

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 426

Taksonomia 26

**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.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 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
Jacek Batóg: Identyfikacja obserwacji odstających w analizie skupień / Influence of outliers on results of cluster analysis	13
Andrzej Bąk: Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza / Linear ordering of objects using Hellwig and TOPSIS methods – a comparative analysis.....	22
Grażyna Dehnel: <i>MM</i> -estymacja w badaniu średnich przedsiębiorstw w Polsce / <i>MM</i> -estimation in the medium-sized enterprises survey in Poland.....	32
Andrzej Dudek: <i>Social network analysis</i> jako gałąź wielowymiarowej analizy statystycznej / Social network analysis as a branch of multidimensional statistical analysis.....	42
Iwona Foryś: Analiza dyskryminacyjna w wyborze obiektów podobnych w procesie szacowania nieruchomości / The discriminant analysis in selection of similar objects in the real estate valuation process	51
Gregory Kersten, Ewa Roszkowska, Tomasz Wachowicz: Ocena zgodności porządkowej systemu oceny ofert negocjatora z informacją preferencyjną / Analyzing the ordinal concordance of preferential information and resulting scoring system in negotiations.....	60
Iwona Konarzewska: Rankingi wielokryteriowe a współzależność liniowa kryteriów / Multi-criteria rankings and linear relationships among criteria	69
Anna Król, Marta Targaszewska: Zastosowanie klasyfikacji do wyodrębniania homogenicznych grup dóbr w modelowaniu hedonicznym / The application of classification in distinguishing homogeneous groups of goods for hedonic modelling.....	80
Marek Lubicz: Problemy doboru zmiennych objaśniających w klasyfikacji danych medycznych / Feature selection and its impact on classifier effectiveness – case study for medical data.....	89
Aleksandra Łuczak: Wpływ różnych sposobów agregacji opinii ekspertów w FAHP na oceny priorytetowych czynników rozwoju / Influence of different methods of the expert judgments aggregation on assessment of priorities for evaluation of development factors in FAHP.....	99
Iwona Markowicz: Tablice trwania firm w województwie zachodniopomorskim według rodzaju działalności / Companies duration tables in Zachodniopomorskie voivodship by the type of activity	108

Małgorzata Markowska, Danuta Strahl: Filary inteligentnego rozwoju a wrażliwość unijnych regionów szczebla NUTS 2 na kryzys ekonomiczny – analiza wielowymiarowa / Smart development pillars and NUTS 2 European regions vulnerability to economic crisis – a multidimensional analysis.....	118
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne deglomeracyjne sieci SOM w analizie skupień / The hierarchical divisive SOM in the cluster analysis	130
Kamila Migdał-Najman, Krzysztof Najman: Hierarchiczne aglomeracyjne sieci SOM w analizie skupień / The hierarchical agglomerative SOM in the cluster analysis	139
Barbara Pawelek, Józef Pocięcha, Jadwiga Kostrzewska, Mateusz Baryła, Artur Lipieta: Problem wartości odstających w prognozowaniu zagrożenia upadłością przedsiębiorstw (na przykładzie przetwórstwa przemysłowego w Polsce) / Problem of outliers in corporate bankruptcy prediction (case of manufacturing companies in Poland)	148
Wojciech Roszka: Syntetyczne źródła danych w analizie przestrzennego zróżnicowania ubóstwa / Synthetic data sources in spatial poverty analysis.....	157
Małgorzata Rószkiewicz: Czynniki różnicujące efektywność pracy ankietera w wywiadach <i>face-to-face</i> w środowisku polskich gospodarstw domowych / Factors affecting the efficiency of face-to-face interviews with Polish households.....	166
Adam Sagan, Marcin Pelka: Analiza wielopoziomowa z wykorzystaniem danych symbolicznych / Multilevel analysis with application of symbolic data	174
Marcin Salamaga: Zastosowanie drzew dyskryminacyjnych w identyfikacji czynników wspomagających wybór kraju alokacji bezpośrednich inwestycji zagranicznych na przykładzie polskich firm / The use of classification trees in the identification of factors supporting the choice of FDI destination on the example of Polish companies.....	185
Agnieszka Stanimir: Pomiar wykluczenia cyfrowego – zagrożenia dla Pokolenia Y / Measurement of the digital divide – risks for Generation Y ...	194
Mirosława Sztemberg-Lewandowska: Grupowanie danych funkcjonalnych w analizie poziomu wiedzy maturzystów / Functional data clustering methods in the analysis of high school graduates' knowledge	206
Tadeusz Trzaskalik: Modelowanie preferencji w wielokryterialnych dyskretnych problemach decyzyjnych – przegląd bibliografii / Preference modeling in multi-criteria discrete decision making problems – review of literature	214

Joanna Trzęsiok: Metody nieparametryczne w badaniu zaufania do instytucji finansowych / Nonparametric methods in the study of confidence in financial institutions	226
Hanna Wdowicka: Analiza sytuacji na lokalnych rynkach pracy w Polsce / Local labour market analysis in Poland.....	235
Artur Zaborski: Zastosowanie skalowania dynamicznego oraz metody wektorów dryfu do badania zmian w preferencjach / The use of dynamic scaling and the drift vector method for studying changes in the preferences.....	245

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. Przewodniczącymi Komitetu Organizacyjnego konferencji byli prof. dr hab. Mirosław Szreder oraz dr hab. Krzysztof Najman, prof. nadzw. UG, sekretarzami naukowymi dr hab. Kamila Migdał-Najman, prof. nadzw. UG oraz dr hab. Anna Zamojska, prof. nadzw. UG, a sekretarzem organizacyjnym Anna Nowicka z Fundacji Rozwoju Uniwersytetu Gdańskiego.

Konferencja Naukowa została dofinansowana ze środków Narodowego Banku Polskiego.

Zakres tematyczny konferencji obejmował takie zagadnienia, jak:

a) teoria (taksonomia, analiza dyskryminacyjna, metody porządkowania liniowego, metody statystycznej analizy wielowymiarowej, metody analizy zmiennych ciągłych, metody analizy zmiennych dyskretnych, metody analizy danych symbolicznych, metody graficzne),

b) zastosowania (analiza danych finansowych, analiza danych marketingowych, analiza danych przestrzennych, inne zastosowania analizy danych – medycyna, psychologia, archeologia, itd., aplikacje komputerowe metod statystycznych).

Zasadniczymi celami konferencji SKAD były prezentacja osiągnięć i wymiana doświadczeń z zakresu teoretycznych i aplikacyjnych zagadnień klasyfikacji i analizy danych. Konferencja stanowi coroczne forum służące podsumowaniu obecnego stanu wiedzy, przedstawieniu i promocji dokonań nowatorskich oraz wskazaniu kierunków dalszych prac i badań.

W konferencji wzięło udział 81 osób. Byli to pracownicy oraz doktoranci następujących uczelni i instytucji: AGH w Krakowie, Politechniki Łódzkiej, Politechniki Gdańskiej, Politechniki Opolskiej, Politechniki Wrocławskiej, Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie, Szkoły Głównej Handlowej w Warszawie, Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Ekonomicznego w Katowicach, Uniwersytetu Ekonomicznego w Krakowie, Uniwersytetu Ekonomicznego w Poznaniu, Uniwersytetu Ekonomicznego we Wrocławiu, Uniwersytetu Gdańskiego, Uniwersytetu Jana Kochanowskiego w Kielcach, Uniwersytetu Łódzkiego, Uniwersytetu Mikołaja Kopernika w Toruniu, Uniwersytetu Przyrodniczego w Poznaniu, Uniwersytetu Szczecińskiego, Uniwer-

sytetu w Białymstoku, Wyższej Szkoły Bankowej w Toruniu, a także przedstawiciele NBP i PBS Sp. z o.o.

W trakcie dwóch sesji plenarnych oraz trzynastu 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. Obradom w poszczególnych sesjach konferencji przewodniczyli profesorowie: Józef Pocięcha, Eugeniusz Gatnar, Tadeusz Trzaskalik, Krzysztof Jajuga, Marek Walesiak, Barbara Pawełek, Feliks Wysocki, Ewa Roszkowska, Andrzej Sokołowski, Andrzej Bąk, Tadeusz Kufel, Mirosław Krzyśko, Krzysztof Najman, Małgorzata Rószkiewicz, Mirosław Szreder.

Teksty 25 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii „Taksonomia” nr 26. Pozostałe recenzowane artykuły znajdują się w „Taksonomii” nr 27.

W pierwszym dniu konferencji odbyło się posiedzenie członków Sekcji Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego, któremu przewodniczył prof. dr hab. Józef Pocięcha. Ustalono plan przebiegu zebrania obejmujący następujące punkty:

- A. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS.
- B. Informacje dotyczące planowanych konferencji krajowych i zagranicznych.
- C. Organizacja konferencji SKAD PTS w latach 2016 i 2017.
- D. Wybór przedstawiciela Rady Sekcji SKAD PTS do IFCS.
- E. Dyskusja nad kierunkami rozwoju działalności Sekcji.

Prof. dr hab. Józef Pocięcha otworzył posiedzenie Sekcji SKAD PTS. Sprawozdanie z działalności Sekcji Klasyfikacji i Analizy Danych PTS przedstawiła sekretarz naukowy Sekcji dr hab. Barbara Pawełek, prof. nadzw. UEK. Poinformowała, że obecnie Sekcja liczy 231 członków. Przypomniała, że na stronie internetowej Sekcji znajdują się regulamin, a także deklaracja członkowska. Poinformowała, że zostały opublikowane zeszyty z serii „Taksonomia” nr 24 i 25 (PN UE we Wrocławiu nr 384 i 385). W „Przeglądzie Statystycznym” (zeszyt 4/2014) ukazało się sprawozdanie z ubiegłorocznej konferencji SKAD, która odbyła się w Międzyzdrojach, w dniach 8–10 września 2014 r. Prof. Barbara Pawełek przedstawiła także informacje dotyczące działalności międzynarodowej oraz udziału w ważnych konferencjach członków i sympatyków SKAD.

W konferencji Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS – International Federation of Classification Societies) w dniach 6–8 lipca 2015 r. w Bolonii, zorganizowanej przez Università di Bologna, udział wzięło 19 osób z Polski (w tym 17 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 79,0%). Ponadto prof. Józef Pocięcha był członkiem Komitetu Naukowego Konferencji z ramienia SKAD, członkiem Międzynarodowego Komitetu Nagród IFCS oraz organizatorem i przewodniczącym sesji nt. „Classification models for forecasting of economic processes”.

W konferencji „European Conference on Data Analysis” (Colchester, 2–4 września 2015 r.) zorganizowanej przez The German Classification Society (GfKI) we współpracy z The British Classification Society (BCS) i Sekcją Klasyfikacji i Analizy Danych PTS (SKAD) udział wzięło 18 osób z Polski (w tym 14 członków Sekcji), które wygłosiły 15 referatów (wkład członków SKAD – 66,0%). Ponadto profesorowie Krzysztof Jajuga oraz Józef Pociecha byli członkami Komitetu Naukowego konferencji, prof. Andrzej Dudek został poproszony przez organizatorów o przygotowanie referatu i wygłoszenie na Sesji Plenarnej „Cluster analysis in XXI century, new methods and tendencies”, prof. Krzysztof Jajuga był przewodniczącym sesji plenarnej, przewodniczącym sesji nt. „Finance and economics II” oraz organizatorem i przewodniczącym sesji nt. „Data analysis in finance”, prof. Józef Pociecha był organizatorem i przewodniczącym sesji nt. „Outliers in classification procedures – theory and practice”, prof. Andrzej Dudek był przewodniczącym sesji nt. „Machine learning and knowledge discovery II”.

Kolejny punkt posiedzenia Sekcji obejmował zapowiedzi najbliższych konferencji krajowych i zagranicznych, których tematyka jest zgodna z profilem Sekcji. Prof. dr hab. Józef Pociecha poinformował o dwóch wybranych konferencjach krajowych (były to XXXIV Konferencja Naukowa „Multivariate Statistical Analysis MSA 2015”, Łódź, 16–18 listopada 2015 r. i X Międzynarodowa Konferencja Naukowa im. Profesora Aleksandra Zeliasia nt. „Modelowanie i prognozowanie zjawisk społeczno-gospodarczych”, Zakopane, 10–13 maja 2016 r.) oraz o trzech wybranych konferencjach zagranicznych. Konferencja „European Conference on Data Analysis” odbędzie się na Uniwersytecie Ekonomicznym we Wrocławiu w dniach 26–28 września 2017 r. W przeddzień tej konferencji, tj. 25.09.2017 r., odbędzie się Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2017”. Następna konferencja Międzynarodowego Stowarzyszenia Towarzystw Klasyfikacyjnych (IFCS) odbędzie się w 2017 r. w Tokio. W 2019 r. Niemiecko-Polskie Sympozjum nt. „Analizy danych i jej zastosowań GPSDAA 2019” organizuje prof. Andreas Geyer-Schultz w Karlsruhe.

W następnym punkcie posiedzenia podjęto kwestię organizacji kolejnych konferencji SKAD. SKAD 2016 zorganizuje Katedra Metod Statystycznych Wydziału Ekonomiczno-Socjologicznego Uniwersytetu Łódzkiego.

W kolejnej części zebrania dokonano wyboru przedstawiciela Rady Sekcji SKAD PTS do IFCS na kadencję 2016–2019. Powołano Komisję Skrutacyjną, której przewodniczącym został prof. Tadeusz Kufel, a członkami dr hab. Iwona Konarzewska i dr Dominik Rozkrut. Profesor Józef Pociecha poprosił zebranych o proponowanie kandydatur zgłaszając jednocześnie prof. Andrzeja Sokołowskiego. Wobec braku następnych kandydatur listę zamknięto. Komisja Skrutacyjna przeprowadziła głosowanie tajne. W głosowaniu uczestniczyło 41 członków Sekcji. Profesor Andrzej Sokołowski został przedstawicielem Rady Sekcji SKAD PTS do

IFCS na kadencję 2016–2019, uzyskując następujący wynik: 39 głosów na „tak”, 1 głos na „nie”, 1 głos był nieważny.

W ostatnim punkcie zebrania dyskutowano nad kierunkami rozwoju działalności Sekcji obejmującymi następujące problemy: udział w międzynarodowym ruchu naukowym (wspólne granty, publikacje), umiędzynarodowienie konferencji SKAD (uczestnicy zagraniczni, dwujęzyczność konferencji), wydawanie własnego czasopisma.

Profesor Józef Pociecha zamknął posiedzenie Sekcji SKAD.

Krzysztof Jajuga, Marek Walesiak

Małgorzata Rószkiewicz

Szkoła Główna Handlowa w Warszawie
e-mail: mroszki@sgh.waw.pl

**CZYNNIKI RÓŻNICUJĄCE EFEKTYWNOŚĆ
PRACY ANKIETERA W WYWIADACH
FACE-TO-FACE W ŚRODOWISKU
POLSKICH GOSPODARSTW DOMOWYCH**
**FACTORS AFFECTING THE EFFICIENCY
OF FACE-TO-FACE INTERVIEWS
WITH POLISH HOUSEHOLDS**

DOI: 10.15611/pn.2016.426.17

Streszczenie: Celem artykułu jest klasyfikacja gospodarstw domowych w Polsce ze względu na gotowość do podejmowania współpracy z ankieterem w badaniach typu *face-to-face* w zależności od intensywności działań stymulujących udział w badaniu. Podstawą analizy były wyniki badania ankietowego zrealizowanego w 2013 r. na losowej próbie gospodarstw domowych. W ocenie wpływu wybranych cech głowy i samego gospodarstwa domowego na prawdopodobieństwo podjęcia współpracy w kolejnych próbach podejmowanych przez ankietera wykorzystano modele regresji logistycznej wielomianowej, tzn. dla prawdopodobieństwa nawiązania kontaktu oraz prawdopodobieństwa zrealizowania wywiadu w kolejnej wizycie ankietera. Uzyskane wyniki wskazują, że przebieg kolejnych prób nawiązania kontaktu w istotny sposób jest powiązany z większością cech społeczno-ekonomicznych gospodarstw domowych. Kluczowe są tu zarówno środowisko wylosowanych jednostek, jak i ich cechy społeczno-ekonomiczne, które kształtują nastawienie do procesu badawczego.

Słowa kluczowe: wskaźnik odpowiedzi, nawiązanie kontaktu, odmowa udziału w badaniu, regresja logistyczna wielomianowa.

Summary: The aim of the article is the identification of factors affecting willingness to cooperate with the interviewer in face-to-face contact depending on the intensity of activities to stimulate participation in the study. The analysis was based on the results of a survey carried out in 2013 on random sample about 34 thousand of Polish households. Non-contact and non-cooperation were considered separately in the study. To assess the impact of selected features of the head of a household on the probability to cooperate in subsequent attempts undertaken by the interviewer (max. 5) multinomial logistic regression models were used. The results indicate that the course of attempts to make contact is associated in a significant way with the majority of socio-economic characteristics of households. It turned out that important are both the environment of randomly selected individuals and their

socio-economic characteristics that shape their attitudes towards the research process but the realization of interview is determined by the demographic characteristics of the units.

Keywords: nonresponse rate, noncontact, noncooperation, multinomial logistic regression, household characteristics.

1. Wstęp

W sytuacjach gdy wyniki pilotażu sugerują, iż w badaniu właściwym można spodziewać się wysokiego poziomu wskaźnika braków odpowiedzi, rozważa się różne strategie umożliwiające ograniczenie wpływu obciążeń z tytułu tego typu błędów nielosowych [De Heer 1999; Groves 2006; Martin, Matheson 1999; Rószkiewicz i in. 2013; Steeh i in. 2001; Stoop 2005]. W pierwszej kolejności rozważa się strategie gwarantujące osiągnięcie wymaganej liczebności próby. Strategie te, z racji charakterystycznych dla nich działań, można określić mianem ekstensywnych lub intensywnych. Strategie ekstensywne obejmują zwielokrotnienie rozmiaru próby lub przygotowanie transz prób rezerwowych. Strategie intensywne obejmują działania, które dążą do poprawy wskaźnika odpowiedzi w założonym rozmiarze próby. Strategie ekstensywne na ogół wiążą się z wysokimi kosztami. Wymagają wylosowania większej liczby jednostek wyboru, zaangażowania i przeszkolenia większej grupy ankieterów, a także wydłużają proces realizacji badania w terenie. Ograniczony budżet oraz założony harmonogram realizacji projektu badawczego skłaniają do wykorzystywania strategii intensywnych. Ich powodzenie zależy od przygotowanych działań stymulujących udział jednostek w badaniu oraz od rozpoznania uwarunkowań postaw tych jednostek wobec samego procesu badawczego. Z tego powodu problem braków danych może być traktowany jako problemem badawczy, odnoszący się do procesu, który ma swoje uwarunkowania i ze względu na skalę występowania może być traktowany jako proces społeczny.

Celem artykułu jest analiza uwarunkowań podejmowania współpracy z ankieterem w badaniach typu *face-to-face* w środowisku gospodarstw domowych w Polsce w zależności od intensywności podejmowanych wysiłków przez realizujących proces badawczy.

2. Model badawczy

Model interakcji ankieter-respondent przyjmuje inną postać w przypadku nawiązania kontaktu oraz w przypadku podejmowania współpracy po nawiązaniu kontaktu. Każda kolejna próba nawiązania kontaktu z tym samym respondentem może być traktowana jako zdarzenie losowe, które ma tylko dwa rozwiązania, tj.: nawiązanie kontaktu lub jego brak. Kiedy kontakt zostanie już nawiązany, sytuacja zmienia się diametralnie, gdyż ta faza procesu badania empirycznego ulega

zakończeniu i rozpoczyna się faza podjęcia współpracy. Wówczas pierwszy kontakt może skończyć się na trzy sposoby, tj. przeprowadzeniem wywiadu, odmową udziału w badaniu lub decyzją o ponownym kontakcie i podjęciu próby przekonania potencjalnego respondenta do udziału w badaniu mimo wszystko. W kolejnym kontakcie sytuacja powtarza się, tworząc sekwencję interakcji badacza z potencjalnym respondentem. Każda kolejna próba podjęcia współpracy z tym samym respondentem, już po nawiązaniu kontaktu, może być traktowana jako zdarzenie losowe, a o przebiegu tej fazy badania decydują postawy respondentów wobec procesu badawczego. Ich zróżnicowanie kształtuje poziom wskaźnika braków odpowiedzi.

Dotychczasowe podejścia badawcze do zagadnienia uczestnictwa jednostek w badaniach terenowych identyfikują szereg kluczowych czynników braków odpowiedzi [Schnell 1997, s.190; Goyder 1987, s. 187]. Należą do nich charakterystyki gospodarstw domowych, cechy ankierców, środowisko społeczne i metodologia badania. Ramy konceptualne dla wyjaśnienia wpływu indywidualnych charakterystyk gospodarstw domowych na ich uczestnictwo w badaniach terenowych opierają się na takich teoriach, jak teoria wymiany społecznej [Goyder 1987; Dillman 2000], koncepcja społeczeństwa obywatelskiego [Brehm 1993] oraz teoria społecznej integracji [Goyder 1987]. Najnowszą w tym nurcie jest teoria dźwigni-wyróżnika (*leverage-saliency theory*) [Groves, Singer, Corning 2000], która koncentruje się na interakcjach między charakterystykami jednostek w próbie i metodologią badania. Teorie te wyjaśniają dostępność i gotowość współpracy jednostek próby wpływem kontekstu społecznego, opisują współgranie wielu czynników uczestnictwa w badaniu oraz definiują mechanizmy oddziaływania cech jednostek na realizację założonej metodologii badania. W badaniach typu *face-to-face* jednym z komponentów przyjętej metodologii, który odgrywa główną rolę w kontaktach z jednostkami badania i osiągnięciu współpracy, jest ankierca. Wyniki analizy przedstawione przez M. Rószkiewicz [2015] dowodzą występowania wpływu cech pracy ankiercy na uczestnictwo w badaniach terenowych realizowanych wśród polskich gospodarstw domowych. Poniżej podjęto próbę rozpoznania uwarunkowań ogniskowania się postaw pozytywnych oraz postaw negatywnych względem procesu badawczego w zależności od liczby prób podejmowanych przez realizujących badanie. Poszukiwano odpowiedzi na pytanie, czy występują odrębne grupy respondentów wśród polskich gospodarstw domowych, które wymagają dodatkowych wysiłków, by podjęły współpracę z ankiercą i na ile wysiłki te powinny być intensywne.

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, koordynowanego przez Instytut Badań Edukacyjnych, którego

metodologię zaproponował zespół badaczy ze Szkoły Głównej Handlowej w Warszawie. Zrealizowano wywiady łącznie w 34 753 gospodarstwach domowych. Wskaźnik odpowiedzi (*response-rate*) osiągnął poziom 40,02%. Trudność w nawiązaniu kontaktu przekładała się na konieczność podejmowania wielokrotnych prób. Kontakt uznawano za niemożliwy w przypadku porażki przy piątej próbie. W kolejnych próbach nawiązania kontaktu ich skuteczność była niższa niż w pierwszym i drugim podejściu. W ramach pierwszej wizyty udawało się nawiązać kontakt w 84% przypadków (wylosowanych adresów), natomiast w ramach dalszych wizyt kontakt nawiązywano już tylko w ok. 59% przypadków. Podczas pierwszej wizyty udało się zrealizować wywiady jedynie wśród 37,4% gospodarstw. Kolejne wizyty po nawiązaniu kontaktu dawały tylko trochę lepsze rezultaty.

W ocenie wpływu wybranych cech głowy i samego gospodarstwa domowego na prawdopodobieństwo podjęcia współpracy w ramach kolejnych wizyt ankietera wykorzystano model regresji logistycznej wielomianowej. W modelowaniu zmiennej zależnej o wielu kategoriach (*multinomial logit model*) zakłada się, że logarytm ilorazu szans (*log odds*) wyraża się modelem liniowym:

$$\eta_{ij} = P(Y_i = j) = \log(p_{ij}/p_{i1}) = \alpha_j + \mathbf{x}_i \beta_j,$$

gdzie: Y – zmienna losowa, która może przyjąć jedną z kategorii $j = 1, 2, \dots, J$,
 j – numer wizyty ankietera, α_j – stała, β_j – wektor współczynników regresji przy zmiennych objaśniających \mathbf{x}_i dla $j = 1, 2, \dots, J - 1$ [Rodriguez 2007].

Oszacowano odrębnie modele dla szansy nawiązania kontaktu w j -tej wizycie ankietera (dla $j = 1, 2, 3, 4, 5$) względem możliwości nawiązania kontaktu przy wcześniejszej wizycie oraz dla szansy zrealizowania wywiadu w j -tej wizycie ankietera (dla $j = 1, 2, 3, 4$) względem możliwości zrealizowania wywiadu przy wcześniejszej wizycie, gdzie wektor zmiennych objaśniających \mathbf{x}_i odnosił się do cech respondenta (głowy gospodarstwa domowego lub osoby najlepiej zorientowanej w sytuacji gospodarstwa) i cech jego gospodarstwa domowego. Zbiór zmiennych objaśniających zawierał następujące charakterystyki respondenta i jego 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.

4. Wyniki

W tab. 1 zestawiono istotne oceny parametrów dla efektów głównych odpowiednio w modelach dla szansy nawiązania kontaktu oraz dla szansy przeprowadzenia wywiadu w kolejnej wizycie ankietera względem ostatniej z podjętych próby, czyli piątej wizyty. W ocenie współczynników wykorzystano kategorie referencyjne,

Tabela 1. Wyniki oszacowania modelu regresji logistycznej wielomianowej odpowiednio dla szansy nawiązania kontaktu oraz szansy zrealizowania wywiadu w kolejnych próbach względem ostatniej, piątej próby

Zmienna niezależna/kategoria referencyjna	Próba kontaktu/ kategoria referencyjna			Próba wywiadu po nawiązaniu kontaktu/ kategoria referencyjna		
	2/5	3/5	4/5	2/5	3/5	4/5
1	2	3	4	5	6	7
Stała	3,476	3,549	2,077			
Kobieta						
Mężczyzna	-,194	-,281	-,080	1,293	1,130	2,281
Wiek/66 lat i więcej						
16–18 lat						
19–26 lat	-,252	-1,072	-,322			
27–49 lat	,249	-,091		8,828	8,442	
50–65 lat	-,199	-,556	-,310	8,713	8,182	
Stan cywilny/rozwód, separacja						
Kawaler, panna	,078	-,126	-,134			
Osoby w związku	,147	,054	-,154			
Osoby owdowiałe	-,490	-,455	-,806			
Aktywność zawodowa/bierni						
Pracujący	-,661	-,388	-,255			
Bezrobotni	-,733	-,210	-,298			
Miejsce zamieszkania/miasta co najmniej 1 mln mieszkańców						
Wieś	,669	,221	,229			
Miasta do 10 tys.	,478	-,385	,543			
Miasta 10–19,9 tys.	,249	,033	,477			
Miasta 20–49,9 tys.	,580	-,208	,218			
50–99,9 tys.	-,720	-1,071	-,915			
Miasta 100–199,9 tys.	-,408	-,108	-,116			
Miasta 200–499,9 tys.	-,413	-,584	-,223			
Miasta 500–999,9 tys.	-1,789	-1,349	-1,016			
Źródło utrzymania/odmowa						
Praca na umowę	-,736	-1,487	-,511			
Praca na własny rachunek		-,905	1,451			
Gospodarstwo rolne	,269	-,779	,344			
Kapitał	1,063	-1,238				
Emerytura	,125	-,714	1,048			
Renta		-,685	1,107			
Niezarobkowe	-,167	-,820	-,142			
Średni miesięczny dochód ekwiwalentny na 1 osobę/powyżej 2000 zł						
do 250 zł na 1 osobę	,478	,114	,275			
250–500 zł	,210	,227	-,115			
500–750 zł	,358	,151	,136			
750–1250 zł	-,781	-1,369	-1,091			

1	2	3	4	5	6	7
1250–2000 zł	,958	,291	,605			
Liczba osób w gospodarstwie/ 5 i więcej						
1	–,154	,900	–,361			
2	–,652	–,143	–1,018	,758	1,504	–2,523
3	,168	,399	–,666	–,812	–,281	–2,286
4	–,247	,155	–,511			
Poziom wykształcenia głowy gospodarstwa/studia II stopnia, stopień naukowy						
Co najwyżej gimnazjalne	1,075	,610	1,356			
Zasadnicze zawodowe	,403	–,164	,139			
Średnie	,717	,769	,870			
Pomaturalne, policealne	–,103	–,694	–,125			
Studia pierwszego stopnia	,604	,406	1,287			

Źródło: obliczenia własne na podstawie danych UDE.

które dotyczyły takich cech, jak: kobieta, grupa wieku 66 lat lub więcej, stan cywilny – po rozwodzie lub w separacji, wykształcenie co najwyżej podstawowe, bierność zawodowa, zamieszkiwanie w gminie wiejskiej oraz utrzymywanie się głównie z niezarobkowych źródeł innych niż kapitał lub nieruchomości, a także średni miesięczny dochód ekwiwalentny na 1 osobę powyżej 2000 zł oraz 5 osób lub więcej dla liczby osób w gospodarstwie domowym.

Wyniki oszacowań wskazują, że wśród respondentów tzw. trudnych, którzy wymagali dodatkowych prób zarówno w nawiązaniu kontaktu, jak i w nakłonieniu do współpracy po nawiązaniu kontaktu w kolejnych próbach oba zdarzenia mające wpływ na wskaźnik odpowiedzi miały inne uwarunkowania. O nawiązaniu kontaktu w dodatkowych próbach podejmowanych przez ankietera w sposób istotny decydowały cechy związane zarówno ze statusem społeczno-ekonomicznym, jak i ze środowiskiem, w którym funkcjonują gospodarstwa domowe. Środowisko wiejskie i miast o liczbie mieszkańców nie większej niż 50 tys. okazały się najłatwiejsze w nawiązywaniu kontaktu niż środowiska wielkomiejskie. Najtrudniejszą grupą w nawiązywaniu kontaktu okazały się ośrodki o liczbie mieszkańców od 50 tys. do 1 mln. Wśród tych grup respondentów konieczne były wielokrotne próby, a skuteczna okazywała się dopiero ostatnia.

Po nawiązaniu kontaktu możliwość podjęcia współpracy, w kolejnych próbach podejmowanych przez ankietera, miała zaś inne uwarunkowania, które w największym stopniu wiązały się w pierwszej kolejności z cechami demograficznymi, a w mniejszym stopniu ze statusem społeczno-ekonomicznym respondentów, zaś środowisko, w którym funkcjonowały jednostki okazało się czynnikiem nieistotnym. Mężczyźni znacznie trudniej nawiązywali kontakt w każdej kolejnej próbie, i wymagali znacznej liczby prób, kobiety zaś były bardziej dostępne, lecz decydowały się na współpracę dopiero po wielokrotnych namowach. Najłatwiej nawiązy-

wano kontakt z osobami najstarszymi. Jeśli już kontakt został nawiązany, to osoby młodsze szybciej decydowały się na współpracę i nie wymagały wielokrotnych prób skłonienia do udziału w badaniu. Najłatwiej nawiązywano kontakt z osobami z gospodarstw trzyosobowych, ale też najtrudniej było je skłonić do współpracy. Najtrudniejszymi respondentami w każdej kolejnej próbie nakłonienia do współpracy okazały się osoby z tych gospodarstw. Z kolei jednostki o wyższej pozycji społecznej albo decydowały się od razu na współpracę, albo w ogóle zrezygnowały z udziału w badaniu. W tym przypadku kolejne próby nawiązania współpracy nie miały sensu.

Wyniki oszacowań wskazują ponadto, że zestaw cech składających się na kategorie referencyjne zwiększał szansę na nawiązanie kontaktu w każdej próbie, lecz nie gwarantował możliwości przeprowadzenia wywiadu po nawiązaniu kontaktu. Najłatwiej dostępna grupa była najmniej skłonna do współpracy. Jest to zatem grupa respondentów, która pochłania najwięcej wysiłków realizujących badanie, gdyż łatwy kontakt skłania do ponawiania prób realizacji badania, które w wielu przypadkach okazują się bezskuteczne, a to obniża efektywność pracy ankietera. Wobec tej grupy konieczne jest zatem wypracowanie odrębnych strategii pozytywnego nastawienia do procesu badawczego, intensywniejszych niż dla pozostałych typów respondentów.

5. Zakończenie

Wyniki analizy historii kontaktów zrealizowanych w ramach badania UDE wskazują, że skuteczność pracy ankietera w kolejnych próbach współpracy z respondentem ma inne uwarunkowania w przypadku samego nawiązywania kontaktu oraz na etapie realizowania wywiadu już po nawiązaniu kontaktu. Jak wskazują wyniki badania, przebieg kolejnych prób nawiązania kontaktu w istotny sposób jest powiązany z większością cech społeczno-ekonomicznych gospodarstw domowych. Kluczowe jest tu zarówno środowisko wylosowanych jednostek jak i ich cechy społeczno-ekonomiczne, które kształtują nastawienie do procesu badawczego. Kiedy kontakt zostaje nawiązany, sytuacja zmienia się diametralnie, nie tylko ze względu na jej istotę, co opisano w modelu badawczym, ale i ze względu na uwarunkowania. W przebiegu tej fazy cechy demograficzne badanych jednostek okazały się zasadnicze dla powodzenia procesu badawczego. Najważniejsze w kształtowaniu skuteczności kolejnych prób okazały się cechy związane z sytuacją rodzinną jednostek. Wyniki sugerują zatem, że strategia standaryzowanych szkoleń ankieterów jest mało skuteczna i może obniżyć efektywność całego przedsięwzięcia. Uwzględnienie zróżnicowania typów respondentów, odrębnie ze względu na dostępność i odrębnie ze względu na gotowość współpracy, w ramach przygotowywania ankieterów do realizacji badania w terenie, wymaga różnicowania strategii nawiązywania kontaktu odrębnie wobec strategii nakłaniania do udziału w badaniu.

Literatura

- Brehm J., 1993, *The Phantom Respondents: Opinion Surveys and Political Representation*, University of Michigan Press, Ann Arbor.
- 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, Wiley, New York.
- 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., Singer E., Corning A., 2000, *Leverage-saliency theory of survey participation. Description and illustration*, Public Opinion Quarterly, vol. 64, s. 299–308.
- Martin J., Matheson J., 1999, *Responses to declining response rate on government surveys*, Survey Methodology Bulletin, vol. 45, s. 33–37.
- Rodriguez G., 2007, *Generalized Linear Model*, chapter 6, Office of Population Research, Princeton University, <http://data.princeton.edu/wws509/notes/c6.pdf> (4.05.2015).
- Rószkiewicz M., 2015, *Milcząca większość. Uwarunkowania poziomu wskaźnika braku odpowiedzi w środowisku gospodarstw domowych w Polsce w 2013 r. Próba diagnozy*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 385, Taksonomia 24, s. 219–228.
- Rószkiewicz M., Prek-Białas J., Węziak-Białowolska D., Zięba-Pietrzak A., 2013, *Projektowanie badań społeczno-ekonomicznych*, Wydawnictwo Naukowe PWN, Warszawa.
- Schnell R., 1997, *Nonresponse in Bevölkerungsumfragen*, Leske + Budrich, Opladen.
- Steeh C., Kirgies N., Cannon B., De Witt J., 2001, *Are they really as bad as they seem? Nonresponse rate at the end of twentieth century*, Journal of Official Statistics, vol. 17, no. 2, s. 227–247.
- 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.