

# 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](http://www.pracnaukowe.ue.wroc.pl)  
[www.wydawnictwo.ue.wroc.pl](http://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](mailto:econbook@ue.wroc.pl)  
[www.ksiegarnia.ue.wroc.pl](http://www.ksiegarnia.ue.wroc.pl)

Druk i oprawa: TOTEM

## Spis treści

<b>Wstęp</b> .....	9
<b>Jacek Batóg:</b> Identyfikacja obserwacji odstających w analizie skupień / Influence of outliers on results of cluster analysis .....	13
<b>Andrzej Bąk:</b> Porządkowanie liniowe obiektów metodą Hellwiga i TOPSIS – analiza porównawcza / Linear ordering of objects using Hellwig and TOPSIS methods – a comparative analysis.....	22
<b>Grażyna Dehnel:</b> <i>MM</i> -estymacja w badaniu średnich przedsiębiorstw w Polsce / <i>MM</i> -estimation in the medium-sized enterprises survey in Poland.....	32
<b>Andrzej Dudek:</b> <i>Social network analysis</i> jako gałąź wielowymiarowej analizy statystycznej / Social network analysis as a branch of multidimensional statistical analysis.....	42
<b>Iwona Foryś:</b> 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
<b>Gregory Kersten, Ewa Roszkowska, Tomasz Wachowicz:</b> Ocena zgodności porządkowej systemu oceny ofert negocjatora z informacją preferencyjną / Analyzing the ordinal concordance of preferential information and resulting scoring system in negotiations.....	60
<b>Iwona Konarzewska:</b> Rankingi wielokryteriowe a współzależność liniowa kryteriów / Multi-criteria rankings and linear relationships among criteria .....	69
<b>Anna Król, Marta Targaszewska:</b> Zastosowanie klasyfikacji do wyodrębniania homogenicznych grup dóbr w modelowaniu hedonicznym / The application of classification in distinguishing homogeneous groups of goods for hedonic modelling.....	80
<b>Marek Lubicz:</b> Problemy doboru zmiennych objaśniających w klasyfikacji danych medycznych / Feature selection and its impact on classifier effectiveness – case study for medical data.....	89
<b>Aleksandra Łuczak:</b> 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
<b>Iwona Markowicz:</b> Tablice trwania firm w województwie zachodniopomorskim według rodzaju działalności / Companies duration tables in Zachodniopomorskie voivodship by the type of activity .....	108

<b>Małgorzata Markowska, Danuta Strahl:</b> 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
<b>Kamila Migdał-Najman, Krzysztof Najman:</b> Hierarchiczne deglomeracyjne sieci SOM w analizie skupień / The hierarchical divisive SOM in the cluster analysis .....	130
<b>Kamila Migdał-Najman, Krzysztof Najman:</b> Hierarchiczne aglomeracyjne sieci SOM w analizie skupień / The hierarchical agglomerative SOM in the cluster analysis .....	139
<b>Barbara Pawelek, Józef Pocięcha, Jadwiga Kostrzewska, Mateusz Baryła, Artur Lipieta:</b> 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
<b>Wojciech Roszka:</b> Syntetyczne źródła danych w analizie przestrzennego zróżnicowania ubóstwa / Synthetic data sources in spatial poverty analysis.....	157
<b>Małgorzata Rószkiewicz:</b> 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
<b>Adam Sagan, Marcin Pelka:</b> Analiza wielopoziomowa z wykorzystaniem danych symbolicznych / Multilevel analysis with application of symbolic data .....	174
<b>Marcin Salamaga:</b> 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
<b>Agnieszka Stanimir:</b> Pomiar wykluczenia cyfrowego – zagrożenia dla Pokolenia Y / Measurement of the digital divide – risks for Generation Y ...	194
<b>Mirosława Sztemberg-Lewandowska:</b> Grupowanie danych funkcjonalnych w analizie poziomu wiedzy maturzystów / Functional data clustering methods in the analysis of high school graduates' knowledge .....	206
<b>Tadeusz Trzaskalik:</b> Modelowanie preferencji w wielokryterialnych dyskretnych problemach decyzyjnych – przegląd bibliografii / Preference modeling in multi-criteria discrete decision making problems – review of literature .....	214

---

<b>Joanna Trzęsiok:</b> Metody nieparametryczne w badaniu zaufania do instytucji finansowych / Nonparametric methods in the study of confidence in financial institutions .....	226
<b>Hanna Wdowicka:</b> Analiza sytuacji na lokalnych rynkach pracy w Polsce / Local labour market analysis in Poland.....	235
<b>Artur Zaborski:</b> 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”. Następną konferencją 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*

**Joanna Trzęsiok**

Uniwersytet Ekonomiczny w Katowicach  
e-mail: joanna.trzesiok@ue.katowice.pl

---

**METODY NIEPARAMETRYCZNE W BADANIU  
ZAUFIANIA DO INSTYTUCJI FINANSOWYCH**  
**NONPARAMETRIC METHODS IN THE STUDY  
OF CONFIDENCE IN FINANCIAL INSTITUTIONS**

---

DOI: 10.15611/pn.2016.426.23

**Streszczenie:** Artykuł ma charakter aplikacyjny i porusza problem zaufania Polaków do instytucji finansowych – banków komercyjnych, giełdy, otwartych funduszy emerytalnych oraz towarzystw ubezpieczeniowych. Celem pracy było zidentyfikowanie, które czynniki społeczno-demograficzne mają wpływ na poziom badanego zaufania. Ponadto sprawdzono, czy zmiany zaufania, obserwowane w ostatnich latach, były istotne. Badanie przeprowadzono na danych pochodzących z *Diagnozy Społecznej*, a postawione hipotezy, ze względu na niespełnienie założeń metod klasycznych, zweryfikowano za pomocą testów Kruskala-Wallisa oraz Friedmana, będących nieparametrycznymi odpowiednikami jednoczynnikowej analizy wariancji. Wyniki przeprowadzonych analiz pokazują, że istotny wpływ na poziom zaufania do instytucji finansowych mają wiek, wykształcenie, klasa zamieszkiwanej miejscowości oraz status społeczno-zawodowy Polaków. Również obserwowane w latach 2009, 2011 oraz 2013 zmiany zaufania okazały się istotne.

**Słowa kluczowe:** testy nieparametryczne, test Kruskala-Wallisa, test Friedmana, diagnoza społeczna.

**Summary:** The paper presents the analysis of Polish people's confidence in financial institutions, e.g. commercial banks, the stock market, open pension funds and insurance companies. The main goal of the paper is to identify which sociodemographic covariates have an impact on the level of this confidence. Moreover we examine whether there are significant changes in confidence observed in recent years. The research includes data from the Polish Social Diagnosis. Hypotheses are verified using Kruskal-Wallis and Friedman tests, which are nonparametric equivalents to one-way analysis of variance. The results show that age, education, class of the inhabited town and socioprofessional status of Poles are the covariates with a significant impact on the level of confidence in financial institutions. The changes in confidence observed in the years 2009, 2011 and 2013 are also confirmed to be significant.

**Keywords:** nonparametric tests, Kruskal-Wallis test, Friedman test, Polish Social Diagnosis.

## 1. Wstęp

Wyniki przeprowadzonych przez Radę Monitoringu Społecznego (pod kierownictwem profesorów: J. Czapińskiego oraz T. Panka) kolejnych diagnoz społecznych pokazują, że ostatnio wiarygodność instytucji finansowych w Polsce wzrasta, choć nadal jest na zbyt niskim poziomie. W 2013 r. wśród badanych osób zaufanie do np. banków komercyjnych, zakładów ubezpieczeń, funduszy inwestycyjnych czy giełdy miała nie więcej niż 1/3 respondentów. Jest to często tłumaczone kryzysem gospodarczym, który objawia się wyraźnym wzrostem nieufności do instytucji finansowych.

Głównym celem artykułu było zbadanie, czy czynniki demograficzne, takie jak wiek, płeć, wykształcenie, sytuacja zawodowa, miejsce zamieszkania, mają wpływ na poziom zaufania do instytucji finansowych w Polsce. Podjęta została również próba określenia, czy wspomniane zmiany badanego zaufania były istotne w ostatnich latach.

Artykuł ma charakter aplikacyjny. Wszystkie analizy przeprowadzono na zbiorze danych rzeczywistych, pochodzącym z Diagnozy Społecznej. Do badania wykorzystano nieparametryczne testy będące odpowiednikami jednoczynnikowej analizy wariancji, zaś obliczenia wykonano przy pomocy programu SPSS.

## 2. Przedmiot badania

Jak już wspomniano, przedmiotem badania był zbiór danych rzeczywistych, pochodzący z Diagnozy Społecznej [Rada Monitoringu Społecznego 2009, 2011, 2013]. W pierwszym etapie analizy wykorzystano wyniki zebrane w 2013 r. dla respondentów indywidualnych, którzy określili poziom swojego zaufania do instytucji finansowych w Polsce. Zmienne, które wykorzystano w analizie, były związane z pytaniami:

- Czy masz zaufanie do banków komercyjnych? – zmienna  $Z_1$ .
- Czy masz zaufanie do giełdy? –  $Z_2$ .
- Czy masz zaufanie do otwartych funduszy emerytalnych? –  $Z_3$ .
- Czy masz zaufanie do towarzystw ubezpieczeń na życie? –  $Z_4$ .
- Czy masz zaufanie do towarzystw ubezpieczeń majątkowych? –  $Z_5$ .

Badani mieli do wyboru cztery odpowiedzi: „nie”, „tak, umiarkowane”, „tak, duże” oraz „nie mam zdania”. Jednak do analizy wybrano jedynie tych respondentów, którzy jasno określili swoje zaufanie i usunięto ze zbioru niezdecydowanych. Tym samym przyjęto, że każda zmienna  $Z_j$  (dla  $j = 1, \dots, 5$ ) będzie miała trzy kategorie, odpowiednio zakodowane liczbowo: 1 = „nie”, 2 = „tak, umiarkowane”, 3 = „tak, duże”.

Badani respondenci byli również opisywani przez zmienne o charakterze społeczno-demograficznym, które nazwano towarzyszącymi. Były to:

- wiek ( $X_1$ ), którego wartości zostały przydzielone do 6 grup reprezentujących przedziały wiekowe: (0,25), (25, 35), (35, 45), (45, 60), (60, 50), (65,  $\infty$ ),
- płeć ( $X_2$ ) z dwiema kategoriami: „kobieta”, „mężczyzna”,
- wykształcenie ( $X_3$ ) z wyróżnionymi 4 kategoriami: „podstawowe i niższe”, „zasadnicze zawodowe i gimnazjalne”, „średnie”, „policealne i wyższe”,
- klasa zamieszkiwanej miejscowości ( $X_4$ ) z sześcioma możliwymi wariantami: „miasto powyżej 500 tys. mieszkańców”, „miasto 200–500 tys. mieszkańców”, „miasto 100–200 tys. mieszkańców”, „miasto 20–100 tys. mieszkańców”, „miasto poniżej 20 tys. mieszkańców” oraz „wieś”,
- status społeczno-zawodowy ( $X_5$ ), gdzie wyróżniono 9 grup: „pracowników sektora publicznego”, „pracowników sektora prywatnego”, „prywatnych przedsiębiorców”, „rolników”, „rencistów”, „emerytów”, „uczniów i studentów”, „bezrobotnych” oraz „innych biernych zawodowo”.

Po usunięciu obserwacji z brakami wartości zmiennych, ostatecznie w pierwszym etapie analizy wykorzystano odpowiedzi 5728 respondentów.

W drugiej części analizy wykorzystano dane z Diagnozy Społecznej z lat: 2009, 2011 i 2013 [Rada Monitoringu Społecznego 2009, 2011, 2013]. Ze względu na porównywalność wyników, wykorzystano tylko tych respondentów, którzy brali udział w każdej z tych trzech fal badania, co spowodowało zmniejszenie zbioru danych do 950 obserwacji. Konieczne było również zredukowanie zestawu zmiennych, gdyż w kwestionariuszu Diagnozy w 2009 r. występowały nieco inne pytania dotyczące zaufania do instytucji finansowych<sup>1</sup>.

### 3. Przebieg analizy

Przeprowadzone badanie przebiegało w dwóch etapach, które odpowiadały postawionym celom.

#### 3.1. Badanie wpływu wybranych czynników społeczno-demograficznych na poziom zaufania

W pierwszym etapie analizy sprawdzono, czy wymienione czynniki demograficzne (zmiennie towarzyszące) mają istotny wpływ na poziom zaufania respondentów do instytucji finansowych. Do zbadania takiej zależności wykorzystywana jest na ogół jednoczynnikowa analiza wariancji. Jednak ze względu na niespełnione założenie o normalności rozkładu badanej zmiennej w grupach wyznaczonych przez kategorie czynnika – zmienną towarzyszącą  $X_l$  (dla  $l = 1, \dots, 5$ ), zastosowano nieparametryczny odpowiednik tej analizy, czyli test Kruskala-Wallisa.

<sup>1</sup> Więcej informacji na ten temat przedstawiono w części 3.2.

Test Kruskala-Wallisa [Kruskal 1952; Kruskal, Wallis 1952] jest rangowym testem służącym do porównywania rozkładów zmiennej w  $K$  populacjach. Za jego pomocą badamy hipotezy

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_K \quad \text{wobec} \quad H_1 : \bigvee_{i \neq j} \theta_i \neq \theta_j, \quad (1)$$

gdzie:  $\theta_1, \theta_2, \dots, \theta_K$  – średni rangi w populacjach.

Do weryfikacji powyższych hipotez służy statystyka Kruskala-Wallisa

$$T = \frac{12}{N(N+1)} \sum_{k=1}^K n_k \left( \bar{R}_k - \frac{N+1}{2} \right)^2, \quad (2)$$

gdzie:  $\bar{R}_k = \frac{1}{n_k} \sum_{m=1}^{n_k} R_{km}$ ,  $R_{km}$  – ranga elementu  $m$  w całej próbie (o liczebności  $n_k$ ) wylosowanej z populacji o numerze  $k$  (dla  $k = 1, \dots, K$ ),

$$N = \sum_{k=1}^K n_k.$$

Statystyka  $T$  ma rozkład chi-kwadrat z  $K - 1$  stopniami swobody.

Przeprowadzona procedura badawcza przebiegała w następujących krokach.

1. Utworzono zmienną sumaryczną  $Y$ , której wartości  $y_i$  (dla  $i = 1, \dots, 5278$ ) były uśrednionymi realizacjami zmiennych  $Z_j$  (dla  $j = 1, \dots, 5$ )

$$y_i = \frac{1}{5} \sum_{j=1}^5 z_{ij}. \quad (3)$$

2. Zbadano rzetelność utworzonej zmiennej sumarycznej  $Y$  za pomocą  $\alpha$ -Cronbacha i otrzymano

$$\alpha = 0,826.$$

Wysoka wartość  $\alpha$  wskazuje na rzetelność skali.

3. Ze względu na niespełnienie założeń klasycznej analizy wariancji, wykorzystano test Kruskala-Wallisa, by sprawdzić, które ze zmiennych towarzyszących  $X_l$  ( $l = 1, \dots, 5$ ) (wiek, płeć, wykształcenie, klasa zamieszkiwanej miejscowości oraz status społeczno-ekonomiczny) mają istotny wpływ na rozkład zmiennej sumarycznej  $Y$ . Zatem dla każdej ze zmiennych  $X_l$  wykonywano analizę, sprawdzając, czy obserwowane są istotne różnice w poziomie zaufania (wyrażonego za pomocą zmiennej  $Y$ ) w grupach wyznaczonych przez kategorie każdego z czynników  $X_l$  z osobna.

Otrzymane wyniki przedstawiono w tab. 1.

**Tabela 1.** Wyniki testu Kruskala-Wallisa dla wszystkich zmiennych towarzyszących

Czynnik	Wartość statystyki Kruskala-Wallisa	Liczba stopni swobody	Prawdopodobieństwo testowe
Wiek	114,5	5	mniejsze od 0,001
Płeć	0,09	1	0,76
Wykształcenie	39,3	3	mniejsze od 0,001
Klasa miejscowości	64,4	5	mniejsze od 0,001
Status społeczno-ekonom.	164,1	8	mniejsze od 0,001

Źródło: opracowanie własne.

Na podstawie wartości prawdopodobieństwa testowego  $p$  (tab. 1) można powiedzieć, że wiek, wykształcenie, klasa zamieszkiwanej miejscowości oraz status społeczno-zawodowy mają istotny wpływ na poziom zaufania do instytucji finansowych<sup>2</sup>. Nie zaobserwowano natomiast różnicy w zaufaniu kobiet i mężczyzn do tych instytucji.

W celu uzupełnienia analizy wykonano testy *post-hoc*. W tabeli 2 wypisano wszystkie te grupy (wyznaczone przez kategorie poszczególnych czynników  $X_l$ , gdzie  $l = 1, 3, 4, 5$ ) dla których obserwujemy istotne różnice w poziomie badanego zaufania.

Na podstawie informacji zamieszczonych w tab. 2 można wyciągnąć następujące wnioski.

#### 1. Dla zmiennej wiek

- zaufanie do instytucji finansowych w grupie osób powyżej 65. roku życia jest istotnie niższe niż w pozostałych grupach wiekowych,
- zaufanie wśród osób od 60 do 65 lat jest istotnie niższe niż wśród osób od 25 do 45 lat,
- również wśród osób z przedziału wiekowego 45–60 lat zaufanie jest istotnie niższe niż wśród osób od 25 do 35 lat.

#### 2. Dla zmiennej wykształcenie

- obserwujemy istotnie wyższe zaufanie do instytucji finansowych w grupie osób z wykształceniem policealnym i wyższym niż w pozostałych grupach,
- ponadto zaufanie wśród osób z wykształceniem średnim jest istotnie wyższe niż dla osób z wykształceniem podstawowym i niższym.

#### 3. Dla zmiennej klasa zamieszkiwanej miejscowości

- zaufanie wśród osób mieszkających na wsi jest istotnie niższe niż wśród osób mieszkających w miastach o liczbie mieszkańców powyżej 100 tys.,
- zaufanie do instytucji finansowych mieszkańców miast do 100 tys. osób jest istotnie niższe niż wśród mieszkańców miast powyżej 500 tys. osób.

<sup>2</sup> W każdym z tych przypadków otrzymano prawdopodobieństwo testowe  $p < 0,05$ , co oznacza, że należy odrzucić hipotezę o równości średnich rang zmiennej  $Y$  w grupach, więc każdy z tych czynników ma istotny wpływ na rozkład  $Y$ .

**Tabela 2.** Zestawienie kategorii zmiennych towarzyszących dla których zaobserwowano istotne różnice w poziomie badanego zaufania

Wiek			Status społeczno-ekonomiczny		
Porównywane kategorie	Wartość statystyki testowej	Skorygowane $p^*$	Porównywane kategorie	Wartość statystyki testowej	Skorygowane $p^*$
$(65, \infty)$ i $(60, 65)$	314,0	0,005	renciści i sektor prywatny	666,7	mniejsze od 0,001
$(65, \infty)$ i $(45, 60)$	432,1	mniejsze od 0,001	renciści i sektor publiczny	736,8	mniejsze od 0,001
$(65, \infty)$ i $(35, 45)$	588,6	Mniejsze od 0,001	renciści i prywatni przedsiębiorcy	814,9	mniejsze od 0,001
$(65, \infty)$ i $(25, 35)$	754,4	mniejsze od 0,001	rolnicy i sektor prywatny	512,8	mniejsze od 0,001
$(65, \infty)$ i $(0, 25)$	545,6	mniejsze od 0,001	rolnicy i sektor publiczny	583,0	mniejsze od 0,001
$(60, 65)$ i $(35, 45)$	274,6	0,016	rolnicy i prywatni przedsiębiorcy	661,1	mniejsze od 0,001
$(60, 65)$ i $(25, 35)$	440,4	mniejsze od 0,001	emeryci i sektor prywatny	505,9	mniejsze od 0,001
$(45, 60)$ i $(25, 35)$	322,4	mniejsze od 0,001	emeryci i sektor publiczny	576,0	mniejsze od 0,001
			emeryci i prywatni przedsiębiorcy	654,2	mniejsze od 0,001
			bezrobotni i sektor prywatny	435,7	mniejsze od 0,001
			bezrobotni i sektor publiczny	505,8	mniejsze od 0,001
			bezrobotni i prywatni przedsiębiorcy	584,0	mniejsze od 0,001
			bierni zawodowo i sektor publiczny	359,7	0,043
			bierni zawodowo i prywatni przedsiębiorcy	437,9	0,024
Wykształcenie			Klasa zamieszkiwanej miejscowości		
Porównywane kategorie	Wartość statystyki testowej	Skorygowane $p^*$	Porównywane kategorie	Wartość statystyki testowej	Skorygowane $p^*$
Podstawowe/niższe i średnie	191,4	0,025	wieś i miasto 100–200 tys.	261,7	0,043
Podstawowe/niższe i wyższe	377,5	mniejsze od 0,001	wieś i miasto 200–500 tys.	312,0	0,001
Zawodowe/gimnazjalne i wyższe	304,3	Mniejsze od 0,001	wieś i miasto powyżej 500 tys.	549,2	mniejsze od 0,001
Średnie i wyższe	186,1	0,008	miasto do 20 tys. i miasto powyżej 500 tys.	510,7	mniejsze od 0,001
			miasto 20–100 tys. i miasto powyżej 500 tys.	390,2	mniejsze od 0,001

\*  $p$  to prawdopodobieństwo testowe.

Źródło: opracowanie własne.

4. Dla zmiennej status społeczno-ekonomiczny

- renciści, rolnicy, emeryci, bezrobotni oraz inni bierni zawodowo charakteryzują się istotnie niższym zaufaniem do instytucji finansowych niż pracownicy sektorów: prywatnego i publicznego oraz prywatni przedsiębiorcy.

### 3.2. Badanie istotności zmian zaufania w czasie

W drugim etapie analizy sprawdzono, czy zaobserwowane w ostatnich latach zmiany zaufania do instytucji finansowych były istotne. Do badania zastosowano nieparametryczny test Friedmana, gdyż, ponownie ze względu na niespełnienie wymaganych założeń, niemożliwe było zastosowanie klasycznej jednoczynnikowej analizy wariancji w schemacie wewnątrzgrupowym (powtarzanych pomiarów).

Rangowy test Friedmana [Friedman 1937] stosuje się w sytuacji, gdy pomiary badanej zmiennej wykonywane są kilkukrotnie w różnych momentach czasu i różnych warunkach. Testuje się więc hipotezy

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_T \quad \text{wobec} \quad H_1 : \bigvee_{i \neq j} \theta_i \neq \theta_j, \quad (4)$$

gdzie:  $\theta_1, \theta_2, \theta_T$  – średnie rangi badanej zmiennej, w kolejnych pomiarach tej samej populacji.

Do weryfikacji powyższych hipotez służy statystyka Friedmana

$$T = \frac{12}{NT(T+1)} \sum_{t=1}^T \left( \sum_{i=1}^N R_{it} \right)^2 - 3N(T+1), \quad (5)$$

gdzie:  $R_{it}$  – rangi przypisane kolejnym pomiarom  $t$  (dla  $t = 1, \dots, T$ ) dla każdego obiektu o numerze  $i$  (gdzie  $i = 1, \dots, N$ ),  $N$  – liczebność próby.

Statystyka  $T$  ma rozkład chi-kwadrat z  $T - 1$  stopniami swobody.

Jak już wspomniano, w tej części badania wykorzystano zredukowany zbiór danych, w którym uwzględniono odpowiedzi tylko 950 respondentów, którzy określili swoje zaufanie do instytucji finansowych<sup>3</sup> i jednocześnie brali udział w trzech falach Diagnozy – w latach 2009, 2011 i 2013. W ciągu tych sześciu lat zmienił się również kwestionariusz Diagnozy, więc analizę przeprowadzono dla zmiennych  $Z_1$  oraz  $Z_3$ , odpowiadającym pytaniom:

- Czy masz zaufanie do banków komercyjnych?
- Czy masz zaufanie do funduszy emerytalnych?

Wśród badanych respondentów zaufanie do banków komercyjnych w 2009 r. miało 52,7% osób, w 2011 r. – tylko 33,6%, natomiast w 2013 r. – 52,5%. Zaufa-

<sup>3</sup> W badaniu nie brano pod uwagę odpowiedzi respondentów, którzy na zadane pytania odpowiedzieli „nie mam zdania”.



nie do otwartych funduszy emerytalnych deklarowało: w 2009 r. – 16,8% respondentów, w 2011 r. – 18,2%, zaś w 2013 r. – 25,9%.

Zbadano, za pomocą testu Friedmana, czy poziom zaufania do banków komercyjnych oraz funduszy emerytalnych zmieniał się istotnie w latach 2009, 2011 i 2013. Wyniki przedstawiono w tab. 3. Ponadto analizę uzupełniono, wykonując testy *post-hoc*, których wartości zamieszczono w tab. 4.

**Tabela 3.** Wyniki testu Friedmana dla wybranych zmiennych

Zmienna	Wartość testu Friedmana	Liczba stopni swobody	Prawdopodobieństwo testowe
Zaufanie do banków komercyjnych	124,1	2	mniejsze od 0,001
Zaufanie do funduszy emerytalnych	36,6	2	mniejsze od 0,001

Źródło: opracowanie własne.

Wartości prawdopodobieństwa testowego, zaprezentowane w tab. 3, pokazują, że w latach 2009–2013 zaufanie zarówno do banków komercyjnych, jak i do funduszy emerytalnych zmieniło się istotnie.

**Tabela 4.** Wyniki testów *post-hoc* dla porównań poziomu zaufania w kolejnych latach

Zaufanie do banków komercyjnych			Zaufanie do OFE		
porównanie dla lat	statystyka testowa	skorygowane $p^*$	porównanie dla lat	statystyka testowa	skorygowane $p^*$
2009 i 2011	0,287	mniejsze od 0,001	2009 i 2011	-0,021	1
2009 i 2013	0,003	1	2009 i 2013	-0,136	0,009
2011 i 2013	-0,284	mniejsze od 0,001	2011 i 2013	-0,115	0,036

\*  $p$  to prawdopodobieństwo testowe.

Źródło: opracowanie własne.

Na podstawie wyników z tab. 4 można dodatkowo wyciągnąć poniższe wnioski.

1. Zaobserwowano, że zaufanie do banków komercyjnych w 2013 r. było istotnie wyższe niż w 2011 r., a w 2011 r. – istotnie niższe niż w 2009 r. Natomiast różnice w poziomie zaufania w latach 2013 i 2009 były nieistotne.

2. Zaufanie do otwartych funduszy emerytalnych w 2013 r. było istotnie wyższe niż w latach 2011, czy 2009. Ponadto poziomy zaufania do OFE w latach 2011 i 2009 nie różniły się istotnie.

## 4. Zakończenie

W artykule badano zaufanie Polaków do banków komercyjnych, giełdy, otwartych funduszy emerytalnych i towarzystw ubezpieczeniowych. Analizy przeprowadzono za pomocą testów Kruskala-Wallisa oraz Friedmana, będących nieparametrycznymi odpowiednikami klasycznej analizy wariancji.

Otrzymane wyniki pokazują, że istotny wpływ na poziom zaufania do wymienionych instytucji finansowych mają: wiek, wykształcenie, klasa zamieszkiwanej miejscowości oraz status społeczno-ekonomiczny.

Natomiast wyniki przeprowadzonych testów *post-hoc* są zgodne z „intuicją” i wskazują, że niższe zaufanie w stosunku do instytucji finansowych wiąże się z wyższym wiekiem, niższym wykształceniem, miejscem zamieszkania o mniejszej liczbie mieszkańców oraz z brakiem aktywności zawodowej.

Drugi etap badań pokazuje, że analizowane zaufanie w 2013 r. istotnie wzrosło w porównaniu z poprzednimi latami.

## Literatura

- Friedman M., 1937, *The use of ranks to avoid the assumption of normality implicit in the analysis of variance*, Journal of the American Statistical Association, vol. 32, no. 200, s. 675–701.
- Kruskal W., 1952, *A nonparametric test for the several sample problem*, Annals of Mathematical Statistics, vol. 23, no. 4, s. 525–540.
- Kruskal W., Wallis W., 1952, *Use of ranks in one-criterion variance analysis*, Journal of the American Statistical Association, vol. 47, no. 260, s. 583–621.
- Rada Monitoringu Społecznego, 2009, 2011, 2013, *Diagnoza społeczna: zintegrowana baza danych*, [www.diagnoza.com](http://www.diagnoza.com) (1.08.2015).