwie domowym za pomocą zmiennej nominalnej Y , dla której wartość 0 oznacza zrealizowanie wywiadu w pierwszym kontakcie, 1 oznacza odmowę udziału w badaniu w pierwszym kontakcie, 2 oznacza zaś brak kontaktu przy pierwszej wizycie. Na tej podstawie formuła wielopoziomowego modelu regresji logistycznej wielomianowej przyjmuje postać modelu z losowym wyrazem wolnym:

$$\log \left(\frac{\pi_{ij}^{(s)}}{\pi_{ij}^{(0)}} \right) = \boldsymbol{\beta}^{(s)T} \mathbf{x}_{ij}^{(s)} + u_j^{(s)},$$

gdzie: $\pi_{ij}^{(s)}$ – prawdopodobieństwo odpowiedzi, czyli $\pi_{ij}^{(s)} = P(y_{ij} = s)$ dla $s = 0, 1, 2$,
 $\mathbf{x}_{ij}^{(s)}$ – wektor predyktorów odnoszących się do cech gospodarstwa domowego oraz ankierów i interakcji międzypoziomowych,
 $\boldsymbol{\beta}^{(s)}$ – wektor współczynników modelu,
 $u_j^{(s)}$ – efekt losowy odnoszący się do nieobserwowalnych charakterystyk ankierów.

Taka specyfikacja modelu pozwala nie tylko na ocenę uwarunkowań każdego ze składników braku odpowiedzi, tj. braku reakcji i odrębnie odmowy udziału w badaniu przy pierwszej wizycie ankiera, ale również pozwala na analizę porównawczą tych uwarunkowań.

4. Wyniki

Wskaźnik odpowiedzi w grupie odnoszącej się do populacji badanej osiągnął poziom 40,02%. Wskaźnik reakcji osiągnął poziom 75,6%, zaś wskaźnik realizacji badania osiągnął poziom 40,04% w grupie należącej do populacji badanej. Jedynie 27,53% jednostek badania, wśród których efektywnie zrealizowano wywiady, pochodziło z próby podstawowej, 22,81% pochodziło z pierwszej rezerwy, 17,94% z drugiej rezerwy, pozostałe zaś 31,73% zbadanych jednostek pochodziło z trzeciej lub kolejnych prób rezerwowych. Najczęściej nie podawano przyczyny odmowy udziału w badaniu.

Wskaźnik reakcji, kształtujący się na dość wysokim poziomie, wykazywał znaczne zróżnicowanie regionalne. Najłatwiej respondenci byli osiągalni w województwie zachodniopomorskim oraz wielkopolskim, gdzie niespełna 16% kontaktów skończyło się definitywnym niepowodzeniem. Najtrudniej zaś w województwach dolnośląskim, podlaskim i opolskim, gdzie częstość braku kontaktu była ponad dwukrotnie wyższa.

Trudność w nawiązaniu kontaktu przekładała się na konieczność podejmowania wielokrotnych prób. Jednakże skuteczność w nawiązywaniu kontaktu przy kolejnych wizytach była niższa niż w pierwszym i drugim podejściu. Jak ilustrują to dane zestawione w tab. 1, w ramach pierwszej wizyty udawało się nawiązać kontakt w 84% przypadków (wylosowanych adresów), natomiast w ramach dalszych wizyt kontakt nawiązywano już tylko w około 59% przypadków.

Tabela 1. Struktura reakcji wylosowanych jednostek z uwzględnieniem liczby wizyt ankierów

Reakcja	Liczba wizyt					Ogółem
	1	2	3	4	5	
Brak kontaktu	16,0%	23,1%	61,3%	49,8%	48,5%	24,4%
Kontakt nawiązany	84,0%	76,9%	38,7%	50,2%	51,5%	75,6%
Ogółem	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Źródło: opracowanie własne na podstawie danych UDE.

Nawiązanie kontaktu nie gwarantowało jeszcze zrealizowania wywiadów. Jak ilustrują to dane zestawione w tab. 2, przy pierwszej wizycie udało się zrealizować wywiady jedynie wśród 37,4% gospodarstw należących do populacji badanej. Kolejne wizyty po nawiązaniu kontaktu dawały tylko trochę lepsze rezultaty.

Wskaźnik realizacji, ujawniający odmawianie współpracy wśród większości jednostek wylosowanych do próby, z którymi nawiązano kontakt, wykazywał znaczne zróżnicowanie regionalne. Najtrudniej respondenci podejmowali współpracę w województwie warmińsko-mazurskim oraz dolnośląskim, gdzie niespełna 33% nawiązanych kontaktów skończyło się podjęciem współpracy. Najłatwiej zaś

Tabela 2. Struktura realizacji wśród nawiązanych kontaktów z jednostek z populacji badanej z uwzględnieniem liczby wizyt ankieterów

Realizacja	Liczba wizyt					Ogółem
	1	2	3	4	5	
Odmowa	62,6%	52,8%	54,8%	52,1%	38,0%	59,9%
Wywiad zrealizowany	37,4%	47,2%	45,2%	47,9%	62,0%	40,1%
Ogółem	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Źródło: opracowanie własne na podstawie danych UDE.

w województwach podlaskim i świętokrzyskim, przy czym tam liczba zrealizowanych wywiadów wśród nawiązanych kontaktów ledwie przekroczyła połowę. Zasadnicza różnica między możliwością nawiązania kontaktu przy pierwszej wizycie względem tej możliwości przy kolejnych wizytach wskazuje, że uzasadnione jest założenie o analogii między respondentami trudnymi pod względem kontaktu i respondentami generującymi błąd braku reakcji. Zróżnicowanie regionalne wskaźnika braku reakcji sugeruje zaś zróżnicowanie reguł dostępności respondentów. Z kolei zróżnicowanie regionalne wskaźnika braku realizacji sugeruje występowanie zróżnicowania reguł uczestnictwa w badaniu.

W rozważanym wielopoziomowym modelu regresji logistycznej wielomianowej dla prawdopodobieństwa braku odpowiedzi przy pierwszym kontakcie zbiór zmiennych niezależnych pierwszego poziomu zawierał następujące charakterystyki głowy i samego gospodarstwa domowego:

- wiek,
- płeć,
- stan cywilny,
- poziom najwyższego uzyskanego wykształcenia formalnego,
- status na rynku pracy,
- charakter miejsca zamieszkania,
- wielkość gospodarstwa domowego,
- główne źródło utrzymania, określające typ gospodarstwa domowego,
- poziom dochodu ekwiwalentnego na osobę w gospodarstwie domowym.

Uzasadnieniem dla wykorzystania modelu dwupoziomowego, uwzględniającego nieobserwowalne efekty oddziaływania ankieterów na prawdopodobieństwo braku odpowiedzi, była wartość wskaźnika dopasowania modelu do danych, wyznaczonego jako bayesowskie kryterium informacyjne BIC, która dla modelu jednopoziomowego wyniosła 42 949,59, zaś dla modelu dwupoziomowego 40 137,89. Różnica w dopasowaniu na korzyść modelu dwupoziomowego, wynosząca 2811,7, sugeruje występowanie zmienności międzygrupowej zmiennej zależnej. Ocena kowariancji $\sigma^{(12)}$, będącej miarą zależności nieobserwowalnych wpływów ankieterów na prawdopodobieństwo braku kontaktu i odmowy udziału w badaniu, była istotna ($p < 0,001$) i wyniosła 0,397069. Wynik ten wskazuje, że ankieterzy, którzy

mieli trudności w nawiązaniu kontaktu, osiągnęli również mniejszą skuteczność w realizacji wywiadów. Wyższy wskaźnik braku reakcji współwystępował z wyższym wskaźnikiem braku realizacji.

W tabeli 3 zestawiono jedynie statystycznie istotne oceny parametrów modelu odnoszące się do wybranych cech głowy i gospodarstwa domowego, odpowiednio dla prawdopodobieństwa braku kontaktu (kolumny 2-4) i odmowy udziału w bada-

Tabela 3. Istotne oszacowania parametrów modelu regresji logistycznej wielomianowej dla zmiennych pierwszego poziomu

Zmienna niezależna	Brak kontaktu			Odmowa		
	współczynnik	błąd standardowy	istotność	współczynnik	błąd standardowy	istotność
	(2)	(3)	(4)	(5)	(6)	(7)
Główne źródło utrzymania						
Praca najemna	-0,071	0,047	0,131	0,37	0,133	0,006
Praca na własny rachunek	-0,013	0,067	0,846	-0,836	0,231	0,000
Rolnictwo	0,152	0,075	0,042	-1,083	0,354	0,002
Kapitał	2,398	0,521	0,000	-17,379	182,1	0,999
Nieruchomość	-20,425	672,503	0,998	0,13	0,642	0,839
Emerytura	0,011	0,055	0,836	0,205	0,154	0,182
Renta	-0,331	0,074	0,000	-0,374	0,228	0,101
Płeć	0,079	0,029	0,006	-0,271	0,071	0,000
Wiek	0,003	0,001	0,01			
Stan cywilny						
Wolny	0,01	0,045	0,828			
Związek	-0,158	0,062	0,010			
Wdowieństwo	-0,221	0,054	0,000			
Wykształcenie						
Podstawowe/gimnazjalne	0,105	0,113	0,352			
Zasadnicze zawodowe	0,239	0,106	0,025			
Średnie	0,263	0,106	0,013			
Pomaturalne/policealne	0,367	0,121	0,002			
Licencjackie/inżynierskie	0,184	0,118	0,117			
Magisterskie	0,359	0,109	0,001			
Stopień naukowy	-0,578	0,273	0,034			
Aktywność zawodowa						
Pracujący	0,175	0,037	0,000			
Bezrobotny	-0,283	0,063	0,000			
Miejsce zamieszkania						
Miasto do 10 tys.	-0,236	0,082	0,004	-0,233	0,293	0,426
miasto 10 tys.-19,9 tys.	0,309	0,052	0,000	0,048	0,198	0,809
miasto 20 tys.-49,9 tys.	0,373	0,053	0,000	1,252	0,145	0,000
miasto 50 tys.-99,9 tys.	0,299	0,055	0,000	1,521	0,139	0,000
miasto 100 tys.-199,9 tys.	0,43	0,048	0,000	1,307	0,136	0,000
miasto 200 tys.-499,9 tys.	1,109	0,046	0,000	1,45	0,140	0,000
miasto 500 tys.-999,9 tys.	0,965	0,067	0,000	0,157	0,310	0,613
Warszawa	0,952	0,058	0,000	2,707	0,130	0,000
Liczba osób	0,084	0,014	0,000	0,167	0,026	0,000
Stała	-2,211	0,141	0,000	-4,816	0,181	0,000

Źródło: opracowanie własne na podstawie danych UDE.

niu (kolumny 5-7) w odniesieniu do grupy referencyjnej, przy kontrolowanym wpływie pozostałych zmiennych. Kategorię referencyjną dla poziomu pierwszego wyznaczał następujący zestaw cech: kobieta, po rozwodzie lub w separacji, o wykształceniu co najwyżej podstawowym, bierna zawodowo, zamieszkująca w gminie wiejskiej oraz utrzymująca się głównie z niezarobkowych źródeł innych niż kapitał lub nieruchomości.

Wyniki oszacowań wskazują, że kombinacja cech składających się na kategorię referencyjną sprzyja zarówno nawiązywaniu kontaktu, jak i realizacji wywiadu. Dla tej grupy ryzyko braku kontaktu w pierwszej próbie było niższe o 89% względem pozostałych gospodarstw, a ryzyko odmowy przy nawiązaniu kontaktu było dwukrotnie mniejsze względem pozostałych gospodarstw domowych. Ponadto przy kontrolowanym wpływie pozostałych cech odnotowano następujące częściowe prawidłowości:

1. Ryzyko braku kontaktu w pierwszym kontakcie wzrasta wraz z wiekiem głowy gospodarstwa domowego.

2. W przypadku, gdy głową gospodarstwa domowego jest mężczyzna, brak kontaktu w pierwszej próbie staje się bardziej prawdopodobny.

3. W porównaniu z grupą referencyjną funkcjonowanie w związkach oraz wdowieństwo zmniejsza ryzyko braku kontaktu w pierwszej próbie.

4. W porównaniu z grupą referencyjną jedynie posiadanie stopni naukowych zmniejsza ryzyko braku kontaktu w pierwszej próbie. Dla wszystkich pozostałych poziomów wykształcenia ryzyko braku takiego kontaktu jest wyższe. Najwyższe ryzyko braku kontaktu w pierwszej próbie względem grupy referencyjnej mają gospodarstwa, których głowa legitymuje się wykształceniem pomaturalnym/policealnym lub ma ukończone studia magisterskie.

5. W porównaniu z grupą referencyjną status pracującego zwiększa ryzyko braku kontaktu w pierwszej próbie, natomiast status bezrobotnego ryzyko to zmniejsza.

6. W porównaniu z grupą referencyjną im większy ośrodek miejski, tym ryzyko braku kontaktu w pierwszej próbie jest wyższe.

7. W porównaniu z grupą referencyjną, jeśli źródłem utrzymania jest kapitał lub rolnictwo, to ryzyko braku kontaktu w pierwszej próbie jest wyższe, zaś mniejsze w przypadku renty. Jedynie te trzy źródła utrzymania istotnie zmieniały ryzyko braku kontaktu w pierwszej próbie.

8. Im większe gospodarstwo domowe, tym wyższe ryzyko braku kontaktu w pierwszej próbie.

Spśród rozważanych cech głowy i gospodarstwa domowego jedynie poziom dochodu ekwiwalentnego na osobę w gospodarstwie domowym okazał się nieistotnym czynnikiem prawdopodobieństwa braku kontaktu. Wynik ten należy interpretować ostrożnie, gdyż w przypadku tej zmiennej występowały znaczne braki danych, co zmusiło do stosowania imputacji, by nie eliminować z analizy znacznej grupy jednostek. Działania te mogły mieć wpływ na uzyskany rezultat.

W przypadku uwarunkowań odmowy udziału w badaniu spośród rozważanych cech głowy i gospodarstwa domowego istotne statystycznie okazały się jedynie 4 cechy, tj. płeć, miejsce zamieszkania, główne źródło utrzymania oraz wielkość gospodarstwa domowego. Ponadto przy kontrolowanym wpływie pozostałych cech odnotowano następujące cząstkowe prawidłowości:

1. W przypadku gdy głową gospodarstwa domowego jest mężczyzna, odmowa udziału w badaniu w pierwszym kontakcie staje się bardziej prawdopodobna.

2. W porównaniu z grupą referencyjną im większy ośrodek miejski, tym ryzyko odmowy w pierwszym kontakcie wyższe.

3. W porównaniu z grupą referencyjną jeśli źródłem utrzymania jest praca na własny rachunek lub rolnictwo, to ryzyko odmowy w pierwszym kontakcie jest niższe, zaś wyższe w przypadku pracy najemnej. Jedynie te trzy źródła utrzymania istotnie zmieniały ryzyko odmowy w pierwszym kontakcie względem grupy referencyjnej.

4. Im większe gospodarstwo domowe, tym wyższe ryzyko odmowy.

5. Zakończenie

Wyniki badania są zbieżne z sędami prezentowanymi w literaturze tematu o występowaniu tzw. trudnych respondentów, których charakteryzuje większe prawdopodobieństwo rezygnacji z udziału w badaniu, co pozwala nie tylko rozszerzyć wiedzę o pochodzeniu braków odpowiedzi, ale również identyfikować segmenty respondentów wymagające większych wysiłków w zakresie perswazyjności i budowania pozytywnego nastawienia do badania.

Wyniki analizy historii kontaktów zrealizowanych w ramach badania UDE potwierdzają, że oba wymiary mają inne uwarunkowania. Stwierdzone zależności między cechami demograficzno-społecznymi badanych gospodarstw i ich dostępnością oraz gotowością do współpracy wskazują, że brak kontaktu jest w pierwszej kolejności powiązany z cechami gospodarstw domowych opisującymi ich status społeczno-ekonomiczny, zaś odmowa udziału w badaniu jest powiązana z cechami gospodarstw domowych opisującymi ich środowisko. Ponadto poszczególne kategorie predyktorów wywołują odmienne efekty w kształtowaniu prawdopodobieństw obu typów braków odpowiedzi. Uzyskane wyniki potwierdzają przypuszczenie o zróżnicowaniu reguł uczestnictwa w badaniu wśród polskich gospodarstw domowych. Podejście wielopoziomowe ujawniło dodatkowo efekt procesu badawczego nakładający się na te reguły, a polegający na współwystępowaniu mniejszej skuteczności w nawiązywaniu kontaktów z mniejszą perswazyjnością w nakłanianiu do współpracy.

Literatura

- Brehm J. (1993), *The Phantom Respondents: Opinion Surveys and Political Representation*, Ann Arbor: University of Michigan Press.
- De Heer W. (1999), *International response trends, results of international survey*, Journal of Official Statistics, vol. 15, no. 2, s. 129-142.
- Dillman D.A. (2000), *Mail and Internet Surveys: The Tailored Design Method*, 2nd edition, New York, Wiley.
- Goyder J. (1987), *Surveys on surveys: limitation and potentialities*, Public Opinion Quarterly, vol. 50, s. 27-41.
- Groves R.M. (2006), *Nonresponse rates on nonresponse bias in household surveys*, Public Opinion Quarterly, vol. 70, s. 646-675.
- Groves R.M., Couper M.P. (1998), *Noresponse in Household Interview Surveys*, New York: Wiley.
- Groves R.M., Singer E., Corning A. (2000), *Leverage-saliency theory of survey participation. Description and illustration*, Public Opinion Quarterly, vol. 64, s. 299-308.
- Lin I-F., Schaeffer N.C. (1995), *Using survey participants to estimate the impact of nonparticipation*, Public Opinion Quarterly, vol. 59, s. 236-258.
- Lynn P. (2002), PEDAKSOIS: Methodology for Collecting Data about Survey Non-respondents, working papers of the Institute for Social and Economic Research, paper 2002-05. Colchester: University of Essex. www.iser.essex.ac.uk/pubs/workpaps/isr/2002-05.pdf
- Martin J., Matheson J. (1999), *Responses to declining response rate on government surveys*, Survey Methodology Bulletin, no. 45, s. 33-37.
- Steeh Ch., Kirgies N., Cannon B. and J. De Witt (2001), *Are they really as bad as they seem? Non-response rate at the end of twentieth century*, Journal of Official Statistics, vol. 17, no. 2, s. 227-247.
- Stinchcombe A.L., Jones C. and Sheatsley P. (1981), *Nonresponse bias for attitude questions*, Public Opinion Quarterly, vol. 45, s. 359-375.
- Stoop I.A.L. (2005), *The Hunt for the Last Respondent. Nonresponse in sample surveys*, Social and Cultural Planning Office of the Netherlands, The Hague.

AN ATTEMPT TO DIAGNOSE THE DETERMINANTS OF NON-RESPONSE RATE IN POLISH HOUSEHOLDS SURVEYS

Summary: The article focuses on the identification of factors affecting the nonresponse in Polish household surveys. The analyse uses data from the survey realized on a random sample of Polish households in 2013. Noncontact and noncooperation were considered separately in the study. Multilevel multinomial logistic regression model was used. The results confirmed that noncontact and noncooperation were two entirely different processes and rules for the participation in the study significantly differentiated regarding subpopulations of Polish households.

Keywords: noresponse rate, noncontact, noncooperation, multilevel multinomial logistic regression.