

Cena 15,00 zł
(VAT 8%)

Indeks 381306
e-ISSN 2543-8476
PL ISSN 0043-518X

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

MAJ / MAY
ROCZNIK / VOLUME 69

2024 | 5

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

POLSKIE TOWARZYSTWO STATYSTYCZNE
POLISH STATISTICAL ASSOCIATION



WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

MAJ / MAY
ROCZNIK / VOLUME 69

2024 | 5 (756)

RADA NAUKOWA / SCIENTIFIC COUNCIL

dr Dominik Rozkrut – przewodniczący/chairman (Uniwersytet Szczeciński, Polska), Prof. Anthony Arundel (Maastricht University, Holandia), Eric Bartelsman, PhD, Assoc. Prof. (Vrije Universiteit Amsterdam, Holandia), prof. dr hab. Czesław Domański (Uniwersytet Łódzki, Polska), prof. dr hab. Elżbieta Gołata (Uniwersytet Ekonomiczny w Poznaniu, Polska), Semen Matkovskyy, PhD, Assoc. Prof. (Ivan Franko National University of Lviv, Ukraina), prof. dr hab. Włodzimierz Okrasa (Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie, Polska), prof. dr hab. Józef Oleński (Polskie Towarzystwo Statystyczne, Polska), prof. dr hab. Tomasz Panek (Szkola Główna Handlowa w Warszawie, Polska), Juan Manuel Rodríguez Poo, PhD, Assoc. Prof. (University of Cantabria, Hiszpania), Iveta Stankovičová, BEng, PhD, Assoc. Prof. (Comenius University in Bratislava, Słowacja), prof. dr hab. Marek Walesiak (Uniwersytet Ekonomiczny we Wrocławiu, Polska), prof. dr hab. Józef Zegar (Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowy Instytut Badawczy, Polska)

sekretarz/secretary: Paulina Kucharska-Singh, Główny Urząd Statystyczny, Polska

KOLEGIUM REDAKCYJNE / EDITORIAL BOARD

Tudorel Andrei, PhD, Assoc. Prof. (Bucharest Academy of Economic Studies, Rumunia), mgr Renata Bielak (Główny Urząd Statystyczny, Polska), dr hab. Marek Cierpień-Wolan, prof. UR (Uniwersytet Rzeszowski, Polska), dr hab. Grażyna Dehnel, prof. UEP (Uniwersytet Ekonomiczny w Poznaniu, Polska), dr Jacek Kowalewski (Uniwersytet Ekonomiczny w Poznaniu, Polska), dr Jan Kubacki (Polskie Towarzystwo Statystyczne, Polska), dr Grażyna Marciniak (Główny Urząd Statystyczny, Polska), dr hab. Andrzej Młodak, prof. Akademii Kaliskiej (Akademia Kaliska im. Prezydenta Stanisława Wojciechowskiego, Polska), prof. dr hab. Mateusz Pipień (Uniwersytet Ekonomiczny w Krakowie, Polska), Marek Rojíček, BEng, PhD (University of Economics, Prague, Czechy), Anna Shostya, PhD, Assoc. Prof. (Pace University in New York, Stany Zjednoczone), dr hab. Małgorzata Tarczyńska-Łuniewska, prof. US (Uniwersytet Szczeciński, Polska), dr Wioletta Wrzaszcz (Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowy Instytut Badawczy, Polska), dr inż. Agnieszka Zgierska (Główny Urząd Statystyczny, Polska)

ZESPÓŁ REDAKCYJNY / EDITORIAL STAFF

redaktor naczelny / editor-in-chief: Marek Cierpień-Wolan

zastępca redaktora naczelnego / deputy editor-in-chief: Andrzej Młodak

redaktorzy tematyczni / thematic editors: Andrzej Młodak, Małgorzata Tarczyńska-Łuniewska, Wioletta Wrzaszcz, Agnieszka Zgierska

sekretarz/secretary: Małgorzata Zygmunt, Główny Urząd Statystyczny, Polska

ADRES REDAKCJI / EDITORIAL OFFICE ADDRESS

Główny Urząd Statystyczny / Statistics Poland, al. Niepodległości 208, 00-925 Warszawa
tel./phone +48 22 608 32 25, e-mail: redakcja.ws@stat.gov.pl

Redakcja językowa: Wydział Czasopism Naukowych, Główny Urząd Statystyczny
Language editing: Scientific Journals Division, Statistics Poland

Redakcja techniczna, skład i łamanie, opracowanie materiałów graficznych, korekta, druk i oprawa:
Zakład Wydawnictw Statystycznych – zespół pod kierunkiem Macieja Adamowicza
Technical editing, typesetting, preparation of graphic materials, proofreading, printing and binding:
Statistical Publishing Establishment – team supervised by Maciej Adamowicz



Zakład Wydawnictw
Statystycznych

Druk i oprawa / Printed and bound by:

Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
al. Niepodległości 208, 00-925 Warszawa, zws.stat.gov.pl

Wersja elektroniczna, stanowiąca wersję pierwotną czasopisma, jest dostępna na ws.stat.gov.pl
The primary version of the journal, issued in electronic form, is available at ws.stat.gov.pl

© Copyright by Główny Urząd Statystyczny and the authors, some rights reserved. CC BY-SA 4.0 licence



Informacje w sprawie sprzedaży i prenumeraty czasopisma / Sales and subscription of the journal:
Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
zws.stat.gov.pl
tel./phone +48 22 608 32 10, +48 22 608 38 10

SPIS TREŚCI
CONTENTS

Od redakcji	IV
From the editorial team	
Studia metodologiczne	
Methodological studies	
Artur Zaborski	
The optimal number of tetrads for measurement of dissimilarities in nonmetric multidimensional scaling	1
Optymalna liczba tetrad do pomiaru niepodobieństw w niemetrycznym skalowaniu wielowymiarowym	
Ryszard Szupiluk, Paweł Rubach	
Log-Quad divergence for Non-negative Matrix Factorization in multi-model prediction	14
Dywergencja Log-Quad dla nieujemnej faktoryzacji macierzy w predykcji za pomocą podejścia wielomodelowego	
Statystyka w praktyce	
Statistics in practice	
Błażej Kochański, Jakub Sochacki	
Nadmiarowe zgony podczas pandemii COVID-19 w Polsce i ocena skuteczności szczepień	25
Excess mortality during the COVID-19 pandemic in Poland and the effectiveness of vaccinating the population	
Dyskusje. Recenzje. Informacje	
Discussions. Reviews. Information	
Elżbieta Zalewska, Kamila Trzcńska	
41st International Conference MSA'2023	45
Dorota Kierska	
Nowości wydawnicze w zbiorach Centralnej Biblioteki Statystycznej	52
New publications in the Central Statistical Library resources	
Joanna Sadowy	
Wydawnictwa GUS. Kwiecień 2024	55
Publications of Statistics Poland. April 2024	
Dla autorów	57
For the authors	
Działy „WS” – tematyka artykułów	68
WS sections – topics of the articles	

OD REDAKCJI

W majowym numerze „Wiadomości Statystycznych. The Polish Statistician” proponujemy Państwu lekturę dwóch artykułów poświęconych metodologii badań statystycznych i pracy o charakterze empirycznym.

Dr hab. Artur Zaborski, prof. UEW, w artykule *The optimal number of tetrads for measurement of dissimilarities in nonmetric multidimensional scaling* eksponuje atuty metody tetrad, stosowanej do wyznaczania niepodobieństw w niemetrycznym skalowaniu wielowymiarowym, np. w badaniach preferencji. Metoda ta polega na wskazywaniu przez badane osoby par najbardziej i najmniej podobnych elementów w czteroelementowych podzbiorach (tetradach) z n -elementowego zbioru obiektów. Autor tworzy cztery macierze odległości dla 7, 9, 11 i 13 obiektów (losowo wybranych miast wojewódzkich) i, posługując się analizą Prokrustesza oraz korelacją rang Spearmana, wykazuje, że metoda tetrad daje korzystne wyniki już wtedy, gdy każda para obiektów pojawia się w zbiorze tetrad tylko raz.

Dr hab. Ryszard Szupiluk, prof. SGH, i dr Paweł Rubach w pracy *Log-Quad divergence for Non-negative Matrix Factorization in multi-model prediction* przedstawiają nowy algorytm nieujemnej faktoryzacji macierzy oparty na dywergencji Log-Quad. Wskazują możliwości wykorzystania go do separacji ukrytych komponentów destrukcyjnych zawartych w predykcjach w ujęciu wielomodelowym na przykładzie prognozowania zużycia energii elektrycznej w Polsce. Porównują prezentowany algorytm z innymi technikami z obszaru ślepej separacji sygnałów i zauważają, że charakteryzuje się on właściwością poprawiania predykcji dla niedużych wolumenów danych.

Nadmiarowe zgony podczas pandemii COVID-19 w Polsce i ocena skuteczności szczepień to temat artykułu dr. Błażeja Kochańskiego i mgr. Jakuba Sochackiego. Autorzy mierzą skalę pandemii w Polsce za pomocą liczby nadmiarowych zgonów w podregionach według klasyfikacji NUTS 3 i w grupach wieku, a następnie określają zależność pomiędzy regionalnym zróżnicowaniem względnej nadwyżki zgonów podczas czwartej fali pandemii a stopniem zaszczepienia populacji. W analizach wykorzystują metodę Karlinsky’ego i Kobaka. Stwierdzają, że odsetek nadmiarowych zgonów w trakcie czwartej fali pandemii był ujemnie i dość silnie skorelowany ze stopniem zaszczepienia, a problem nadmiarowych zgonów dotyczył przede wszystkim osób w wieku 60 lat i więcej.

Ponadto w numerze zamieszczamy sprawozdanie z 41st International Conference MSA’2023, która odbyła się w dniach 6–8 listopada 2023 r. na Uniwersytecie Łódzkim. Dr Elżbieta Zalewska i dr Kamila Trzcńska relacjonują przebieg konferencji i przedstawiają tematy wygłoszonych referatów.

Zachęcamy także do zapoznania się z nowościami wydawniczymi, które w ostatnim okresie wzbogaciły zbiory Centralnej Biblioteki Statystycznej, i najnowszymi publikacjami GUS. Ich omówienie znajdą Państwo w końcowej części numeru.

Życzymy miłej lektury.

FROM THE EDITORIAL TEAM

The May issue of *Wiadomości Statystyczne. The Polish Statistician* features two articles on the methodology of statistical research and a study of an empirical character.

Artur Zaborski, PhD, DSc, professor at the Wrocław University of Economics and Business, in his article entitled *The optimal number of tetrads for measurement of dissimilarities in nonmetric multidimensional scaling*, presents the advantages of the tetrad method, which is applied to determine dissimilarities in nonmetric multidimensional scaling, e.g. in preference studies. The approach consists in respondents indicating pairs of the most and the least similar elements in four-element subsets (tetrads) from a set of n objects. The author creates four distance matrices for 7, 9, 11 and 13 objects (randomly selected voivodship cities in Poland) and, using Procrustes analysis and Spearman's rank correlation, he shows that the tetrad method produces favourable results already at the point when each pair of objects appears in the set of tetrads only once.

Ryszard Szupiluk, PhD, DSc, professor at the SGH Warsaw School of Economics, and Paweł Rubach, PhD, in their paper *Log-Quad divergence for Non-negative Matrix Factorization in multi-model prediction*, present a new algorithm for Non-negative Matrix Factorization based on Log-Quad divergence. The authors show the possibility of using it to separate latent destructive components contained in predictions in a multi-model approach, using forecasting electricity consumption in Poland as an example. They compare the presented algorithm with other techniques from the area of blind signal separation and find that it is able to improve predictions for small volumes of data.

Excess mortality during the COVID-19 pandemic in Poland and the effectiveness of vaccinating the population is the subject of the article by Błażej Kochański, PhD, and Jakub Sochacki, MSc. The authors measure the magnitude of the pandemic in Poland on the basis of the number of excess deaths in NUTS 3 subregions and by age group. They then determine the correlation between the regional variation in relative excess mortality during the fourth wave of the pandemic and the vaccination rate of the population. They use the Karlinsky and Kobak method in their analyses. They conclude that the percentage of excess deaths recorded during the fourth wave of the pandemic was negatively and quite strongly correlated with the vaccination rate of the population, and that the problem of excess mortality appeared primarily among people aged 60 and over.

In addition, this issue includes a report on the 41st International Conference MSA'2023, held on 6–8 November 2023 at the University of Łódź. Elżbieta Zalewska, PhD, and Kamila Trzcińska, PhD, provide information about the conference proceedings and present the topics of the delivered papers.

The May issue of *Wiadomości Statystyczne. The Polish Statistician* closes with a presentation of and discussion on the new releases that have recently been added to the resources of the Central Statistical Library and the latest publications of Statistics Poland.

We wish you a pleasant reading.

The optimal number of tetrads for measurement of dissimilarities in nonmetric multidimensional scaling¹

Artur Zaborski^a

Abstract. The direct determination of dissimilarities is the most popular and most frequently used method for raising input data in nonmetric multidimensional scaling, i.e. when variables are measured on an ordinal scale (e.g. in preference studies). Most methods for the direct measurement of similarities, including ranking, sorting, pairwise comparison, conditional ranking of similarities are, however, very laborious, especially when a large number of objects is tested. Thus, the research described in this article is based on the tetrad method, which is uncomplicated and less burdensome for the respondents.

In the proposed method, respondents are asked to evaluate four-element subsets (tetrads) from a set of n objects. The respondent is asked to indicate the pair with the most and the least similar elements in each tetrad. As the number of tetrads rapidly increases along with the number of objects, it becomes necessary to use the incomplete variant of the method, in which only some four-element subsets are presented to the respondents for evaluation.

The aim of the research presented in the article is to determine the size of the tetrad set that is sufficient to create a dissimilarity matrix used to perform nonmetric multidimensional scaling. The study was based on four distance matrices for 7, 9, 11 and 13 objects that were randomly selected voivodship capitals in Poland. The distances between the cities were expressed in kilometres. The Procrustes analysis and Spearman's rank correlation were used in the study. The findings show that the use of the tetrad method for the measurement of dissimilarities produces beneficial results already at the point when each pair of objects appears in the set of tetrads only once, which allows the number of opinions provided by the respondents to be significantly reduced.

Keywords: measurement of dissimilarities, direct indication of dissimilarities, tetrad method, nonmetric multidimensional scaling, Procrustes analysis

JEL: C38, C63, M31

Optymalna liczba tetrad do pomiaru niepodobieństw w niemetrycznym skalowaniu wielowymiarowym

Streszczenie. Bezpośrednie wyznaczanie niepodobieństw jest najpopularniejszym i najczęściej stosowanym sposobem uzyskiwania danych wejściowych w niemetrycznym skalowaniu wielowymiarowym, czyli gdy zmienne są mierzone na skali porządkowej (np. w badaniach preferencji).

¹ Artykuł został opracowany na podstawie referatu wygłoszonego na konferencji Multivariate Statistical Analysis MSA'2022, która odbyła się w dniach 7–9 listopada 2022 r. w Łodzi. / The article is based on a paper delivered at the Multivariate Statistical Analysis MSA'2022 Conference, held on 7–9 November 2022 in Łódź, Poland.

^a Uniwersytet Ekonomiczny we Wrocławiu, Wydział Ekonomii i Finansów, Polska / Wrocław University of Economics and Business, Faculty of Economics and Finance, Poland.

ORCID: <https://orcid.org/0000-0003-1374-2268>. E-mail: artur.zaborski@ue.wroc.pl.

Stosowanie większości metod bezpośredniego pomiaru podobieństwa, takich jak rangowanie, sortowanie, porównania parami czy warunkowe porządkowanie podobieństw, jest jednak bardzo pracochłonna, zwłaszcza przy dużej liczbie obiektów. Z tego powodu w badaniu omawianym w niniejszym artykule posłużono się metodą tetrad – nieskomplikowaną i nieuciążliwą dla respondentów.

Proponowana metoda polega na przedstawieniu respondentom do oceny czteroelementowe podzbiory (tetrazy) z n -elementowego zbioru obiektów. Respondent proszony jest o wskazanie pary najbardziej i najmniej podobnych elementów w każdej tetradzie. Ponieważ liczba tetrad rośnie bardzo szybko wraz z liczbą obiektów, niezbędne jest zastosowanie niepełnego wariantu tej metody, w której respondentom przedstawia się do oceny jedynie część czteroelementowych podzbiorów.

Badanie omawiane w artykule ma na celu ustalenie wielkości zbioru tetrad, która jest wystarczająca do wyznaczenia macierzy niepodobieństw służącej do wykonania niemetrycznego skalowania wielowymiarowego. Badanie przeprowadzono na podstawie czterech macierzy odległości dla 7, 9, 11 i 13 obiektów, którymi były losowo wybrane miasta wojewódzkie w Polsce. Odległości między miastami wyrażono w kilometrach. Wykorzystano analizę Prokrustesza oraz współczynnik korelacji rang Spearmana. Wykazano, że zastosowanie metody tetrad do pomiaru niepodobieństw daje korzystne wyniki już wtedy, gdy każda para obiektów pojawia się w zbiorze tetrad tylko raz, co pozwala na znaczne ograniczenie liczby opinii wyrażanych przez respondentów.

Słowa kluczowe: pomiar niepodobieństw, bezpośrednie wyznaczanie niepodobieństw, metoda tetrad, niemetryczne skalowanie wielowymiarowe, analiza Prokrustesza

1. Introduction

Nonmetric multidimensional scaling is used when variables are measured on an ordinal scale. The development of nonmetric methods of classical scaling took place in the 1960s. The first algorithm was proposed by Shepard (1962). Shepard presented, for the first time, an iterative computer procedure of the proximity matrix analysis, in which the distances between points were measured on an ordinal scale. The following researchers also contributed extensively to the development of multidimensional scaling methodology in their studies: Carroll and Chang (1970), Guttman (1968), Kruskal (1964a, 1964b), de Leeuw and Heiser (1980), Takane et al. (1977), and others. The goal of nonmetric multidimensional scaling is to find such a mapping φ of a set of objects $O = (O_1, O_2, \dots, O_n)$, e.g. people, products, companies, brands, advertisements, etc., with dissimilarities δ_{ij} ($i, j = 1, 2, \dots, n$) into a set of points in an r -dimensional space (r is usually equal to 2 or 3) with d_{ij} distances between them, so that $\hat{d}_{ij} \approx d_{ij}$, where \hat{d}_{ij} is a monotone regression of d_{ij} on δ_{ij} (see e.g. Borg & Groenen, 2005, p. 34; Kruskal, 1964a, 1964b). In nonmetric multidimensional scaling, dissimilarities are measured on an ordinal scale, so \hat{d}_{ij} must satisfy the following condition:

$$\delta_{ij} < \delta_{i'j'} \Rightarrow \hat{d}_{ij} \leq \hat{d}_{i'j'}. \quad (1)$$

The starting point in multidimensional scaling is the creation of a dissimilarity matrix. There are two ways of obtaining input dissimilarities in multidimensional scaling: when they are directly obtained from empirical subjective measurements of objects performed by subjects, they are called direct dissimilarities; by contrast, when they are not obtained from subjects, but calculated from a data matrix associated with these objects, they are labeled as derived dissimilarities. This article focuses only on direct dissimilarities.

So far, many more or less popular methods used for direct similarity measurement have been developed, e.g. sorting, pair comparisons, ranking, ranking of pairs, conditional ranking, and others. The differences in the application of various measurement methods may depend on the number of objects simultaneously presented to the respondents, the difficulty in assessing similarities and the total number of required ratings (see e.g. Bijmolt, 1996). When the number of objects is high, the number of direct assessments made by respondents becomes too large, and makes the task of dissimilarities measurement more difficult.

In view of the fact that the results of multidimensional scaling based on different collecting methods are similar (see e.g. Bijmolt, 1996; Humphreys, 1982; Zaborski, 2003), the choice of the method of measurement should be guided primarily by two criteria: the method should not be labour-intensive and expressing judgements should not be problematic for the respondents.

A method that largely meets these two criteria is the incomplete method of tetrads, proposed by Zaborski (2020). The idea is based on the method of triads (Burton & Nerlove, 1976) and on the theory of balanced incomplete block designs (see e.g. Burton & Nerlove, 1976; Morris, 2010; Rink, 1987). A preliminary comparative analysis of the method of triads and the method of tetrads for nine objects (Zaborski, 2022) showed that obtaining comparable results when using the method of triads requires over three times more respondents' assessments than in the case of the method of tetrads.

The purpose of the article is to determine the size of the tetrad set that is sufficient to create a dissimilarity matrix and to perform nonmetric multidimensional scaling.

2. Measurement of dissimilarities in nonmetric multidimensional scaling

There are three main approaches to collecting dissimilarities in nonmetric multidimensional scaling (Tsogo et al., 2000). The first approach is based on rankings and similarities ratings of pairs of objects. The second group of methods uses grouping and sorting tasks in order to calculate similarities. Finally, the third approach consists of pairwise comparisons of similarities.

The above-mentioned selection of methods tend to evoke subjective feelings of the respondents, i.e. fatigue, weariness resulting from making numerous assessments, or experiencing difficulties in expressing differences between objects. As a result, the collected data may be incomplete or the assessments made do not always fully reflect the opinions of the respondents.

Many studies, including those of Bijmolt (1996) and Humphreys (1982), have considered this problem. Although the research above did not confirm any significant influence of the measurement of dissimilarities on the multidimensional scaling results, other aspects differentiating the methods were identified. It is much more difficult to determine dissimilarities for an entire set of objects than for two or three of them, especially when some objects are only slightly different from each other. With a large number of objects, the respondents focus on extreme assessments, i.e. they notice the objects that are the most similar, the least similar, while ignoring the other objects in their assessments. As a result, the respondents' assessments are incomplete or often random. In turn, methods in which the respondents are presented with only two or three options require providing multiple answers, which results in fatigue and weariness. In such cases, incomplete analyses may offer a solution. A review of the research related to incomplete methods of measuring dissimilarities in multivariate scaling was presented by e.g. Tsogo et al. (2000).

3. The method of tetrads under incomplete block design

3.1. The methodological basis of the method of tetrads

In the complete method of tetrads (Zaborski, 2020, 2022), the subject is asked to consider all the possible groups of four objects (O_i, O_j, O_k, O_l) , $i, j, k, l = 1, 2, \dots, n$, where $i \neq j \neq k \neq l \neq i \neq k$ and $j \neq l$, taken from the full set of n objects $O = (O_1, O_2, \dots, O_n)$. The respondent has to indicate the most similar pair and the least similar pair. On this basis the tetrad is formed, where the most similar objects are placed as the first and the second, and the least similar as the first and the fourth. If the object from the most similar pair (O_i, O_j) is not present in a pair of the least similar objects, then the most similar objects are placed as the second and the third. In this situation, the respondent should also be asked to indicate the second most similar pair of objects. For example, if (O_i, O_j) is the most similar pair, (O_i, O_k) is the second most similar pair and (O_k, O_l) is the least similar pair, the tetrad is $(O_k, O_i, O_j, O_l)^*$. The asterisk next to the tetrad indicates a case where the most similar objects are placed as the second and the third.

The number of ratings which a respondent must make for n objects in the method of tetrads is equal to the number of four-element combinations of the n -element set, and amounts to:

$$C_n^4 = \frac{n(n-1)(n-2)(n-3)}{24}, \tag{2}$$

where each pair appears in tetrads $(n-1)(n-2)/2$ times. Therefore, when the number of tetrads is considered too large to be practical, according to the theory of balanced incomplete block designs (see e.g. Rink, 1987), it can be reduced in such a way that all pairs of objects in tetrad sets are presented equally frequently, but less frequently than their potential maximum number. If λ denotes the number of tetrads in which each pair of objects occurs (e.g. using a $\lambda = 2$ design, each pair of objects appears together only twice in the questionnaire), then the reduced number of blocks L_λ equals:

$$L_\lambda = C_n^4 \cdot \frac{2\lambda}{(n-2)(n-3)} = \frac{\lambda n(n-1)}{12}, \tag{3}$$

and equation (3) must satisfy both of these defining relations (see e.g. Rink, 1987):

$$\begin{cases} nr = kL_\lambda \\ (n-1)\lambda = (k-1)r \end{cases}, \tag{4}$$

where:

k is the number of objects in one block (for tetrads $k = 4$),

r is the number of replications of each object in the reduced blocks,

$\lambda = 1, \dots, (n-1)(n-2)/2$.

The number of tetrads for different values of λ and n is shown in Table 1.

Table 1. Number of tetrads for different values of λ and n

n	λ						Full set of tetrads
	1	2	3	4	5	6	
6	15	15
7	7	.	14	.	21	35
8	14	.	.	28	70
9	18	.	.	36	126
10	15	.	30	.	45	210
11	55	330
12	33	.	.	66	495
13	13	26	39	52	65	78	715
14	91	1001
15	105	1365
16	20	40	60	80	100	120	1820
17	68	.	.	136	2380

Source: author's work.

As it is not possible to define a reduced number of blocks for all combinations of λ and n , not all the cells in Table 1 are filled.

3.2. Determining the dissimilarities based on tetrads

It is possible to include the evaluation of paired comparisons into a similarity matrix. The creation of a triangular similarity matrix is possible by assigning the number of points to the pair of objects in the tetrads. The number of points is equal to the number of pairs, for which it can be assumed that the similarity is smaller than the similarity of a given pair. The number of points assigned to pairs from the set of hypothetical tetrads marked with the consecutive letters of the alphabet are presented in Table 2.

Table 2. Number of points for pairs of objects A, B, C, D in example tetrads

Tetrad	Pairs in tetrads					
	AB	AC	AD	BC	BD	CD
ABCD	5	1	0	3	1	3
CABD*	5	4	1	1	3	0
CBAD*	5	1	3	4	1	0
DABC*	5	1	4	3	1	0
DBAC*	5	3	1	1	4	0

* The most similar objects are placed as the second and the third.

Note. The most similar pairs are marked in green, the second most similar pairs in light green, the least similar pairs in grey.

Source: author’s work.

The value of element p_{ij} in the i -th row and the j -th column of the similarity matrix is equal to the sum of points awarded to a pair consisting of the i -th and the j -th objects in all blocks.

To discover the perceptual map by using nonmetric multidimensional scaling, the similarity matrix should be transformed into a matrix of dissimilarities. Dissimilarities δ_{ij} are determined in accordance with the formula:

$$\delta_{ij} = \begin{cases} 1 - \frac{p_{ij}}{\max r \cdot m_{ij}} & \text{for } m_{ij} \neq 0 \\ 0,5 & \text{for } m_{ij} = 0 \end{cases}, \tag{5}$$

where m_{ij} is the number of pairs (O_i, O_j) in blocks and $\max r$ is the maximum number of points that can be obtained by a pair of objects in a block. For tetrads, $\max r$ is equal to 5. The denominator in the second component for $m_{ij} \neq 0$ of equation (5) indicates the maximum possible number of points for pair (O_i, O_j) , i.e. when in all blocks it was considered to be a pair of the most similar objects. Zaborski (2020, 2022) showed that owing to the use of formula (5), performing scaling on the basis of the incomplete tetrad method is possible, even when all pairs of objects cannot be presented equally frequently.

4. Results of the study

In order to make the study results independent of the respondents' subjective effects, the analysis was made on the basis of four distance matrices for 7, 9, 11, and 13 objects that were randomly selected voivodship capitals in Poland. The distances between cities were expressed in kilometres. Using the ALSCAL algorithm, multidimensional scaling was performed for each matrix. As a result, four configurations of n points representing cities ($n = 7, 9, 11, 13$) were obtained. Moreover, for each n , the distances of all pairs of objects (cities) were ranked from the smallest to the largest.

To determine the minimum number of tetrads necessary to recreate the known structure of the objects, for each set of n objects, 18 blocks of L_n tetrads were generated (three different blocks for each L_n value). For $n = 7$, the number of tetrads in a block was $L_7 = 4, 5, 6, 7, 9, 11$, for $n = 9$, $L_9 = 4, 6, 8, 10, 12, 14$, for $n = 11$, $L_{11} = 10, 11, 12, 13, 14, 16$ and for $n = 13$, $L_{13} = 7, 9, 11, 13, 15, 17$. As a result, 72 sets of tetrads were obtained. All L_n -element blocks, for which tetrads were created, were generated using the `ibd` function of the `ibd` package in the R programme (Mandal, 2019).

For each set of objects and for each block of tetrads, similarity matrices were calculated based on the ranking of distances for all pairs of objects. Next, they were transformed into a dissimilarity matrix according to formula (5) and multidimensional scaling was performed with the use of the MINISSA programme. MINISSA performs the basic model of nonmetric multidimensional scaling and it is available in the New MDSX multidimensional scaling package (Coxon & Davies, 1982).

The Procrustes statistic was used to test the quality of matching the resulting configurations of points to the configuration determined based on the distance matrices (Borg & Groenen, 2005; Cox & Cox, 2001):

$$R^2 = \frac{\left\{ \text{tr}(\mathbf{X}^{*\text{T}} \mathbf{Y} \mathbf{Y}^{\text{T}} \mathbf{X}^*)^{\frac{1}{2}} \right\}^2}{\text{tr}(\mathbf{X}^{*\text{T}} \mathbf{X}^*) \text{tr}(\mathbf{Y}^{\text{T}} \mathbf{Y})}, \quad (6)$$

where:

- $\mathbf{X}^* = \mathbf{X}(\mathbf{X}^{\text{T}} \mathbf{Y} \mathbf{Y}^{\text{T}} \mathbf{X})^{\frac{1}{2}} (\mathbf{Y}^{\text{T}} \mathbf{X})^{-1}$ – optimally rotated configuration \mathbf{X} , where:
- $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^{\text{T}}$ – the configuration of points determined on the basis of the incomplete blocks,
- $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]^{\text{T}}$ – the configuration of points determined on the basis of the distance matrix. $R^2 \in (0, 1)$, where 1 means a perfect match.

The quality of the matching configurations of points obtained on the basis of T_L^k (T_L^k – k -th set of L tetrads) to the configuration obtained on the basis of the distance matrix tested with the Procrustes statistics is presented in Table 3.

Table 3. Procrustes statistics for different sets of tetrads

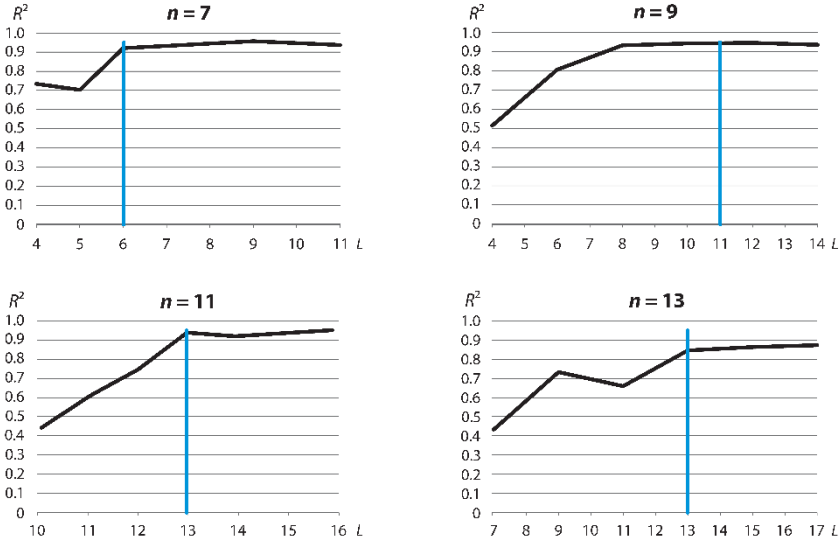
T_L^k	R^2	T_L^k	R^2	T_L^k	R^2	T_L^k	R^2
$n = 7$		$n = 9$		$n = 11$		$n = 13$	
T_{11}^1	0.91709	T_{14}^1	0.93574	T_{16}^1	0.96095	T_{17}^1	0.92804
T_{11}^2	0.96246	T_{14}^2	0.92967	T_{16}^2	0.93460	T_{17}^2	0.91933
T_{11}^3	0.93234	T_{14}^3	0.94632	T_{16}^3	0.96023	T_{17}^3	0.88515
$\overline{T_{11}^k}$	0.93730	$\overline{T_{14}^k}$	0.93724	$\overline{T_{16}^k}$	0.95193	$\overline{T_{17}^k}$	0.91084
T_9^1	0.97065	T_{12}^1	0.96024	T_{14}^1	0.93327	T_{15}^1	0.92088
T_9^2	0.98040	T_{12}^2	0.94652	T_{14}^2	0.92061	T_{15}^2	0.91940
T_9^3	0.92021	T_{12}^3	0.93471	T_{14}^3	0.90692	T_{15}^3	0.91683
$\overline{T_9^k}$	0.95709	$\overline{T_{12}^k}$	0.94716	$\overline{T_{14}^k}$	0.92027	$\overline{T_{15}^k}$	0.91904
T_7^1	0.96827	T_{10}^1	0.96051	T_{13}^1	0.94660	T_{13}^1	0.90023
T_7^2	0.91936	T_{10}^2	0.93733	T_{13}^2	0.93274	T_{13}^2	0.91622
T_7^3	0.91009	T_{10}^3	0.93470	T_{13}^3	0.93682	T_{13}^3	0.89817
$\overline{T_7^k}$	0.93257	$\overline{T_{10}^k}$	0.94418	$\overline{T_{13}^k}$	0.93872	$\overline{T_{13}^k}$	0.90487
T_6^1	0.90302	T_8^1	0.94211	T_{12}^1	0.74459	T_{11}^1	0.73392
T_6^2	0.92113	T_8^2	0.95326	T_{12}^2	0.67872	T_{11}^2	0.69141
T_6^3	0.93950	T_8^3	0.90749	T_{12}^3	0.81894	T_{11}^3	0.66545
$\overline{T_6^k}$	0.92122	$\overline{T_8^k}$	0.93429	$\overline{T_{12}^k}$	0.74742	$\overline{T_{11}^k}$	0.69693
T_5^1	0.74699	T_6^1	0.83260	T_{11}^1	0.68085	T_9^1	0.82126
T_5^2	0.65601	T_6^2	0.89545	T_{11}^2	0.51807	T_9^2	0.67311
T_5^3	0.70576	T_6^3	0.69519	T_{11}^3	0.62874	T_9^3	0.89331
$\overline{T_5^k}$	0.70292	$\overline{T_6^k}$	0.80775	$\overline{T_{11}^k}$	0.60922	$\overline{T_9^k}$	0.79589
T_4^1	0.56308	T_4^1	0.82258	T_{10}^1	0.47152	T_7^1	0.41954
T_4^2	0.84216	T_4^2	0.48776	T_{10}^2	0.39246	T_7^2	0.34305
T_4^3	0.79820	T_4^3	0.23341	T_{10}^3	0.46272	T_7^3	0.31264
$\overline{T_4^k}$	0.73448	$\overline{T_4^k}$	0.51458	$\overline{T_{10}^k}$	0.44223	$\overline{T_7^k}$	0.35841

Note. $\overline{T_L^k}$ – mean of T_L^k , where $k = (1, 2, 3)$, L – number of tetrads in a block.

Source: author's calculations based on the ranking of the distances between all pairs of objects (cities).

Figure 1 shows how, depending on the number of tetrads, the mean values of the Procrustes statistics change for the different number of objects included in the study. The vertical line in the graphs of Figure 1 indicates the minimal number of tetrads generated so that each pair of objects appears in the set at least once.

Figure1. Relationship between the number of tetrads for different n and the value of R^2



Note. L – number of tetrads in a block. The vertical line indicates the minimal number of tetrads in a set, where each pair occurs at least once.

Source: author's calculations based on the ranking of the distances between all pairs of objects (cities).

Based on the test results presented in Table 3 and Figure 1, it can be concluded that the matching of the configurations obtained on the basis of tetrads to the configurations obtained on the basis of the distance matrix is very good if each pair of objects appears in the tetrad set at least once. The mean values of the Procrustes statistics in the study for these tetrad sets range from 0.905 for the set of 13 tetrads for 13 objects to 0.957 for the set of 9 tetrads for 7 objects, with a coefficient of variance equaling 0.00047. The value of the matching function clearly decreases only when each pair of objects does not appear at least once in the tetrad set.

In the configuration of points obtained as a result of nonmetric multidimensional scaling, it is not the distances between the points that are important, but their rank order. Therefore, for all point configurations obtained by the tetrad method, the distances between points were ranked from the smallest to the largest. Then, the Spearman rank correlations for all distances obtained on the basis of tetrads to the distances for the input data were determined. The analysis results are presented in Table 4 and Figure 2. All the correlations presented in Table 4 are statistically significant.

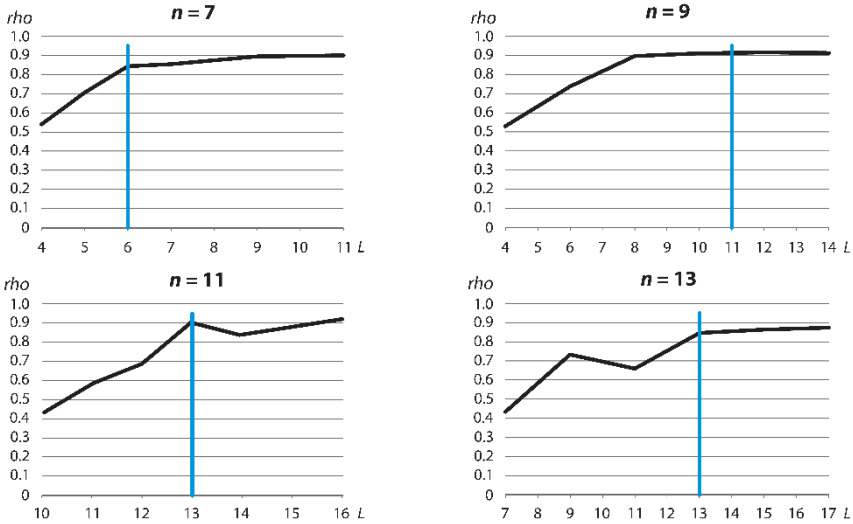
Table 4. Spearman rank correlations for different sets of tetrads

T_L^k	Spearman rho	T_L^k	Spearman rho	T_L^k	Spearman rho	T_L^k	Spearman rho
n = 7		n = 9		n = 11		n = 13	
T_{11}^1	0.88182	T_{14}^1	0.91762	T_{16}^1	0.93088	T_{17}^1	0.89380
T_{11}^2	0.91299	T_{14}^2	0.90147	T_{16}^2	0.89387	T_{17}^2	0.90078
T_{11}^3	0.90650	T_{14}^3	0.91873	T_{16}^3	0.92193	T_{17}^3	0.82757
$\overline{T_{11}^k}$	0.90044	$\overline{T_{14}^k}$	0.91261	$\overline{T_{16}^k}$	0.91556	$\overline{T_{17}^k}$	0.87405
T_9^1	0.87403	T_{12}^1	0.92517	T_{14}^1	0.87684	T_{15}^1	0.86280
T_9^2	0.94026	T_{12}^2	0.91586	T_{14}^2	0.85729	T_{15}^2	0.86525
T_9^3	0.87143	T_{12}^3	0.90672	T_{14}^3	0.78023	T_{15}^3	0.86259
$\overline{T_9^k}$	0.89524	$\overline{T_{12}^k}$	0.91592	$\overline{T_{14}^k}$	0.83812	$\overline{T_{15}^k}$	0.86355
T_7^1	0.92468	T_{10}^1	0.92513	T_{13}^1	0.91320	T_{13}^1	0.82875
T_7^2	0.84156	T_{10}^2	0.90522	T_{13}^2	0.90433	T_{13}^2	0.86181
T_7^3	0.79741	T_{10}^3	0.90253	T_{13}^3	0.89279	T_{13}^3	0.84767
$\overline{T_7^k}$	0.85455	$\overline{T_{10}^k}$	0.91096	$\overline{T_{13}^k}$	0.90344	$\overline{T_{13}^k}$	0.84608
T_6^1	0.89351	T_8^1	0.91946	T_{12}^1	0.72965	T_{11}^1	0.67137
T_6^2	0.84156	T_8^2	0.89762	T_{12}^2	0.59322	T_{11}^2	0.63571
T_6^3	0.79740	T_8^3	0.87264	T_{12}^3	0.73889	T_{11}^3	0.67228
$\overline{T_6^k}$	0.84416	$\overline{T_8^k}$	0.89657	$\overline{T_{12}^k}$	0.68725	$\overline{T_{11}^k}$	0.65979
T_5^1	0.78182	T_6^1	0.78251	T_{11}^1	0.58146	T_9^1	0.70356
T_5^2	0.64416	T_6^2	0.77982	T_{11}^2	0.52460	T_9^2	0.65961
T_5^3	0.69221	T_6^3	0.65137	T_{11}^3	0.65202	T_9^3	0.83685
$\overline{T_5^k}$	0.70606	$\overline{T_6^k}$	0.73790	$\overline{T_{11}^k}$	0.58603	$\overline{T_9^k}$	0.73334
T_4^1	0.28442	T_4^1	0.69294	T_{10}^1	0.45203	T_7^1	0.50639
T_4^2	0.64416	T_4^2	0.57243	T_{10}^2	0.39271	T_7^2	0.36737
T_4^3	0.69221	T_4^3	0.32431	T_{10}^3	0.45733	T_7^3	0.42828
$\overline{T_4^k}$	0.54026	$\overline{T_4^k}$	0.52989	$\overline{T_{10}^k}$	0.43402	$\overline{T_7^k}$	0.43401

Note. As in Table 3.

Source: author's calculations based on the ranking of the distances between all pairs of objects (cities).

Figure 2. Relationship between the number of tetrads for different n and Spearman's rank correlation coefficient



Note. As in Figure 1.

Source: author's calculations based on the ranking of the distances between all pairs of objects (cities).

The conclusions of the correlation analysis are in line with those of Procrustes analysis. If each pair appears in the tetrad set at least once, the ranks are in strong agreement. For the analysed examples, when each pair appears in the tetrad set at least once, the mean value of Spearman's rank correlation coefficient ranges from 0.844 for the set of 6 tetrads for 7 objects to 0.916 for the set of 12 tetrads for 9 objects. The coefficient of variation for these results is 0.001571, which proves that the selection of the tetrad set for a given L_n is not significant.

5. Conclusions

The tetrad method can be classified as one of the group of methods in which dissimilarities are assessed on the basis of comparisons with pairs of objects. These methods require the respondents to make many assessments, causing fatigue and weariness, as a result of which the obtained results do not always fully reflect the respondents' attitudes.

The study confirms that, despite the above, the tetrad method can be an alternative to obtaining dissimilarity data when the variables are measured on an ordinal scale. It has been shown that satisfactory results in determining dissimilarities by means of the method of tetrads under incomplete block designs can be obtained already at the point when each pair of objects appears in the tetrad set at least once. This allows for a significant reduction in the number of multiple opinions provided by the respondents.

The results of the study also indicate that the choice of the incomplete set of tetrads has no significant effect on the results of nonmetric multidimensional scaling.

Despite the advantages, the main limitation of the incomplete tetrad method, as well as most methods of direct dissimilarity determination, is the possibility of using it for a large number of objects. For even a dozen or so objects, the need for respondents to make assessments may cause fatigue and boredom. In such a situation, the solution may be to aggregate the results of different respondents (or groups of respondents), provided that different subsets of the full set of objects are presented to individual respondents.

References

- Bijmolt, T. H. A. (1996). *Multidimensional Scaling in Marketing: Towards Integrating Data Collection and Analysis*. Capelle a/d Ussel: Labyrinth Publication.
- Borg, I., & Groenen, P. J. F. (2005). *Modern Multidimensional Scaling. Theory and Applications* (2nd edition). Springer. <https://doi.org/10.1007/0-387-28981-X>.
- Burton, M. L., & Nerlove, S. B. (1976). Balanced Design for Triads Tests: Two Examples from English. *Social Science Research*, 5(3), 247–267. [https://doi.org/10.1016/0049-089X\(76\)90002-8](https://doi.org/10.1016/0049-089X(76)90002-8).
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3), 283–319. <https://doi.org/10.1007/BF02310791>.
- Cox, T. F., & Cox, M. A. A. (2001). *Multidimensional Scaling* (2nd edition). Chapman & Hall/CRC. <https://doi.org/10.1201/9780367801700>.
- Coxon, A. P. M., & Davies, P. M. (1982). *The User's Guide to Multidimensional Scaling with Special Reference to the MDS(X) Library of Computer Programs*. Heinemann Educational Books.
- Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33(4), 469–506. <https://doi.org/10.1007/BF02290164>.
- Humphreys, M. A. (1982). Data collecting effects on nonmetric multidimensional scaling solutions. *Educational and Psychological Measurement*, 42(4), 1005–1022. <https://doi.org/10.1177/001316448204200408>.
- Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1–27. <https://doi.org/10.1007/BF02289565>.
- Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115–129. <http://dx.doi.org/10.1007/BF02289694>.
- de Leeuw, J., & Heiser, W. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate Analysis V* (pp. 501–522). North-Holland.
- Mandal, B. N. (2019). *ibd: Incomplete Block Designs*. www.r-project.org.
- Morris, M. D. (2010). *Design of Experiments. An Introduction Based on Linear Models*. Chapman & Hall/CRC.

- Rink, D. R. (1987). An improved preference data collection method: Balanced incomplete block designs. *Journal of the Academy of Marketing Science*, 15(3), 54–61. <https://doi.org/10.1007/BF02721954>.
- Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I. *Psychometrika*, 27, 125–140. <https://doi.org/10.1007/BF02289630>.
- Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences in multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika*, 42, 7–67. <https://doi.org/10.1007/BF02293745>.
- Tsogo, L., Masson, M. H., & Bardot, A. (2000). Multidimensional Scaling Methods for Many-Object Sets: A Review. *Multivariate Behavioral Research*, 35(3), 307–319. https://doi.org/10.1207/S15327906MBR3503_02.
- Zaborski, A. (2003). Wpływ alternatywnych metod pomiaru preferencji na wyniki skalowania wielowymiarowego. *Prace Naukowe Akademii Ekonomicznej w Katowicach: Analiza i prognozowanie zjawisk o charakterze niemetrycznym*, 55–69.
- Zaborski, A. (2020). The Use of the Incomplete Tetrad Method for Measuring the Similarities in Nonmetric Multidimensional Scaling. *Folia Oeconomica Stetinensia*, 20(1), 519–530. <https://doi.org/10.2478/fofi-2020-0030>.
- Zaborski, A. (2022). Triads or tetrads? Comparison of two methods for measuring the similarity in preferences under incomplete block design. *Statistics in Transition new series*, 23(3), 185–198. <https://doi.org/10.2478/stattrans-2022-0037>.

Log-Quad divergence for Non-negative Matrix Factorization in multi-model prediction¹

Ryszard Szupiluk,^a Paweł Rubach^b

Abstract. The aim of this paper is to present a new Non-negative Matrix Factorization (NMF) algorithm based on Log-Quad divergence, and to demonstrate its application to the separation of latent destructive components contained in prediction results in a multi-model approach. We provide an example of its application to a real economic problem, i.e. forecasting electricity consumption on the basis of information about hourly use of electricity in Poland in the period of 1988–1997. We evaluated and compared this method with other blind signal (source) separation techniques, such as Independent Component Analysis (ICA) and Algorithm for Multiple Unknown Signals Extraction (AMUSE). The results show that the NMF algorithm based on Log-Quad divergence has an interesting ability to improve predictions for small volumes of data.

Keywords: Non-negative Matrix Factorization, NMF, latent components identification, blind source separation, blind signal separation, prediction, ICA, AMUSE

JEL: C02, C50

Dywergencja Log-Quad dla nieujemnej faktoryzacji macierzy w predykcji za pomocą podejścia wielomodelowego

Streszczenie. Celem niniejszego artykułu jest przedstawienie nowego algorytmu nieujemnej faktoryzacji macierzy (Non-negative Matrix Factorization – NMF) opartego na dywergencji Log-Quad i możliwości zastosowania go do separacji ukrytych komponentów destrukcyjnych zawartych w predykcjach w ujęciu wielomodelowym. Jako przykład wykorzystania go w praktyce wybrano prognozowanie zużycia energii elektrycznej na podstawie danych dotyczących godzinowego zużycia energii elektrycznej w Polsce w latach 1988–1997. Proponowany algorytm oceniono i porównano z innymi technikami z obszaru ślepej separacji sygnałów (źródeł), takimi jak analiza składowych niezależnych (Independent Component Analysis – ICA) oraz algorytm

¹ Artykuł został opracowany na podstawie referatu wygłoszonego na konferencji Multivariate Statistical Analysis MSA'2022, która odbyła się w dniach 7–9 listopada 2022 r. w Łodzi. / The article is based on a paper delivered at the Multivariate Statistical Analysis MSA'2022 Conference, held on 7–9 November 2022 in Łódź, Poland.

^a Szkoła Główna Handlowa w Warszawie, Instytut Informatyki i Gospodarki Cyfrowej, Polska / Warsaw School of Economics, Institute of Information Systems and Digital Economy, Poland.
ORCID: <https://orcid.org/0000-0002-4416-7131>. E-mail: ryszard.szupiluk@sgh.waw.pl.

^b Szkoła Główna Handlowa w Warszawie, Instytut Informatyki i Gospodarki Cyfrowej, Polska / Warsaw School of Economics, Institute of Information Systems and Digital Economy, Poland.
ORCID: <https://orcid.org/0000-0001-5487-609X>. Autor korespondencyjny / Corresponding author, e-mail: pawel.rubach@sgh.waw.pl.

dekorelacji wieloetapowej (Algorithm for Multiple Unknown Signals Extraction – AMUSE). Wyniki pokazują, że algorytm NMF oparty na dywergencji Log-Quad charakteryzuje się ciekawą właściwością poprawiania predykcji w przypadku niedużych wolumenów danych.

Słowa kluczowe: nieujemna faktoryzacja macierzy, NMF, identyfikacja ukrytych komponentów, ślepa separacja źródeł, ślepa separacja sygnałów, predykcja, ICA, AMUSE

1. Introduction

One of the common tasks in data analysis is to find a new interesting or useful representation of data. For this purpose, various transformations can be used. Over the last two decades, the area of multivariate methods developed particularly intensively, especially the methods related to the problem of blind signal separation. In such a problem, the goal is to separate (reproduce, estimate, reconstruct, identify) *a priori* unknown signals mixed in an unknown mixing system. Separation takes place using mixed data only. The application of a problem formulated in such a way may be found in many practical areas, such as: telecommunications, medicine, geology, acoustics, economics and business (Comon & Jutten, 2010). A classic example of blind source separation is the ‘cocktail party’ problem. In this kind of a party, many individuals are speaking simultaneously, thus generating noise, which we record using microphones. Based on such mixed signals, we want to isolate individual conversations. Classic signal filtering techniques based on the Fourier transform (FT) have limited application, because people speak in approximately the same frequency band (ca. 1 kHz), and there is no frequency diversity necessary to use the Fourier transform. In this type of problems, techniques currently referred to as blind signal separation, such as the Independent Component Analysis (ICA) and the Non-negative Matrix Factorization (NMF), have proven to be effective. They are based on other mathematical criteria.

As a result of intense research, in addition to the classic methods such as Principal Component Analysis (PCA) or factor analysis, a long list of new methods has emerged, such as the above-mentioned ICA and NMF, Smooth Principal Component Analysis (SmPCA), Sparse Component Analysis (SCA) and Independent Factor Analysis (IFA). The whole set of these methods is also referred to as Latent Component Analysis (LCA; Cichocki & Amari, 2002).

Although each of these methods is based on different mathematical criteria with accompanying assumptions specific to the given method, they have certain common features. First of all, they are most often considered in the context of a machine-learning approach, which entails the numerical optimization of the objective function that is the basis of a given method. Consequently, much of the literature on statistical methods, e.g. on the ICA or algebraic methods such as the NMF, concern the context of neural networks and/or machine learning.

Another similarity between the above-mentioned methods is their performance in the area of blind signal separation, namely the fact that they can be applied to the reconstruction of mixed-source signals. It is expected that all these methods, despite being mathematically distinct, lead to similar or even identical results. An example illustrating the whole issue can be the ‘cocktail party’ problem mentioned before, where the goal is to separate a single conversation from all the sounds produced by many people talking simultaneously. A success or a failure in solving such a problem is whether the statement is understandable or not.

At the same time, there are significant differences in the implementation of individual methods within the LCA when they are applied to blind separation. These differences exist both between the methods and within a given method between particular numerical algorithms. One of these differences is the amount of data necessary for a given method to operate effectively.

The aim of this paper is to present a new NMF algorithm based on the Log-Quad divergence (Cichocki et al., 2009), and to demonstrate its application to the separation of latent destructive components contained in prediction results in a multi-model approach. The test was performed on the basis of the forecast of energy consumption based on a dataset of hourly electricity consumption in Poland in the years 1988–1997. The goal was to compare the operation of the Log-Quad NMF-based system with the ICA-based one, as well as showing the differences in the operation of other NMF algorithms, i.e. the Image Space Reconstruction Algorithm (ISRA) developed by Lee and Seung (1999) and the version derived from it by Cichocki and Févotte (Cichocki et al., 2008; Févotte et al., 2009).

In machine learning systems and models, there are usually many arbitrarily selected parameters that affect the efficiency of algorithms. They include, for example, the selection of learning coefficients, nonlinearity forms, or the assumed number of iterations. Since the LCA methods belong to unsupervised learning, we generally cannot control the quality of operation of such algorithms except for easily-interpretable physical signals.

2. Non-negative Matrix Factorization

In its basic form, Non-negative Matrix Factorization can be defined as expressing a matrix \mathbf{X} , with elements $x_{ik} \geq 0$, as a multiplication of two non-negative matrices \mathbf{A} , \mathbf{S} (Berry et al., 2007), i.e.

$$\mathbf{X} \approx \mathbf{AS}, \quad (1)$$

where the elements of matrix \mathbf{A} are non-negative: $a_{ij} \geq 0$, which is also denoted as: $\mathbf{A} \geq 0$. An analogous condition is applied to matrix \mathbf{S} .

To assess the extent to which matrices \mathbf{X} , \mathbf{A} , \mathbf{S} fulfill equation (1), the following objective function $J = L(\mathbf{X}, \mathbf{AS})$ is introduced, which at the same time is the objective function used to formulate the NMF algorithm. This function may take the classical form of a measure of the distance between \mathbf{X} and \mathbf{AS} expressed by formula $J = \|\mathbf{X} - \mathbf{AS}\|_p$, where $\|\cdot\|_p$ is the p -norm. For $p = 2$, we obtain the Euclidean distance, on the basis of which the ISRA algorithm, considered to be the first solution of the NMF in its current form, was derived by Lee and Seung (1999). This algorithm can be represented in a matrix form as:

$$\mathbf{A} \leftarrow \mathbf{A} \times \mathbf{X} \mathbf{S}^T / \mathbf{A} \mathbf{S} \mathbf{S}^T \quad (2)$$

and

$$\mathbf{S} \leftarrow \mathbf{S} \times \mathbf{A}^T \mathbf{X} / \mathbf{A}^T \mathbf{A} \mathbf{S}, \quad (3)$$

where \times and $/$ mean element-wise multiplication (Hadamard product) and element-wise division, respectively.

The ISRA algorithm does not guarantee convergence to a local minimum, but rather to a certain stationary point (which does not have to be a minimum). For this reason, it is currently rarely used in real applications; more often it plays the role of a reference point, a solution against which new algorithms are compared. It also serves as an inspiration for the development of other approaches. The disadvantages of using this algorithm include its slow convergence (if it converges at all) and the fact that once it is set to 0, it stays that way. These limitations necessitated the search for alternative solutions. One of the possibilities is to use the divergence function $D(y||z)$ which became the basis for deriving NMF algorithms such that $J = D(\mathbf{X}||\mathbf{AS})$.

3. Log-Quad divergence

Divergence functions are among the most popular criteria for the derivation of NMF algorithms. Unlike ordinary distance measures, a divergence function does not have to satisfy the triangle inequality, and is usually asymmetric. Historically, the development of divergence measures involved assessing the similarity between distributions. Currently, they are used to assess the similarity (or lack thereof) between non-negative variables, vectors, matrices or functions. Divergences can be defined for both continuous and discrete quantities. Divergence can be accepted or interpreted as a specific measure of distance (quasi-distance; Amari, 1985;

Cichocki et al., 2009; Csiszár, 1978). One of the broader families of divergences are Bregman divergences defined as (Bregman, 1967):

$$D_{\varphi}(\mathbf{X}|\mathbf{AS}) = \sum_{i=1}^m \sum_{t=1}^n (\varphi(x_{it}) - \varphi([\mathbf{AS}]_{it}) - \varphi'(x_{it})(x_{it} - [\mathbf{AS}]_{it})), \quad (4)$$

where $[\mathbf{AS}]_{it}$ denotes the (i, t) -th element of the $m \times n$ matrix \mathbf{AS} , and $\varphi(u)$ is a strictly convex function that has a continuous derivative $\varphi'(u)$.

For non-linearity and $\varphi(u) = u^2 + u \ln(u)$, $u > 0$, the Log-Quad algorithm obtains the following divergence (Cichocki et al., 2009):

$$D_{\text{LQ}}(\mathbf{X}|\mathbf{AS}) = \sum_{i=1}^m \sum_{t=1}^n \left((x_{it} - [\mathbf{AS}]_{it})^2 + x_{it} \ln \frac{x_{it}}{[\mathbf{AS}]_{it}} - x_{it} + [\mathbf{AS}]_{it} \right). \quad (5)$$

Log-Quad divergence is one of the measures of similarity that are relatively rarely discussed or used. It is specific in that it is actually a combination of some popular similarity measures such as squared differences, Kullback-Leibler entropy and ordinary differences. While these are usually standalone measures, here they are combined in one formula. We define the NMF problem for divergence (5) as $\min_{\mathbf{A}, \mathbf{S}} D_{\text{LQ}}(\mathbf{X}|\mathbf{AS})$, and use the formula of the alternating multiplicative NMF algorithm for Bregman divergence (Dhillon & Sra, 2005) in the form of:

$$\mathbf{A} \leftarrow \mathbf{A} \cdot \frac{(\varphi''(\mathbf{AS}) \cdot \mathbf{X}) \mathbf{S}^T}{(\varphi''(\mathbf{AS}) \cdot (\mathbf{AS})) \mathbf{S}^T}, \quad (6)$$

and for matrix \mathbf{S} :

$$\mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\mathbf{A}^T (\varphi''(\mathbf{AS}) \cdot \mathbf{X})}{\mathbf{A}^T (\varphi''(\mathbf{AS}) \cdot (\mathbf{AS}))}. \quad (7)$$

In the case of assuming the objective function in the form of Log-Quad divergence, the second derivative of the function $\varphi(u)$ takes the form of $\varphi''(u) = 2 + 1/u$, which leads to the following algorithm for estimating matrices \mathbf{A} and \mathbf{S} :

$$\mathbf{A} \leftarrow \mathbf{A} \cdot \frac{((2 + 1./(\mathbf{AS})) \cdot \mathbf{X}) \mathbf{S}^T}{((2 + 1./(\mathbf{AS})) \cdot (\mathbf{AS})) \mathbf{S}^T}, \quad (8)$$

and

$$\mathbf{S} \leftarrow \mathbf{S} \cdot \frac{\mathbf{A}^T((2 + 1./(\mathbf{AS})) \times \mathbf{X})}{\mathbf{A}^T((2 + 1./(\mathbf{AS})) \times (\mathbf{AS}))}. \quad (9)$$

The algorithm obtained in models (8)–(9) is one of many variants possible to obtain on the basis of the objective function in the form of Log-Quad divergence. These forms depend on the adopted method of minimizing the objective function.

4. Multi-model system for improving quality of prediction

The NMF model (1) can be interpreted in terms of a prediction-improvement system. Let us assume that the rows of matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$ contain the results of forecasts from different models. Another assumption is that they are a linear combination $\mathbf{X} = \mathbf{AS}$ of hidden components responsible for the correct results of the forecast as well as the components responsible for errors. So, we can write that matrix $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_n]^T = [\hat{\mathbf{s}}_1, \dots, \hat{\mathbf{s}}_p, \tilde{\mathbf{s}}_{p+1}, \dots, \tilde{\mathbf{s}}_n]^T$. This means it consists of constructive components $\hat{\mathbf{s}}_i$ and destructive components $\tilde{\mathbf{s}}_j$. By identifying matrices \mathbf{A} and \mathbf{S} and then eliminating the destructive components (i.e. assuming $\tilde{\mathbf{s}}_j = \mathbf{0}$), we should obtain improvement in the prediction quality

$$\hat{\mathbf{X}} = \mathbf{A}[\hat{\mathbf{s}}_1, \dots, \hat{\mathbf{s}}_p, \mathbf{0}_{p+1}, \dots, \mathbf{0}_n]^T, \quad (10)$$

where $\hat{\mathbf{X}}$ denotes a matrix whose rows contain corrected forecasts (Szupiluk et al., 2007).

5. Practical experiment

The verification and demonstration of the new NMF Log-Quad algorithm was performed using hourly energy consumption data for Poland in the period of 1988–1997. The dataset contains observations of hourly electricity consumption and forecasts generated every 24 hours on the basis of previous hourly consumption. There are six independent forecasts that were generated using Multilayer Perceptron neural networks, with 12, 18, 21, 24, 27 and 30 neurons in the hidden layer. Overall, the dataset contains 43,824 predictions and is divided into three arbitrarily divided subsets (Dataset A: 17,544 predictions, Dataset B: 17,544 predictions, and Dataset C: 8,736 predictions).

The goal of the experiment was to improve the forecasts of energy consumption by applying our multi-model prediction improvement schema to a given dataset. According to this schema, the initial predictions are decomposed using the NMF Log-Quad algorithm into six latent components. In the next step, a selected base latent component or a set of components is eliminated (replaced with zeros), and subsequently the latent components are recomposed into six forecasts. The Mean Squared Error (MSE) is computed for initial predictions as well as on the recomposed ones, and then compared.

In order to observe the performance of the algorithm in more detail, the whole dataset was divided into smaller frames with predicted consumption values of various sizes, namely 20, 40, 50, 75, 100, 500, 1,000, 2,000, 3,000, 4,000 and 5,000. Overall, there were 5,355 frames. For each frame, a full combination of different decompositions using different sets of base components were tested – starting from the elimination of a single base component up to the replacement of five out of six base components with zeros. There are 62 such combinations that were evaluated for each frame.

Table 1. Prediction improvements using the NMF Log-Quad algorithm

Frame size	Predictions range (dataset)	Identifier of base component removed	Initial MSE 10 ⁻³	Final MSE 10 ⁻³	Improvement in %
20	8560–8580 (C)	2, 6	24.32	19.50	19.8
40	12400–12440 (B)	4	14.06	11.33	10.6
50	5400–5450 (A)	5	3.79	3.36	11.2
75	8550–8625 (C)	6	9.51	6.67	29.9

Source: authors’ computations based on an energy dataset, performed using Python and matlab implementations of the NMF Log-Quad algorithm.

Table 1 shows examples where improvement in forecasts was observed. The table features particular frames, ranges of predictions, the dataset identifier (A, B or C), the identifier of the base component that was removed, the values of initial and final MSE, and a percentage drop in the prediction error. Overall, 39 cases were found where a forecast improvement was observed. Interestingly, all of these results were achieved on relatively small frames (up to 75 values) and with just a single or at most two base components eliminated. These results prove that the NMF Log-Quad algorithm can be used in such a prediction-improvement system; however, as is shown in Table 2, it does not perform as effectively as other algorithms, especially those that use the ICA, e.g. the Joint Approximate Diagonalization of Eigenmatrices (JADE; Cardoso & Souloumiac, 1996; Rutledge & Jouan-Rimbaud Bouveresse, 2013) or Algorithm for Multiple Unknown Signals Extraction (AMUSE; Szupiluk & Rubach, 2020; Tong et al., 1990).

Table 2. Comparison of prediction improvement combinations for different algorithms

Frame size	NMF LQ		JADE (ICA)		AMUSE (SOS BSS)	
	number of cases	average improvement in %	number of cases	average improvement in %	number of cases	average improvement in %
20	27	28.2	1,330	25.3	975	6.8
40	9	21.7	702	21.9	993	6.9
50	1	11.2	408	20.4	841	6.6
75	2	31.2	320	16.0	1,016	5.9
100	302	13.1	1,215	6.0
500	50	7.7	1,410	4.8
1,000	18	6.0	957	4.5
2,000	18	3.1	508	4.3
3,000	5	2.2	332	4.6
4,000	10	1.8	295	4.0
5,000	4	2.1	208	4.4

Source: authors' computations based on an energy dataset, performed using standard Python, SciKit Learn and matlab implementations of the algorithm.

The same prediction improvement schema was tested with three other algorithms: the classical NMF implementation by Cichocki and Févotte (Cichocki et al., 2008; Févotte et al., 2009), available e.g. in the Python, SciKit Learn library, the ICA method using the JADE algorithm, and the Second Order Statistics (SOS) Blind Source Separation (BSS) AMUSE algorithm.

It is very important to underline that unlike the proposed NMF Log-Quad (NMF LQ), the classical NMF algorithm did not achieve a prediction improvement in any of the explored subframes or base component combinations. On the other hand, the other two algorithms achieved better results than the NMF LQ in the same setup. Table 2 shows the number of cases of frames and base component combinations that lead to lower values of the MSE. Not only were there more combinations that resulted in prediction improvement, but the JADE algorithm achieved a more significant MSE drop for all the frame sizes. Table 3, on the other hand, demonstrates that both JADE and AMUSE were often able to achieve prediction improvements in the cases where many base components were removed (sometimes four or even five components out of six). This proves their ability to separate the prediction from the noise that lowers its quality.

Table 3. Number of base components that were removed to obtain a prediction improvement – comparison of three algorithms: NMF Log-Quad, JADE and AMUSE

Number base components removed	NMF LQ (all cases: 39)	JADE (all cases: 3,148)	AMUSE (all cases: 9,025)
1	36	1,958	3,473
2	3	871	2,810
3	263	1,855
4	52	751
5	4	136

Source: authors’ computations based on an energy dataset, performed using standard Python, SciKit Learn and matlab implementations of the algorithm.

6. Conclusions

This paper introduces a new Non-negative Matrix Factorization algorithm derived from Log-Quad divergence, and proposes its application to a prediction-improvement system. This system can be applied in various areas where different prediction techniques are possible, and might also be used as a model aggregation method. In particular, our approach can be used in specific cases where data contains physical type of noise (e.g. measurement errors). It is particularly adequate for models created in the data-mining methodology (using machine learning), where we do not assume that the model represents the true nature of the analysed phenomenon, but is only intended to fulfill an instrumental purpose, giving the best possible forecast.

We implemented the new NMF Log-Quad algorithm and the prediction improvement system to energy consumption forecasting, and its performance was compared with other methods used in blind signal separation. It should be noted that the NMF Log-Quad algorithm is characterized by a relatively high efficiency for signals with a small amount of data and, within our study, it showed an advantage over methods such as Independent Component Analysis (ICA) or Algorithm for Multiple Unknown Signals Extraction (AMUSE). However, it performed worse than the above-mentioned methods when processing larger amounts of data. These approaches can be considered complementary and used jointly to effectively improve prediction for both large and small data volumes. Compared to other NMF algorithms, such as the Image Space Reconstruction Algorithm (ISRA) and its derivatives, the NMF Log-Quad improves predictions to a greater degree. It is also a more stable and effective algorithm in terms of the Non-negative Matrix Factorization itself.

References

- Amari, S. (1985). Differential Geometry of Statistical Models. In S. Amari (Ed.), *Differential-Geometrical Methods in Statistics* (pp. 11–65). Springer-Verlag. https://doi.org/10.1007/978-1-4612-5056-2_2.
- Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P., & Plemmons, R. J. (2007). Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, 52(1), 155–173. <https://doi.org/10.1016/j.csda.2006.11.006>.
- Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3), 200–217. [https://doi.org/10.1016/0041-5553\(67\)90040-7](https://doi.org/10.1016/0041-5553(67)90040-7).
- Cardoso, J.-F., & Souloumiac, A. (1996). Jacobi Angles for Simultaneous Diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 17(1), 161–164. <https://doi.org/10.1137/S0895479893259546>.
- Cichocki, A., & Amari, S. (2002). *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley & Sons. <https://doi.org/10.1002/0470845899>.
- Cichocki, A., Lee, H., Kim, Y.-D., & Choi, S. (2008). Non-negative matrix factorization with α -divergence. *Pattern Recognition Letters*, 29(9), 1433–1440. <https://doi.org/10.1016/j.patrec.2008.02.016>.
- Cichocki, A., Zdunek, R., Phan, A. H., & Amari, S. (2009). *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. John Wiley & Sons. <https://doi.org/10.1002/9780470747278>.
- Comon, P., & Jutten, C. (Ed.). (2010). *Handbook of Blind Source Separation: Independent Component analysis and Applications*. Academic Press. <https://doi.org/10.1016/C2009-0-19334-0>.
- Csiszár, I. (1978). Information measures: A critical survey. In J. Kožešnik (Ed.), *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians held in Prague, 18 to 23 August 1974* (pp. 73–86). Springer. <https://www.fuw.edu.pl/~kostecki/scans/csiszar1978.pdf>.
- Dhillon, I. S., & Sra, S. (2005). Generalized Nonnegative Matrix Approximations with Bregman Divergences. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems*, 18 (pp. 283–290). <https://proceedings.neurips.cc/paper/2005/hash/d58e2f077670f4de9cd7963c857f2534-Abstract.html>.
- Févotte, C., Bertin, N., & Durrieu, J.-L. (2009). Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis. *Neural Computation*, 21(3), 793–830. <https://doi.org/10.1162/neco.2008.04-08-771>.
- Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788–791. <https://doi.org/10.1038/44565>.
- Rutledge, D. N., & Jouan-Rimbaud Bouveresse, D. (2013). Independent Components Analysis with the JADE algorithm. *TrAC Trends in Analytical Chemistry*, 50, 22–32. <https://doi.org/10.1016/j.trac.2013.03.013>.
- Szupiluk, R., & Rubach, P. (2020). Identification of Delays in AMUSE Algorithm for Blind Signal Separation of Financial Data. In L. Rutkowski, R. Scherer, M. Korytkowski, W. Pedrycz, R. Tadeusiewicz, & J. M. Zurada (Eds.), *Artificial Intelligence and Soft Computing. ICAISC 2020. Lecture Notes in Computer Science* (pp. 253–261). Springer. https://doi.org/10.1007/978-3-030-61534-5_23.

- Szupiluk, R., Wojewnik, P., & Ząbkowski, T. (2007). Smooth Component Analysis as Ensemble Method for Prediction Improvement. In M. E. Davies, C. J. James, S. A. Abdallah, & M. D. Plumbley (Eds.), *Independent Component Analysis and Signal Separation. ICA 2007. Lecture Notes in Computer Science* (pp. 277–284). Springer. https://doi.org/10.1007/978-3-540-74494-8_35.
- Tong, L., Soon, V. C., Huang, Y. F., & Liu, R. (1990). AMUSE: A new blind identification algorithm. In *IEEE International Symposium on Circuits and Systems* (pp. 1784–1787). <https://doi.org/10.1109/ISCAS.1990.111981>.

Nadmiarowe zgony podczas pandemii COVID-19 w Polsce i ocena skuteczności szczepień

Błażej Kochański^a, Jakub Sochacki^b

Streszczenie. Z powodu pandemii COVID-19 zmarły miliony ludzi na całym świecie. Jak wynika z wielu badań, szczepienia przeciw chorobie wywołanej wirusem SARS-CoV-2 okazały się środkiem ograniczającym skalę zachorowań i liczbę zgonów. Celem badania omawianego w artykule jest pomiar skali pandemii w Polsce za pomocą liczby nadmiarowych zgonów w podregionach według klasyfikacji NUTS 3 i w grupach wieku, a następnie określenie zależności pomiędzy różnicowaniem regionalnym względnej nadwyżki zgonów podczas czwartej fali pandemii a stopniem zaszczepienia populacji. Nadmiarowe zgony są rozumiane jako nadwyżka zgonów zarejestrowanych w stosunku do przewidywanego poziomu. Badaniem objęto okres od marca 2020 r. do lutego 2022 r. Korzystano z zasobów Eurostatu i Głównego Urzędu Statystycznego oraz danych pobranych z rządowego portalu Otwarte Dane.

Analiza liczby nadmiarowych zgonów w podziale na grupy wieku wskazuje, że niemal 90% to zgony osób starszych (w wieku 60 lat i więcej). Z kolei pomiar w podziale na podregiony pozwala stwierdzić, że korelacja nadmiarowych zgonów i udziału zaszczepionych jest ujemna i dość silna. Przy założeniu występowania prostej regresji liniowej okazuje się, że dodatkowy 1% osób zaszczepionych w starszych grupach wieku przełożyłby się średnio na spadek liczby zgonów podczas czwartej fali pandemii o blisko 2000 w skali kraju. Ten wynik świadczy o tym, że szczepienia przeciw COVID-19 były skuteczne w zapobieganiu zgonom z powodu tej choroby wśród osób starszych.

Słowa kluczowe: nadmiarowe zgony, szczepienia, pandemia, COVID-19

JEL: I18, J10, J14, J18

Excess mortality during the COVID-19 pandemic in Poland and the effectiveness of vaccinating the population

Abstract. The COVID-19 pandemic caused millions of deaths worldwide. According to several pieces of research, vaccinations against the illness caused by the SARS-CoV-2 proved to be an effective measure in reducing the infection and death rates. The aim of the study is to measure the scale of the COVID-19 pandemic in Poland on the basis of excess mortality rates in NUTS 3 subregions and in age groups, and then to determine the correlation between

^a Politechnika Gdańska, Wydział Zarządzania i Ekonomii, Polska / Gdańsk University of Technology, Faculty of Management and Economics, Poland. ORCID: <https://orcid.org/0000-0001-8502-931X>. Autor korespondencyjny / Corresponding author, e-mail: blazej.kochanski@pg.edu.pl.

^b Politechnika Gdańska, Wydział Elektroniki, Telekomunikacji i Informatyki, Polska / Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Informatics, Poland. ORCID: <https://orcid.org/0000-0001-8500-4390>. E-mail: jakub.sochacki@wp.pl.

the regional variation in relative excess mortality during the fourth wave of the pandemic and the vaccination rate of the population. Excess mortality is understood as the number of registered deaths that exceeds the expected level. Our study covers the period from March 2020 to February 2022. We used data from Eurostat, Statistics Poland and the governmental Poland's Data portal.

The analysis of the number of excessive deaths by age groups indicates that almost 90% of them are the deaths of older people (aged 60 or over), while the breakdown by NUTS 3 subregion demonstrates that the correlation of excessive deaths and the vaccination rate of the population is negative and fairly strong. Assuming a simple linear regression, it turns out an additional 1% of vaccinated people in the older age groups would have translated to the fall in deaths during the fourth wave of the pandemic by around 2,000 nationwide. This result indicates that vaccinations against COVID-19 were effective in preventing deaths among older people.

Keywords: excess mortality, vaccination, pandemic, COVID-19

1. Wprowadzenie

Według oficjalnych danych w wyniku pandemii COVID-19 zmarły miliony ludzi na całym świecie. Mathieu i in. (2020) podają, że do końca lutego 2022 r. zarejestrowano 437 mln zakażeń koronawirusem SARS-COV-2 i 6 mln zgonów z powodu COVID-19. Dokładne oszacowanie wpływu tej choroby na umieralność na świecie oraz w poszczególnych krajach i regionach może być utrudnione z uwagi na zróżnicowane możliwości przeprowadzania testów diagnostycznych i nieujednoliconą praktykę raportowania (Riffe i in., 2021). Beaney i in. (2020) wskazują, że pomiaru umieralności wynikającej z pandemii można dokonać nie na podstawie zdiagnozowanych przypadków choroby lub zakażenia, ale na podstawie liczby zgonów ogółem. W takiej sytuacji wyznacza się liczbę nadmiarowych zgonów definiowanych jako nadwyżka zarejestrowanych zgonów w stosunku do przewidywanego poziomu zgonów. Takie podejście jest praktyką znaną w epidemiologii. Na przykład Alling i in. (1981), Collins i Lehmann (1953), Housworth i Langmuir (1974), Reichert i in. (2004) oraz Simonsen i in. (2013) stosowali tę metodę do szacowania umieralności spowodowanej epidemiami grypy. W odniesieniu do COVID-19 tego typu badania prowadzili m.in. Böttcher i in. (2021), Islam i in. (2021), Konstantinoudis i in. (2022) oraz Kontis i in. (2020). To zagadnienie było również przedmiotem publikacji Organisation for Economic Co-Operation and Development (2021).

W polskiej literaturze nadmiarowe zgony w związku z pandemią COVID-19 analizował m.in. Murkowski (2021), który wykazał, że w 2020 r. dotyczyły one przede wszystkim osób starszych oraz częściej mężczyzn niż kobiet. Liczba nadmiarowych zgonów znacznie przekraczała oficjalną liczbę zgonów z powodu COVID-19, co najprawdopodobniej w dużej mierze można wyjaśnić niewłaściwym określeniem przyczyny zgonu jako innej niż COVID-19 oraz pośrednim wpływem pandemii na zgony w związku z ograniczeniem dostępu do systemu ochrony zdrowia w czasie

szczytów zachorowań. Podobne wyjaśnienie różnicy pomiędzy liczbą nadmiarowych zgonów a liczbą zgonów przypisanych COVID-19 przedstawili Czerwiński (2021) oraz Jaroszewska i Ołdak (2022). Kuroпка i in. (2021) również podjęli temat nadmiarowych zgonów w 2020 r. i – podobnie jak Murkowski (2021) – wykazali, że nadmiarowe zgony częściej dotyczyły mężczyzn niż kobiet. Stwierdzili ponadto, że nie można wskazać jednoznacznie obszarów o nadmiernie zwiększonej umieralności. Podobne wnioski sformułował Śleszyński (2022), który na podstawie analizy danych na poziomie gmin stwierdził, że w 2020 r. wzrost natężenia zgonów był „dość egalitarny”.

Istnieje wiele metod wyznaczania liczby nadmiarowych zgonów. Przegląd takich metod oraz porównanie sześciu z nich w kontekście COVID-19 przedstawili Levitt i in. (2022). W niniejszym artykule zastosowano metodę zaproponowaną przez Karliński'ego i Kobaka (2021) – przystępną koncepcyjnie, niewymagającą złożonych założeń ani skomplikowanego aparatu matematycznego oraz uwzględniającą sezonowość zgonów i trendy umieralności. Jest to pierwsze zastosowanie tej metody dla Polski na poziomie podregionów według klasyfikacji NUTS 3, a także pierwsza prezentacja korelacji nadmiarowych zgonów i poziomu zaszczepienia dla tego podziału terytorialnego.

Szczepienia przeciw COVID-19 wprowadzono na przełomie lat 2020 i 2021. Zgodnie z szacunkami na podstawie randomizowanych testów klinicznych (Korang i in., 2022) skuteczność szczepień w zapobieganiu zachorowaniom wynosiła od 61% dla (niestosowanych w Polsce) szczepionek z inaktywowanym wirusem do 95% dla szczepionek mRNA. Na podstawie metaanalizy danych zgromadzonych w rzeczywistych warunkach Zheng i in. (2022) wykazali, że szczepienia zapobiegają hospitalizacji ze skutecznością 89,1%, a zgonom – ze skutecznością 99,0%. Danych o skuteczności szczepień dostarczają również badania regionalne, czyli pokazujące współzależność pomiędzy stopniem zaszczepienia a poziomem zachorowań związanych z COVID-19 w poszczególnych regionach, przeprowadzone przede wszystkim dla Stanów Zjednoczonych. Na przykład Harris (2022), opierając się na danych pochodzących ze 112 okręgów (ang. *counties*) o największej populacji, stwierdził, że istnieje ujemna korelacja stopnia zaszczepienia oraz zachorowań i hospitalizacji mających związek z COVID-19. Do podobnych wniosków doszli McLaughlin i in. (2022), którzy wykorzystali dane ze wszystkich okręgów, uwzględniając kontrolne zmienne socjodemograficzne, środowiskowe, ekonomiczne i zdrowotne. Wyniki badań potwierdzających skuteczność szczepień na poziomie krajów przedstawili również Haider i in. (2023), Jabłońska i in. (2021), Ning i in. (2022) oraz Papadopoulos i in. (2022). Dla Polski próbę wyznaczenia korelacji liczby zgonów i stopnia zaszczepienia na poziomie powiatów podjęli Lewandowski i Madoń (2022). Autorzy porównali różnicę w liczbie zgonów między okresami marzec–maj i październik–grudzień

w 2021 r. z odsetkiem osób zaszczepionych i stwierdzili silną ujemną korelację tych dwóch wielkości ($-0,41$).

Celem badania omawianego w niniejszym artykule jest pomiar skali pandemii w Polsce za pomocą liczby nadmiarowych zgonów w podregionach według klasyfikacji NUTS 3 i w grupach wieku oraz określenie zależności pomiędzy zróżnicowaniem regionalnym względnej nadwyżki zgonów podczas czwartej (ostatniej) fali pandemii a stopniem zaszczepienia populacji.

2. Metoda badania

W badaniu wyznaczono liczbę nadmiarowych zgonów na poziomie ogólnopolskim, a także w podregionach i grupach wieku. Uzyskane wielkości wykorzystano do zbadania zależności statystycznej pomiędzy stopniem zaszczepienia populacji a nadmiarowymi zgonami, definiowanymi jako procentowa nadwyżka w stosunku do liczby zgonów przewidywanej w modelu. Związek między poziomem zaszczepienia a nadmiarowymi zgonami w podregionach określono za pomocą współczynnika korelacji Pearsona i prostego modelu regresji liniowej. Badaniem objęto okres, w którym zaobserwowano podwyższony poziom zgonów – pomiędzy sierpniem 2020 r. a lutym 2022 r. Tygodniowe dane o zgonach w podziale na podregiony i grupy wieku pochodzą z zasobów Eurostatu (b.r.).

Po wybuchu pandemii wzrosło zainteresowanie danymi o umieralności – media, opinia społeczna, analitycy i badacze oczekiwali, że informacje będą się pojawiać szybko, z większą częstotliwością oraz że będą dostępne również w podziale na wiek, płeć i regiony. W odpowiedzi na to zapotrzebowanie Eurostat zdecydował się regularnie zbierać i publikować dane o tygodniowej liczbie zgonów w poszczególnych krajach i regionach. Dane dla Polski, wykorzystane w tym artykule, zostały przygotowane przez GUS (2022); są one również podane w przystępnej formie na stronach Eurostatu. Dane tygodniowe o liczbie zgonów dla Polski (podobnie jak dla wielu innych krajów Unii Europejskiej) są dostępne na poziomie podregionów grupujących powiaty, dlatego badanie przeprowadzono właśnie na takim poziomie agregacji danych. Z jednej strony dane o liczbie zgonów (jak i wykorzystywane w tym badaniu dane o szczepieniach) wydają się kompletne i wiarygodne, z drugiej strony analizę tę może zaburzać to, że ze względu na niezarejestrowane migracje najbardziej atrakcyjne aglomeracje są niedoszacowane pod względem liczby stałych mieszkańców (Śleszyński i Niedzielski, 2018). Błąd ten może zaburzać otrzymane wyniki, podobnie jak to się dzieje w przypadku wszelkich innych analiz ludności opartych na danych regionalnych GUS. Niemniej w analizie korelacji nadmiarowych zgonów i udziału osób zaszczepionych jest on przynajmniej częściowo samoczynnie niwelowany, ponieważ obie zmienne mają charakter względny, a wspomniane zaburzenie w podobnym stopniu dotyczy obu zmiennych.

Dane o odsetku osób zaszczepionych zaczerpnięto z rządowego portalu Otwarte Dane (b.r.). W badaniu wykorzystano dane o zaszczepieniu według stanu na 13 października 2021 r., mimo że za początek czwartej fali uznano w tym badaniu pierwszy dzień 37. tygodnia (13 września 2021 r.). Rozbieżność wynika z tego, że dopiero od 13 października na stronach rządowych zaczęto udostępniać znacznie dokładniejsze informacje o odsetku osób zaszczepionych, tj. oprócz danych dla całej populacji również dane dla poszczególnych grup wieku.

Do oszacowania nadmiarowych zgonów wykorzystano podejście zaproponowane przez Karlinsky'ego i Kobaka (2021). Dla każdej z grup (u Karlinsky'ego i Kobaka grupami są kraje, a w badaniu omawianym w niniejszym artykule są to podregiony lub grupy wieku) nadmiarowe zgony w poszczególnych tygodniach lat 2020–2022 zostały obliczone jako nadwyżka zaobserwowanych zgonów w stosunku do wartości bazowej (referencyjnej) dla każdego tygodnia osobno. Wartość bazową wyznaczono na podstawie regresji liniowej zbudowanej osobno dla każdej grupy z wykorzystaniem danych historycznych z lat 2015–2019 według wzoru:

$$D_{t,Y} = \alpha_t + \beta \cdot Y + \varepsilon, \quad (1)$$

gdzie:

$D_{t,Y}$ – liczba zgonów w tygodniu t w roku Y ,

β – nachylenie linii regresji obrazujące roczny trend,

α_t – wyrazy wolne dla każdego tygodnia,

ε – szum losowy.

Taka konstrukcja modelu uwzględnia w prosty sposób zarówno trend roczny (wynikający m.in. – w szczególności w przypadku regionów – ze struktury demograficznej), jak i tygodniową sezonowość zgonów w ciągu roku. Należy zwrócić uwagę, że porównanie liczby zgonów z jednego tygodnia w określonym roku z danymi z tego samego tygodnia w kilku poprzednich latach jest uproszczeniem. W rzeczywistości sezonowość i jej powtarzalność w kolejnych latach dobrze widać na sumach uwzględniających kilka–kilkanaście następujących po sobie tygodni, np. na wynikach miesięcznych. W omawianym badaniu wykorzystano wyniki tygodniowe, łącząc je w okresy zwiększonej liczby zgonów („fale”).

Wartość bazowa $B_{t,Y}$ to wartość przewidywana w modelu regresji dla każdego z regionów. Jest wyrażona wzorem:

$$B_{t,Y} = \hat{\alpha}_t + \hat{\beta} \cdot Y, \quad (2)$$

gdzie Y to 2020, 2021 lub 2022 r.

Ze względu na to, że według normy ISO 53. tydzień ma charakter przestępny i występuje tylko w niektórych latach, jako wartość bazową dla niego w 2020 r. przyjęto wartość teoretyczną z modelu obliczoną dla 52. tygodnia tego roku (identycznie – Karlinsky i Kobak, 2021).

Mając wartości bazowe, można wyznaczyć liczbę nadmiarowych zgonów dla kolejnych tygodni poczynając od 1. tygodnia 2020 r. na podstawie wzoru:

$$\Delta_{t,Y} = D_{t,Y} - B_{t,Y}. \quad (3)$$

Zastosowana metoda ułatwia przedstawienie badanego zjawiska w rozbiciu na składowe (podregiony i grupy wieku), ponieważ suma liczby nadmiarowych zgonów dla poszczególnych składowych równa się liczbie nadmiarowych zgonów wyznaczonej dla całego kraju. Wynika to z tego, że zarówno liczba rzeczywistych zgonów, jak i ich liczba przewidywana w modelu (wartość bazowa) stanowi sumę odpowiednio rzeczywistych i przewidywanych wartości dla odpowiednich grup. W przypadku rzeczywistej liczby zgonów jest to oczywiste, a w przypadku wartości bazowych wynika z liniowej zależności pomiędzy wartościami przewidywanymi w modelu regresji dla całego kraju \hat{y}_t i wartościami przewidywanymi dla części składowych (\hat{y}_i), co można zilustrować następująco:

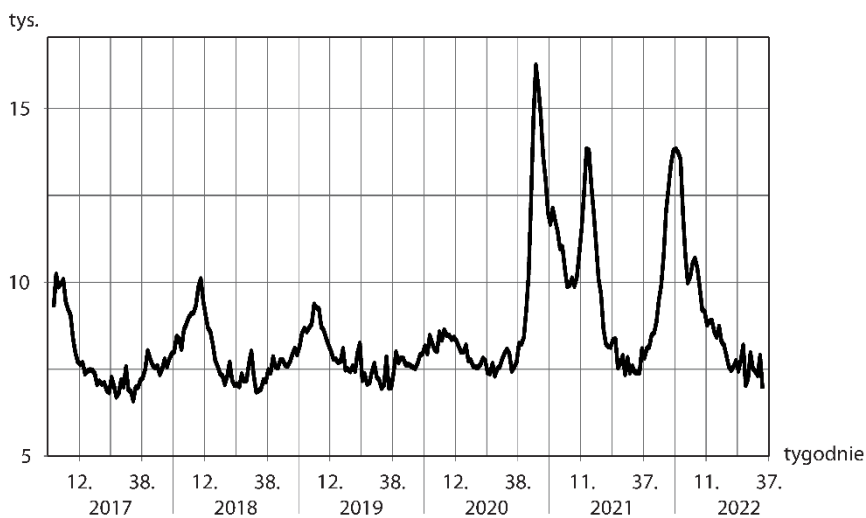
$$\mathbf{y}_t = \Sigma \mathbf{y}_i \Rightarrow \hat{\mathbf{y}}_t = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \mathbf{y}_t = \mathbf{H}\mathbf{y}_t = \mathbf{H}\Sigma \mathbf{y}_i = \Sigma \mathbf{H}\mathbf{y}_i = \Sigma \hat{\mathbf{y}}_i. \quad (4)$$

3. Wyniki badania

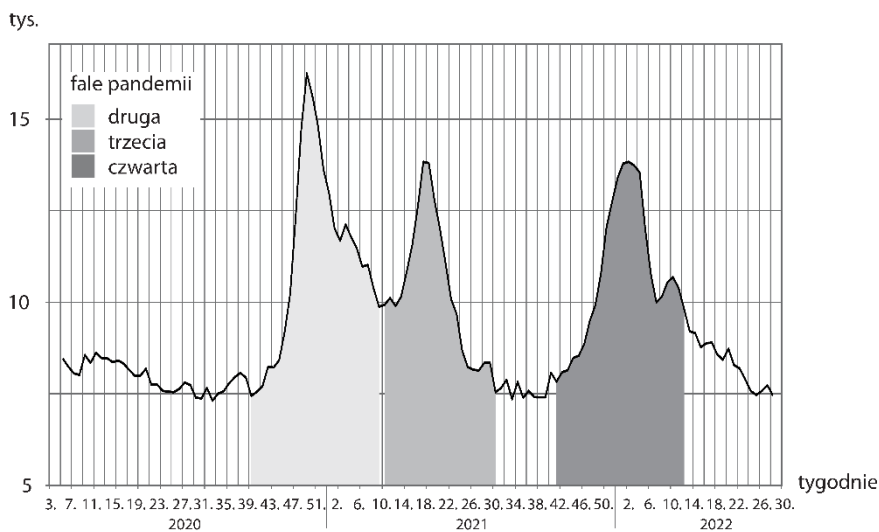
3.1. Nadmiarowe zgony podczas fal pandemii

Na wyk. 1 przedstawiono liczbę zgonów w Polsce w kolejnych tygodniach od 2017 r. Można zauważyć, że dane o całkowitej liczbie zgonów charakteryzują się roczną sezonowością – zimą, w sezonie grypowym, liczba zgonów była wyższa niż latem. Podobnie jak w innych krajach, w Polsce zaobserwowano wyraźne fale związane z pandemią COVID-19 pomiędzy październikiem 2020 r. a kwietniem 2021 r. i od października 2021 r. do lutego 2022 r.

W mediach nadano falom pandemii i nadmiarowej umieralności kolejne numery (Jassat i in., 2021; zob. też np. Taboada i in., 2021). Zwykle przyjmuje się, że pierwsza fala rozpoczęła się wczesną wiosną 2020 r. W Europie, np. we Włoszech, zaobserwowano wtedy wzrost liczby zgonów (Ciminelli i Garcia-Mandicó, 2020), jednak w Polsce nie było wyraźnej nadmiarowej umieralności.

Wykr. 1. Liczba pandemii w Polsce w kolejnych tygodniach

Źródło: obliczenia własne na podstawie: Eurostat (b.r.).

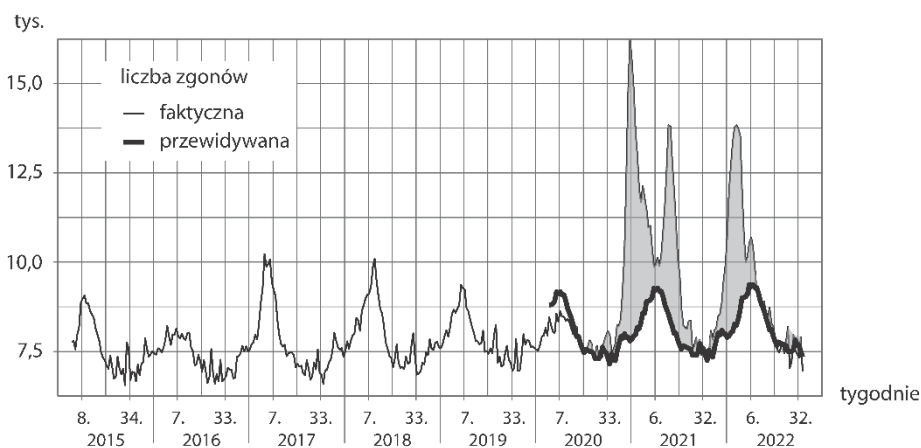
Wykr. 2. Tygodniowa liczba zgonów w Polsce podczas pandemii COVID-19

Źródło: obliczenia własne na podstawie: Eurostat (b.r.).

Na wyk. 2 zilustrowano kolejne fale pandemii. Na podstawie przedstawionych danych można przyjąć, że druga fala pandemii w Polsce wystąpiła między 35. tygodniem 2020 r. a 5. tygodniem 2021 r., trzecia – między 6. a 26. tygodniem 2021 r., a czwarta – między 37. tygodniem 2021 r. a 8. tygodniem 2022 r.

Na wyk. 3 przedstawiono liczbę nadmiarowych zgonów w stosunku do wartości bazowych.

Wykr. 3. Tygodniowa liczba zgonów w Polsce – porównanie z wartościami bazowymi



Źródło: obliczenia własne na podstawie: Eurostat (b.r.).

Na podstawie obliczeń z wykorzystaniem wzoru (3) można stwierdzić, że w poszczególnych falach pandemii w Polsce zarejestrowano następujące liczby nadmiarowych zgonów: w drugiej fali – 73 663, w trzeciej – 42 812, a w czwartej – 55 371 (łącznie – 171 846 zgonów).

3.2. Struktura nadmiarowych zgonów według wieku

W tabl. 1 przedstawiono strukturę nadmiarowych zgonów podczas kolejnych fal pandemii, z podziałem na grupy wieku.

Tabl. 1. Nadmiarowe zgony z podziałem na grupy wieku

Grupa wieku	Fale pandemii					
	druga		trzecia		czwarta	
		w %		w %		w %
Ogółem	73 663	100,0	42 812	100,0	55 371	100,0
Poniżej 10 lat
10–19	30	0,0	49	0,1	33	0,1
20–29	230	0,3	104	0,2	252	0,5
30–39	477	0,6	422	1,0	659	1,2
40–49	1628	2,2	1588	3,7	1723	3,1
50–59	4111	5,6	3581	8,4	3921	7,1
60–69	11 024	15,0	8891	20,8	6629	12,0
70–79	21 814	29,6	16 085	37,6	17 521	31,6
80–89	24 251	32,9	9140	21,3	16 298	29,4
90 lat i więcej	10 190	13,8	3018	7,0	8370	15,1

Uwaga. Ze względu na zaokrąglenia i pominięcie wartości ujemnych (najmłodsza grupa wieku) wartości procentowe nie sumują się do 100%.

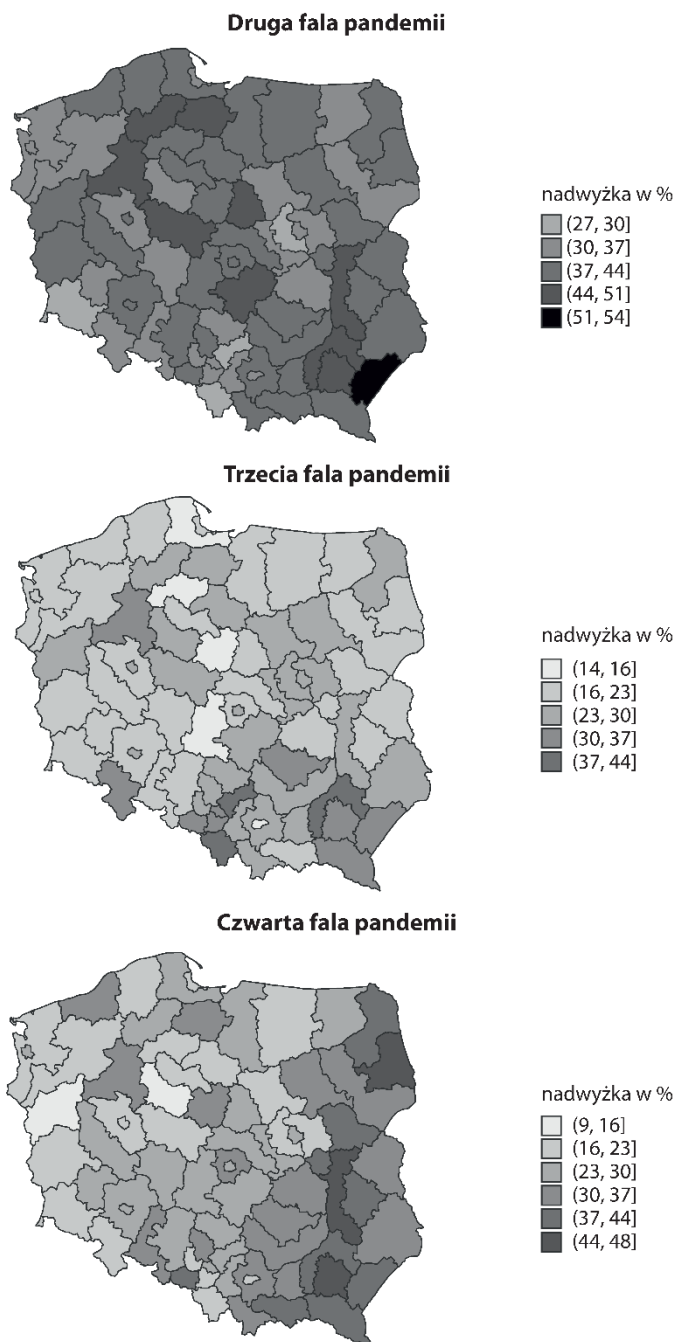
Źródło: obliczenia własne na podstawie: Eurostat (b.r.).

Nadmiarowe zgony, podobnie jak zgony rozpoznane jako spowodowane COVID-19, występowały przede wszystkim w najstarszych grupach wieku (Ghisolfi i in., 2020; Green i in., 2021; Levin i in., 2020). W trzech falach pandemii łącznie nadmiarowe zgony osób w wieku 60 lat i więcej stanowią 89,2% wszystkich nadmiarowych zgonów, a osób w wieku 70 lat i więcej – 73,6%.

3.3. Zróżnicowanie regionalne nadmiarowych zgonów

Informacje o tygodniowej liczbie zgonów są dostępne do poziomu podregionów (w Polsce – 73), więc możliwe jest wyznaczenie nadmiarowych zgonów podczas fal pandemii dla takiego poziomu agregacji danych. Z uwagi na to, że podregiony różnią się wielkością populacji, właściwym podejściem jest wyrażenie nadmiarowych zgonów w postaci nadwyżki procentowej w stosunku do wartości bazowej.

Nadmiarowe zgony dla całej Polski przewyższały wartość bazową podczas drugiej fali pandemii o 37,8%, trzeciej – o 24,7%, czwartej – o 28,3%. Wyniki obliczeń w podziale na podregiony – osobno dla każdej fali – przedstawiono na mapie 1.

Mapa 1. Regionalne zróżnicowanie liczby nadmiarowych zgonów

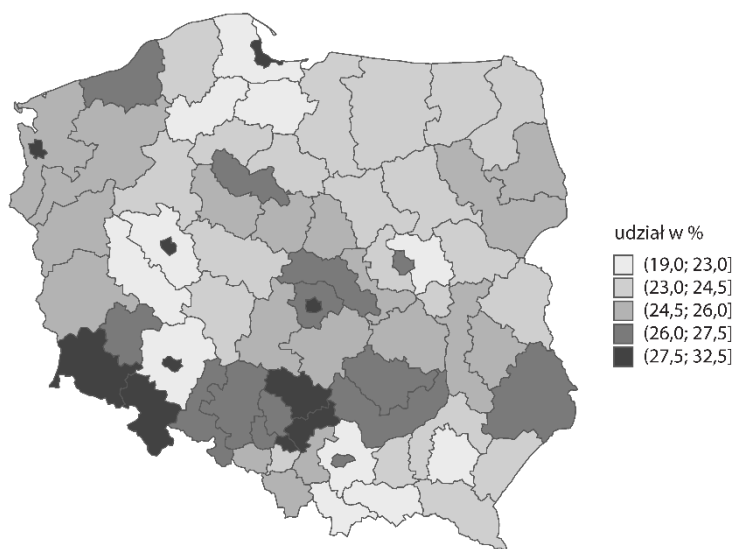
Źródło: obliczenia własne na podstawie: Eurostat (b.r.).

Na kartogramie ilustrującym czwartą falę pandemii można zaobserwować wyraźnie wyższy odsetek nadmiarowych zgonów we wschodniej części kraju, sięgający niemal 48% w niektórych podregionach, w porównaniu z częścią zachodnią, w której nadwyżka zgonów wyniosła kilkanaście procent w stosunku do wartości bazowej. Zróżnicowanie między podregionami było wtedy największe – współczynnik zmienności wyniósł 28,9% (dla porównania: 14,4% dla drugiej fali i 25,8% dla trzeciej).

Na kartogramie dotyczącym drugiej fali pandemii, która miała najbardziej gwałtowny charakter i skutkowała najwyższą liczbą zgonów, widoczne jest pewne zróżnicowanie przestrzenne odsetka nadmiarowych zgonów. Pomimo to jest ono najniższe spośród analizowanych faz, a zanotowany odsetek nadmiarowych zgonów był wysoki w całej Polsce.

Można przypuszczać, że nadzwyczajnie wysoka liczba nadmiarowych zgonów we wschodniej Polsce może mieć związek z dużym udziałem osób w wieku 60 lat i więcej w liczbie ludności ogółem tego obszaru. Jednak korelacja udziału osób starszych w populacji podregionów i odsetka nadmiarowych zgonów jest niewielka i wynosi dla drugiej fali pandemii – 0,32, dla trzeciej – 0,11 i dla czwartej – 0,15. Niska wartość współczynnika korelacji Pearsona i wyniki analizy danych przedstawionych na mapie 2 nie wskazują, aby wyższy udział osób starszych w populacji podregionów przed wybuchem pandemii w Polsce zaważył liczbę nadmiarowych zgonów.

Mapa 2. Udział osób w wieku 60 lat i więcej w liczbie ludności podregionów w 2019 r.



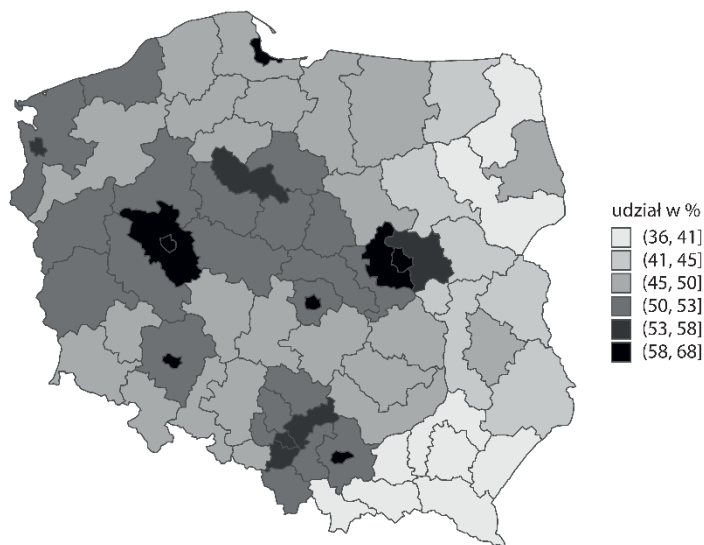
3.4. Zaszczepienie ludności

Czwarta fala pandemii, podczas której regionalne zróżnicowanie nadmiarowych zgonów było największe, nadeszła, gdy dość duża część ludności Polski była już zaszczepiona. Na mapach 3 i 4 przedstawiono zróżnicowanie regionalne odsetka osób zaszczepionych w całej populacji, a także w populacji osób w wieku 60 lat i więcej oraz 70 lat i więcej. Za osobę zaszczepioną uznano w omawianym badaniu osobę zaszczepioną w pełni, tzn. dwoma dawkami w przypadku szczepionki dwudawkowej i jedną dawką w przypadku szczepionki jednodawkowej.

Porównanie regionalnego zróżnicowania udziału zaszczepionych i nadmiarowych zgonów w trakcie czwartej fali pandemii pozwala stwierdzić, że mapy te w dużej mierze się pokrywają – w podregionach charakteryzujących się większym odsetkiem zaszczepionych zaobserwowano niższy poziom nadmiarowych zgonów, natomiast w podregionach o mniejszym stopniu zaszczepienia nadwyżkowa umieralność była znacznie większa.

Na mapie 3 zilustrowano odsetek osób zaszczepionych w populacji podregionów według stanu na 13 października 2021 r.

Mapa 3. Regionalne zróżnicowanie udziału zaszczepionych (stan na 13.10.2021 r.)

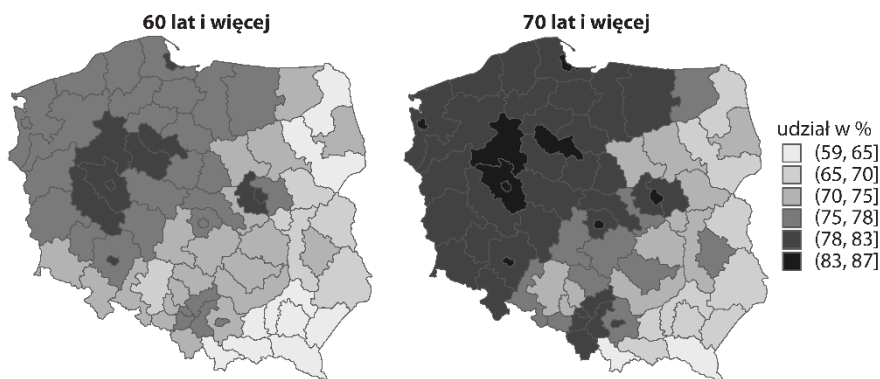


Źródło: obliczenia własne na podstawie: Otwarte Dane (b.r.).

Można zaobserwować wyraźnie wyższy poziom zaszczepienia w regionach z dużymi miastami, a także relatywnie niski odsetek zaszczepienia na wschodzie Polski w porównaniu z centrum i zachodem kraju. Warto zwrócić uwagę na odwrotną zależność pomiędzy regionalnym zróżnicowaniem udziału zaszczepionych (mapa 3) a regionalnym zróżnicowaniem nadmiarowych zgonów (mapa 1, kartogram dotyczący czwartej fali pandemii). Na wschodzie kraju zanotowano niższy odsetek osób zaszczepionych oraz poziom zgonów przewyższający wartość bazową.

Na mapie 4 przedstawiono zróżnicowanie przestrzenne poziomu zaszczepienia osób starszych.

Mapa 4. Regionalne zróżnicowanie udziału zaszczepionych w grupach osób starszych (stan na 13.10.2021 r.)



Źródło: obliczenia własne na podstawie: Otwarte Dane (b.r.).

Analizując powyższe kartogramy, można zauważyć, że niższy odsetek zaszczepienia osób starszych – podobnie jak dla całej populacji – występuje szczególnie na wschodzie kraju. Należy pamiętać, że osoby z grupy 70 lat i więcej znajdują się także w grupie 60 lat i więcej, co niewątpliwie wpłynęło na wyższy poziom zaszczepienia ogółu populacji osób starszych (powyżej 60. roku życia).

Obserwacje zależności pomiędzy poziomem zaszczepienia a liczbą zgonów poczynione na podstawie kartogramów można skwantyfikować za pomocą współczynnika korelacji i modelu regresji liniowej. Zbadano możliwość występowania korelacji odsetka nadmiarowych zgonów podczas trzech fal pandemii i poziomu zaszczepienia ludności w podregionach według stanu na 13 października 2021 r.

Tabl. 2. Wartość współczynnika korelacji odsetka nadmiarowych zgonów i odsetka zaszczepionych osób

Odsetek nadmiarowych zgonów w grupach wieku	Odsetek osób zaszczepionych w grupach wieku			
	20–39 lat	40–59	60–69	70 lat i więcej
Druaga fala pandemii				
20–39 lat	0,17	0,16	0,13	0,14
40–59	0,09	0,06	-0,02	-0,02
60–69	-0,20	-0,17	-0,18	-0,25
70 lat i więcej	-0,54	-0,47	-0,33	-0,37
Trzecia fala pandemii				
20–39 lat	-0,19	-0,20	-0,19	-0,17
40–59	0,23	0,18	0,12	0,18
60–69	0,07	0,06	-0,02	-0,04
70 lat i więcej	-0,21	-0,22	-0,25	-0,28
Czwarta fala pandemii				
20–39 lat	-0,06	-0,05	-0,04	-0,06
40–59	-0,01	-0,05	-0,14	-0,09
60–69	-0,49	-0,50	-0,55	-0,55
70 lat i więcej	-0,69	-0,70	-0,73	-0,71

Źródło: obliczenia własne na podstawie: Eurostat (b.r.) i Otwarte Dane (b.r.).

Z danych w tabl. 2 wynika, że podczas drugiej i trzeciej fali pandemii korelacje odsetka osób zaszczepionych i odsetka nadmiarowych zgonów nie były silne, ale w przypadku czwartej fali pandemii można było zaobserwować silne powiązania pomiędzy odsetkiem nadmiarowych zgonów w starszych grupach wieku a poziomem zaszczepienia ludności we wszystkich badanych grupach. Współczynnik korelacji powyższych zmiennych wyniósł odpowiednio $-0,55$ i $-0,71$ dla grup wieku 60–69 lat oraz 70 lat i więcej, a dla obu tych grup łącznie współczynnik korelacji poziomu zaszczepienia osób w wieku 60 lat i więcej w poszczególnych podregionach i procentowej nadwyżki zgonów dla wszystkich grup wieku łącznie okazał się ujemny i ukształtował się na poziomie $-0,744$. Współzależność w tym zakresie jest więc dość silna. Warto zwrócić uwagę, że we wszystkich falach pandemii w Polsce niższy udział osób zaszczepionych w populacji podregionu był związany głównie z wyższym odsetkiem nadmiarowych zgonów wśród osób w wieku 70 lat i więcej.

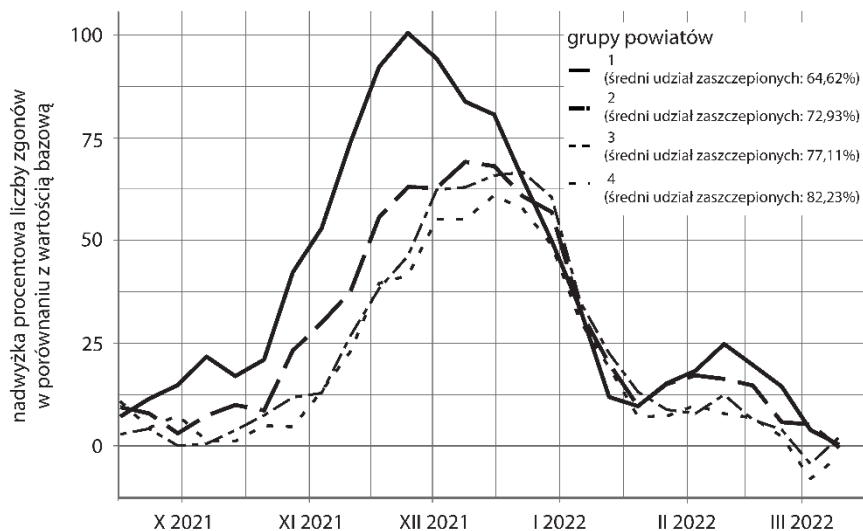
Korelacja nie oznacza zależności przyczynowo-skutkowej, jednak uwzględniając wiele badań wskazujących na efektywność szczepionek przeciwko COVID-19 (zob. np. Andrews i in., 2021; Nordström i in., 2021; Nunes i in., 2021), można przyjąć, że zaobserwowana korelacja nie jest przypadkiem i potwierdza skuteczność szczepień w przeciwdziałaniu zgonom spowodowanym zakażeniem koronawirusem.

Aby zilustrować zależność pomiędzy poziomem zaszczepienia ludności a przebiegiem nadmiarowej umieralności podczas czwartej fali pandemii, podzielono podregiony na cztery grupy w zależności od udziału osób zaszczepionych w wieku 60 lat i więcej:

- grupa 1 – mniej niż 70%;
- grupa 2 – 70–74,99%;
- grupa 3 – 75–79,99%;
- grupa 4 – 80% i więcej.

Następnie obliczono odsetek nadmiarowych zgonów w tych grupach. Wyniki przedstawiono na wyk. 4.

Wykr. 4. Nadmiarowe zgony podczas czwartej fali pandemii COVID-19 z podziałem na grupy podregionów w zależności od udziału zaszczepionych osób w wieku 60 lat i więcej

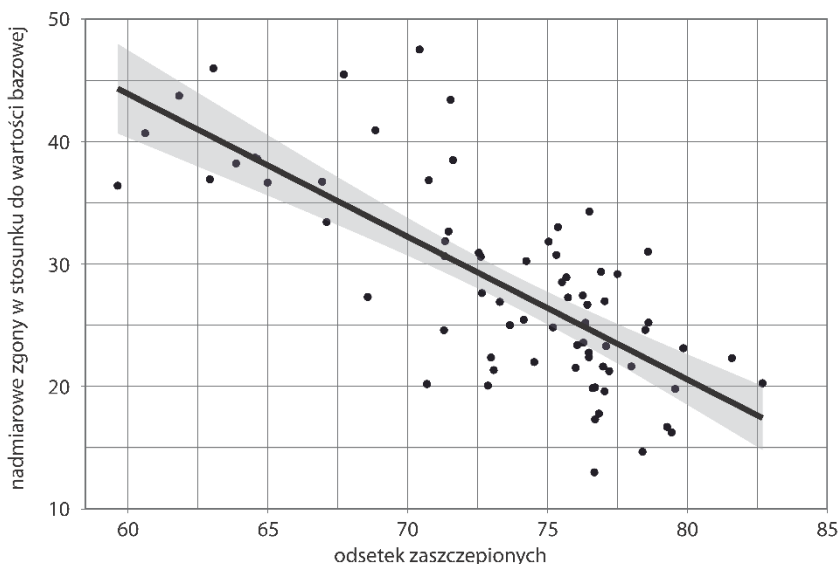


Źródło: obliczenia własne na podstawie: Eurostat (b.r.).

Dane wskazują na znacznie większą nadwyżkę zgonów w stosunku do przewidywanego poziomu, sięgającą ponad 100% w listopadzie 2021 r. w grupie o najniższym udziale zaszczepionych osób w wieku 60 lat i więcej, wynoszącym średnio 64,62%. Jednak w powiatach o najwyższym udziale seniorów, wynoszącym średnio 82,23%, obserwuje się najniższą nadwyżkę zgonów, wynoszącą w szczytowym momencie niecałe 61%. Zależność ta skłania do zbudowania modelu wskazującego, jak poziom zaszczepienia osób w wieku 60 lat i więcej przekładałby się na odsetek nadmiarowych zgonów w stosunku do wartości bazowej.

Z modelu prostej regresji liniowej ($F(1; 71) = 88,23$, $p < 0,001$, $R^2 = 0,554$) wynika, że współczynnik nachylenia linii regresji, która została przedstawiona na wykr. 5, wynosi $-1,17$ ($\pm 0,12$, $t = 9,39$, $p < 0,001$). Oznaczałoby to, że kolejny 1 p.proc. wartości wskaźnika zaszczepienia osób w wieku 60 lat i więcej przekłada się na redukcję nadmiarowych zgonów średnio o 1,17 p.proc. ($\pm 0,12$ p.proc.).

Wykr. 5. Zależność nadmiernej umieralności od zaszczepienia osób w wieku 60 lat i więcej podczas czwartej fali pandemii COVID-19



Źródło: obliczenia własne na podstawie: Eurostat (b.r.) i Otwarte Dane (b.r.).

Wynik uzyskany za pomocą modelu prostej regresji liniowej można przełożyć na liczbę zgonów w skali kraju, których można byłoby uniknąć, gdyby odsetek zaszczepionych osób starszych był wyższy. Zgodnie z przedstawionymi wcześniej obliczeniami w trakcie czwartej fali pandemii w Polsce doszło do 55 371 nadmiarowych zgonów, co stanowiło 27,4% wartości bazowej. Jeżeli przedstawiony model regresji rzeczywiście ilustruje zależność przyczynowo-skutkową, to zaszczepienie kolejnego 1% osób w wieku 60 lat i więcej oznacza o 2354 (± 251) zgony mniej.

4. Podsumowanie

W badaniu przedstawionym w artykule oszacowano liczbę nadmiarowych zgonów w trakcie pandemii COVID-19 dla populacji Polski oraz dla podregionów według klasyfikacji NUTS 3 i grup wieku za pomocą metody Karlinsky'ego i Kobaka (2021).

Według wiedzy autorów jest to pierwsze tego typu zastosowanie wspomnianej metody.

Na podstawie analizy danych na poziomie podregionów stwierdzono, że odsetek nadmiarowych zgonów w trakcie czwartej fali pandemii był ujemnie skorelowany ze stopniem zaszczepienia populacji. Niestety negatywne skutki pandemii w Polsce były ogromne – nadmiarowe zgony między sierpniem 2020 r. a lutym 2022 r. wyniosły ok. 172 000. Z obliczeń przeprowadzonych w ramach badania wynika, że ten problem dotknął przede wszystkim osoby ze starszych grup wieku (60 lat i więcej), co jest spójne z szacunkami dotyczącymi śmiertelności w wyniku zakażenia wirusem SARS-COV-2. Analiza regionalna wskazuje, że podczas czwartej fali pandemii regionalne zróżnicowanie nadmiarowych zgonów było większe niż podczas wcześniejszych fal.

W badaniu podjęto próbę oceny występowania korelacji regionalnego poziomu zaszczepienia i liczby nadmiarowych zgonów. Nadmiarowe zgony okazały się dość silnie skorelowane z poziomem zaszczepienia w grupie osób w wieku 60 lat i więcej, która jest najbardziej narażona na zachorowanie i zgon spowodowany zakażeniem koronawirusem. Przy założeniu występowania przyczynowo-skutkowej zależności liniowej w tym zakresie każdy kolejny 1% osób zaszczepionych w starszych grupach wieku przełożyłby się na spadek liczby zgonów podczas czwartej fali pandemii średnio o 2000 w skali kraju. Prosta regresja liniowa na zebranych danych dostarcza dodatkowej przesłanki do stwierdzenia, że szczepienia przeciwko COVID-19 skutecznie przeciwdziałają zgonom spowodowanym przez tę chorobę.

Bibliografia

- Alling, D. W., Blackwelder, W. C., Stuart-Harris, C. H. (1981). A study of excess mortality during influenza epidemics in the United States, 1968–1976. *American Journal of Epidemiology*, 113(1), 30–43. <https://doi.org/10.1093/oxfordjournals.aje.a113063>.
- Andrews, N., Tessier, E., Stowe, J., Gower, C., Kirsebom, F., Simmons, R., Gallagher, E., Chand, M., Brown, K., Ladhani, S. N., Ramsay, M., Bernal, J. L. (2021). *Vaccine effectiveness and duration of protection of Comirnaty, Vaxzevria and Spikevax against mild and severe COVID-19 in the UK*. Medrxiv. <https://doi.org/10.1101/2021.09.15.21263583>.
- Beaney, T., Clarke, J. M., Jain, V., Golestaneh, A. K., Lyons, G., Salman, D., Majeed, A. (2020). Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide?. *Journal of the Royal Society of Medicine*, 113(9), 329–334. <https://doi.org/10.1177/0141076820956802>.
- Böttcher, L., D’Orsogna, M. R., Chou, T. (2021). Using excess deaths and testing statistics to determine COVID-19 mortalities. *European Journal of Epidemiology*, 36(5), 545–558. <https://doi.org/10.1007/s10654-021-00748-2>.
- Ciminelli, G., Garcia-Mandicó, S. (2020). COVID-19 in Italy: An Analysis of Death Registry Data. *Journal of Public Health*, 42(4), 723–730. <https://doi.org/10.1093/pubmed/fdaa165>.

- Collins, S. D. W., Lehmann, J. (1953). *Excess Deaths From Influenza and Pneumonia and From Important Chronic Diseases During Epidemic Periods, 1918–1951*. Federal Security Agency, Public Health Service.
- Czerwiński, A. (2021). *Nadmierna śmiertelność w Polsce w 2020 r.* (Working Paper nr 1). https://pie.net.pl/wp-content/uploads/2018/07/PIE-WP_1-2021.pdf.
- Eurostat. (b.r.). *Deaths by week and NUTS 3 region* [zbiór danych]. Pobrane 13 grudnia 2022 r. z https://ec.europa.eu/eurostat/databrowser/view/demo_r_mwk3_t/default/table?lang=en&category=demo.demomwk.
- Ghisolfi, S., Almás, I., Sandefur, J. C., von Carnap, T., Heitner, J., Bold, T. (2020). Predicted COVID-19 fatality rates based on age, sex, comorbidities and health system capacity. *BMJ Global Health*, 5(9), 1–8. <http://dx.doi.org/10.1136/bmjgh-2020-003094>.
- Główny Urząd Statystyczny. (2022). *Zgony według tygodni oraz wieku, płci zmarłych i 73 podregionów w 2023 roku* [zbiór danych]. <https://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/zgony-wedlug-tygodni,39,2.html>.
- Green, M. S., Nitzan, D., Schwartz, N., Niv, Y., Peer, V. (2021). Sex differences in the case-fatality rates for COVID-19 – a comparison of the age-related differences and consistency over seven countries. *PLoS ONE*, 16(4), 1–13. <https://doi.org/10.1371/journal.pone.0250523>.
- Haider, N., Hasan, M. N., Guitian, J., Khan, R. A., McCoy, D., Ntoumi, F., Dar, O., Ansumana, R., Uddin, M. J., Zumla, A., Kock, R. A. (2023). The disproportionate case-fatality ratio of COVID-19 between countries with the highest vaccination rates and the rest of the world. *IJID Regions*, 6, 159–166. <https://doi.org/10.1016/j.ijregi.2023.01.011>.
- Harris, J. E. (2022). COVID-19 Incidence and hospitalization during the delta surge were inversely related to vaccination coverage among the most populous U.S. Counties. *Health Policy and Technology*, 11(2), 1–11. <https://doi.org/10.1016/j.hlpt.2021.100583>.
- Housworth, J., Langmuir, A. D. (1974). Excess mortality from epidemic influenza, 1957–1966. *American Journal of Epidemiology*, 100(1), 40–48. <https://doi.org/10.1093/oxfordjournals.aje.a112007>.
- Islam, N., Shkolnikov, V. M., Acosta, R. J., Klimkin, I., Kawachi, I., Irizarry, R. A., Alicandro, G., Khunti, K., Yates, T., Jdanov, D. A., White, M., Lewington, S., Lacey, B. (2021). Excess deaths associated with Covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. *BMJ*, 373(1137), 1–14. <https://doi.org/10.1136/bmj.n1137>.
- Jabłońska, K., Aballéa, S., Toumi, M. (2021). The real-life impact of vaccination on COVID-19 mortality in Europe and Israel. *Public Health*, 198, 230–237. <https://doi.org/10.1016/j.puhe.2021.07.037>.
- Jaroszewska, E., Oldak, M. (2022). Ochrona zdrowia, zdrowie i życie ludzkie jako kluczowe obszary zagrożenia w czasie pandemii COVID-19 w Polsce. *Studia Politologiczne*, 65, 29–57. <https://doi.org/10.33896/SPolit.2022.65.2>.
- Jassat, W., Mudara, C., Ozougwu, L., Tempia, S., Blumberg, L., Davies, M.-A., Pillay, Y., Carter, T., Morewane, R., Wolmarans, M., von Gottberg, A., Bhiman, J. N., Walaza, S., Cohen, C., DATACOV author group. (2021). Difference in mortality among individuals admitted to hospital with COVID-19 during the first and second waves in South Africa: a cohort study. *The Lancet Global Health*, 9(9), e1216–e1225. [https://doi.org/10.1016/S2214-109X\(21\)00289-8](https://doi.org/10.1016/S2214-109X(21)00289-8).
- Karlinsky, A., Kobak, D. (2021). Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *Elife*, 10, 1–21. <https://doi.org/10.7554/eLife.69336>.

- Konstantinoudis, G., Cameletti, M., Gómez-Rubio, V., Gómez, I. L., Pirani, M., Baio, G., Larrauri, A., Riou, J., Egger, M., Vineis, P., Blangiardo, M. (2022). Regional excess mortality during the 2020 COVID-19 pandemic in five European countries. *Nature Communications*, 13(1), 1–11. <https://doi.org/10.1038/s41467-022-28157-3>.
- Kontis, V., Bennett, J. E., Rashid, T., Parks, R. M., Pearson-Stuttard, J., Guillot, M., Asaria, P., Zhou, B., Battaglini, M., Corsetti, G., McKee, M., Di Cesare, M., Mathers, C. D., Ezzati, M. (2020). Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. *Nature Medicine*, 26(12), 1919–1928. <https://doi.org/10.1038/s41591-020-1112-0>.
- Korang, S. K., von Rohden, E., Veroniki, A. A., Ong, G., Ngalamika, O., Siddiqui, F., Juul, S., Nielsen, E. E., Feinberg, J. B., Petersen, J. J., Legart, C., Kokogho, A., Maagaard, M., Klingenberg, S., Thabane, L., Bardach, A., Ciapponi, A., Thomsen, A. R., Jakobsen, J. C., Gluud, C. (2022). Vaccines to prevent COVID-19: A living systematic review with Trial Sequential Analysis and network meta-analysis of randomized clinical trials. *PLoS ONE*, 17(1), 1–23. <https://doi.org/10.1371/journal.pone.0260733>.
- Kuropka, I., Rossa, A., Wróblewska, W., Wojtyński, B., Śleszyński, P. (2021). *Pandemia i jej skutki zdrowotne i demograficzne*. Komitet Nauk Demograficznych PAN. <https://doi.org/10.24425/140474>.
- Levin, A. T., Hanage, W. P., Owusu-Boaitey, N., Cochran, K. B., Walsh, S. P., Meyerowitz-Katz, G. (2020). Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *European Journal of Epidemiology*, 35(12), 1123–1138. <https://doi.org/10.1007/s10654-020-00698-1>.
- Levitt, M., Zonta, F., Ioannidis, J. P. A. (2022). Comparison of pandemic excess mortality in 2020–2021 across different empirical calculations. *Environmental Research*, 213, 1–11. <https://doi.org/10.1016/j.envres.2022.113754>.
- Lewandowski, P., Madoń, K. (2022). *Skuteczność szczepień przeciw COVID-19 w Polsce* (IBS Policy Paper 1/2022). <https://ibs.org.pl/publications/skutecznosc-szczepien-przeciw-covid-19-w-polsce/>.
- Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Gavrilov, D., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., Roser, M. (2020). *Coronavirus pandemic (COVID-19)*. <https://ourworldindata.org/coronavirus>.
- McLaughlin, J. M., Khan, F., Pugh, S., Swerdlow, D. L., Jodar, L. (2022). County-level vaccination coverage and rates of COVID-19 cases and deaths in the United States: An ecological analysis. *The Lancet Regional Health – Americas*, 9, 1–11. <https://doi.org/10.1016/j.lana.2022.100191>.
- Murkowski, R. (2021). Nadmierna umieralność w Polsce podczas pandemii COVID-19 w 2020 roku. *Wiadomości Statystyczne. The Polish Statistician*, 66(7), 7–23. <https://doi.org/10.5604/01.3001.0015.0351>.
- Ning, C., Wang, H., Wu, J., Chen, Q., Pei, H., Gao, H. (2022). The COVID-19 Vaccination and Vaccine Inequity Worldwide: An Empirical Study Based on Global Data. *International Journal of Environmental Research and Public Health*, 19(9), 1–13. <https://doi.org/10.3390/ijerph19095267>.

- Nordström, P., Ballin, M., Nordström, A. (2022). Risk of infection, hospitalisation, and death up to 9 months after a second dose of COVID-19 vaccine: a retrospective, total population cohort study in Sweden. *The Lancet*, 399(10327), 814-823. [https://doi.org/10.1016/S0140-6736\(22\)00089-7](https://doi.org/10.1016/S0140-6736(22)00089-7).
- Nunes, B., Rodrigues, A. P., Kislaya, I., Cruz, C., Peralta-Santos, A., Lima, J., Leite, P. P., Sequeira, D., Dias, C. M., Machado, A. (2021). mRNA vaccine effectiveness against COVID-19-related hospitalisations and deaths in older adults: A cohort study based on data linkage of national health registries in Portugal, February to August 2021. *Eurosurveillance*, 26(38), 1-7. <https://doi.org/10.2807/1560-7917.ES.2021.26.38.2100833>.
- Organisation for Economic Co-operation and Development. (2021). Excess mortality. W: *Health at a Glance 2021: OECD Indicators* (s. 84-85). OECD Publishing. <https://doi.org/10.1787/ec2de914-en>.
- Otwarte Dane. (b.r.). *Odsetek osób zaszczepionych przeciwko COVID-19 w gminach* [zbiór danych]. Pobrane 22 lutego 2022 r. z <https://dane.gov.pl/pl/dataset/2476>.
- Papadopoulos, V. P., Emmanouilidou, A., Yerou, M., Panagaris, S., Souleiman, C., Varela, D., Avramidou, P., Melissopoulou, E., Pappas, C., Iliadou, Z., Piperopoulos, I., Somadis, V., Partsalidis, A., Metaxa, E., Feresiadis, I., Filippou, D. (2022). SARS-CoV-2 Vaccination Coverage and Key Public Health Indicators May Explain Disparities in COVID-19 Country-Specific Case Fatality Rate Within European Economic Area. *Cureus*, 14(3). <https://doi.org/10.7759/cureus.22989>.
- Reichert, T. A., Simonsen, L., Sharma, A., Pardo, S. A., Fedson, D. S., Miller, M. A. (2004). Influenza and the Winter Increase in Mortality in the United States, 1959-1999. *American Journal of Epidemiology*, 160(5), 492-502. <https://doi.org/10.1093/aje/kwh227>.
- Riffe, T., Acosta, E., COVERAGE-DB team. (2021). Data Resource Profile: COVERAGE-DB: a global demographic database of COVID-19 cases and deaths. *International Journal of Epidemiology*, 50(2), 390-390f. <https://doi.org/10.1093/ije/dyab027>.
- Simonsen, L., Spreeuwenberg, P., Lustig, R., Taylor, R. J., Fleming, D. M., Kroneman, M., Van Kerkhove, M. D., Mounts, A. W., Paget, W. J., the GLaMOR Collaborating Teams. (2013). Global Mortality Estimates for the 2009 Influenza Pandemic from the GLaMOR Project: A Modeling Study. *PLOS Medicine*, 10(11). <https://doi.org/10.1371/journal.pmed.1001558>.
- Śleszyński, P. (2022). Wpływ pandemii COVID-19 na przestrzenne struktury demograficzne i osadnicze Polski – wstępne wnioski i hipotezy. W: K. Markowski (red.), *Konsekwencje zmian demograficznych. Materiały z III Kongresu Demograficznego. Część 1* (s. 32-52). Rządowa Rada Ludnościowa, Główny Urząd Statystyczny. <https://kd.stat.gov.pl/publikacje>.
- Śleszyński, P., Niedzielski, M. (2018). Zastosowanie danych telemetrycznych do szacunku ludności dziennej i nocnej w Warszawie. *Czasopismo Geograficzne*, 89(1/2), 43-60.
- Taboada, M., González, M., Alvarez, A., Eiras, M., Costa, J., Álvarez, J., Seoane-Pillado, T. (2021). First, second and third wave of COVID-19. What have we changed in the ICU management of these patients?. *Journal of Infection*, 82(6), e14-e15. <https://doi.org/10.1016/j.jinf.2021.03.027>.
- Zheng, C., Shao, W., Chen, X., Zhang, B., Wang, G., Zhang, W. (2022). Real-world effectiveness of COVID-19 vaccines: A literature review and meta-analysis. *International Journal of Infectious Diseases*, 114, 252-260. <https://doi.org/10.1016/j.ijid.2021.11.009>.

41st International Conference MSA'2023

W dniach 6–8 listopada 2023 r. na Uniwersytecie Łódzkim (UŁ) odbyła się 41st International Conference MSA'2023 – kolejna z corocznych konferencji poświęconych metodom analizy statystycznej. Głównym organizatorem wydarzenia była Katedra Metod Statystycznych UŁ, a współorganizatorami: Instytut Statystyki i Demografii UŁ, łódzki oddział Polskiego Towarzystwa Statystycznego oraz Urząd Statystyczny w Łodzi (US w Łodzi). Honorowy patronat nad konferencją objęli: prof. dr hab. Elżbieta Żądzińska, rektor UŁ, dr Dominik Rozkrut, prezes Głównego Urzędu Statystycznego (GUS), oraz Komitet Statystyki i Ekonometrii Polskiej Akademii Nauk. Partnerami konferencji były firmy StatSoft Polska i LabMasters. Komitetowi Naukowemu przewodniczył prof. dr hab. Czesław Domański, a Komitetowi Organizacyjnemu – dr hab. Alina Jędrzejczak, prof. UŁ. Funkcję sekretarzy naukowych konferencji pełniły dr Kamila Trzcicka i dr Elżbieta Zalewska.

Konferencję otworzyli: Alina Jędrzejczak, Czesław Domański i dr Piotr Gabrielczak, prodziekan ds. współpracy z zagranicą Wydziału Ekonomiczno-Socjologicznego UŁ. W wydarzeniu wzięło udział 100 osób, w tym przedstawiciele ośrodków akademickich z Gdańska, Katowic, Krakowa, Łodzi, Poznania, Szczecina, Warszawy, Wrocławia oraz GUS i US w Łodzi, a także goście z Czech, Włoch i Ukrainy. Podczas wydarzenia odbyło się 18 sesji plenarnych i panelowych, na których wygłoszono łącznie 60 referatów.



Fot. E. Zalewska

Otwarcie konferencji (od lewej): Alina Jędrzejczak, Czesław Domański i Piotr Gabrielczak

Pierwszego dnia konferencji odbyło się osiem sesji. Nazwiska prelegentów i tytuły wystąpień z pierwszych siedmiu sesji przedstawiono w zestawieniu 1.

Zestawienie 1. Referaty wygłoszone podczas pierwszego dnia konferencji

Autor, afiliacja	Tytuł
Sesja I plenarna	
przewodniczący: Bronisław Ceranka, Uniwersytet Przyrodniczy w Poznaniu	
Jerzy Tadeusz Kowaleski, Uniwersytet Łódzki	Maciej z Miechowa (1457–1523) – statystyk państwowostwa
Czesław Domański, Uniwersytet Łódzki	Tadeusz Pilat (1844–1923) – pierwsze szacunki parametrów na podstawie próby
Krzysztof Jajuga, Uniwersytet Ekonomiczny we Wrocławiu	Prof. Stanisława Bartosiewicz – wspomnienie
Edyta Rutkowska-Tomaszewska, Uniwersytet Wrocławski	Prof. Jerzy Jakubczyc – wspomnienie
Krzysztof Jajuga, Uniwersytet Ekonomiczny we Wrocławiu	
Czesław Domański, Uniwersytet Łódzki	Wiesław Łagodziński – wspomnienie
Sesja II plenarna (wykład przygotowany na zaproszenie)	
przewodniczący: Krzysztof Jajuga, Uniwersytet Ekonomiczny we Wrocławiu	
Przemysław Biecek, Politechnika Warszawska, Uniwersytet Warszawski	Explanatory Models Analysis
Sesja IIIA	
przewodniczący: Włodzimierz Okrasa, Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie	
Elżbieta Gołata, Uniwersytet Ekonomiczny w Poznaniu	Konsekwencje pomiaru ludności w ocenie płodności kobiet w Polsce
Grażyna Trzpiot, Uniwersytet Ekonomiczny w Katowicach	Wybrane środowiskowe implikacje na jakość życia mieszkańców miast
Beata Bieszk-Stolorz, Uniwersytet Szczeciński	Wielowymiarowa ocena warunków życiowych i dobrobytu ludzi starszych w Unii Europejskiej
Krzysztof Dmytrów, Uniwersytet Szczeciński	
Bronisław Ceranka, Uniwersytet Przyrodniczy w Poznaniu	Estymacja nieznanych miar obiektów w chemicznym układzie wagowym ze skorelowanymi błędami pomiarów
Małgorzata Graczyk, Uniwersytet Przyrodniczy w Poznaniu	
Sesja IIIB	
przewodniczący: Eugeniusz Gatnar, Uniwersytet Ekonomiczny w Katowicach	
Jacek Wesołowski, Główny Urząd Statystyczny, Politechnika Warszawska	Discrete parametric statistical graphical models
Marcin Szymkowiak, Uniwersytet Ekonomiczny w Poznaniu	Inference for non-probability samples using calibration estimators for totals and quantiles
Maciej Beręsewicz, Uniwersytet Ekonomiczny w Poznaniu	
Andrzej Dudek, Uniwersytet Ekonomiczny we Wrocławiu	Image and audio neural style transfer – statistical foundations
Second Bwanakare, Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie	Energy efficiency forecast based on sparse and conflicting Information

Zestawienie 1. Referaty wygłoszone podczas pierwszego dnia konferencji (dok.)

Autor, afiliacja	Tytuł
Sesja IVA	
przewodnicząca: Małgorzata Markowska, Uniwersytet Ekonomiczny we Wrocławiu	
Agnieszka Sompolska-Rzechuła, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie	Realizacja celów zrównoważonego rozwoju – Polska na tle krajów UE
Iwona Bąk, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie	
Elżbieta Antczak, Uniwersytet Łódzki	Zielone miejsca pracy w Polsce – analiza przekrojowo-czasowa
Artur Gajdos, Uniwersytet Łódzki	
Marta Kuc-Czarnecka, Politechnika Gdańska	Czynniki utrudniające i przyspieszające realizację SDG7 w krajach UE
Iwona Markowicz, Uniwersytet Szczeciński	
Agnieszka Sompolska-Rzechuła, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie	
Alina Stundziene, Kaunas University of Technology	
Sesja IVB	
przewodnicząca: Beata Bieszk-Stolorz, Uniwersytet Szczeciński	
Artur Zaborski, Uniwersytet Ekonomiczny we Wrocławiu	Analiza unfolding obiektów symbolicznych na przykładzie zewnętrznej oceny reklam samochodów
Marcin Pełka, Uniwersytet Ekonomiczny we Wrocławiu	
Mariusz Doszyń, Uniwersytet Szczeciński	Wiedza ekspercka w ekonometrycznych modelach wyceny nieruchomości
Dominik Krężolek, Uniwersytet Ekonomiczny w Katowicach	Semiparametryczna optymalizacja portfela inwestycyjnego z wykorzystaniem funkcji kopula. Zastosowanie na rynku metali i surowców energetycznych
Szymon Stolarski, Uniwersytet Ekonomiczny w Katowicach	

Ósmą sesję, której przewodniczył prof. dr hab. Jerzy Tadeusz Kowaleski, poświęcono 55-leciu pracy na UŁ i jubileuszowi 80. urodzin Czesława Domańskiego. Alina Jędrzejczak (UŁ) zaprezentowała sylwetkę jubilata w wystąpieniu *Czesław Domański – 55 lat dla statystyki*, a Mirosław Szreder (Uniwersytet Gdański) wygłosił referat *Precyzja i uniwersalność języka statystyki jako elementy zachwyty nad nią w pracach i działalności Profesora Czesława Domańskiego*.

Drugiego dnia konferencji odbyło się siedem sesji. Tematy wystąpień zebrano w zestawieniu 2.

Zestawienie 2. Referaty wygłoszone podczas drugiego dnia konferencji

Autor, afiliacja	Tytuł
Sesja VA	
przewodniczący: Grzegorz Kończak, Uniwersytet Ekonomiczny w Katowicach	
Tomasz Żądło, Uniwersytet Ekonomiczny w Katowicach	Bootstrapowa ocena dokładności oszacowań dla prób nieprostych i nielosowych
Stanisław Jaworski, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie	Optymalny rozmiar próby w modelu trójkątnym dla pytań drażliwych

Zestawienie 2. Referaty wygłoszone podczas drugiego dnia konferencji (cd.)

Autor, afiliacja	Tytuł
Sesja VA (dok.)	
Małgorzata Krzciuk, Uniwersytet Ekonomiczny w Katowicach	O estymatorach MSE predyktorów typu plugin dla liniowych modeli mieszanych ze skorelowanymi efektami losowymi
Justyna Majewska, Uniwersytet Ekonomiczny w Katowicach	O odpornym na obserwacje odstające prognozowaniu umieralności
Sesja VB przewodnicząca: Grażyna Trzpiot, Uniwersytet Ekonomiczny w Katowicach	
Katarzyna Frodyma, Uniwersytet Ekonomiczny w Krakowie	Metoda DEA w analizie efektywności energetycznej
Monika Zielińska-Sitkiewicz, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie	Analiza konsumpcji energii elektrycznej w Polsce z wykorzystaniem modeli predykcji i sieci neuronowych
Mariola Chrzanowska, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie	
Klaudia Lenart, Szkoła Doktorska Uniwersytetu Ekonomicznego w Katowicach	Zastosowanie Google Trends jako źródła danych w modelach statystycznych
Magdalena Kawecka, Uniwersytet Ekonomiczny w Katowicach	Kwalifikacje i umiejętności a kompetencje przyszłości u dorosłych Polaków
Sesja VIA przewodniczący: Józef Pocięcha, Uniwersytet Ekonomiczny w Krakowie	
Eugeniusz Gatnar, Uniwersytet Ekonomiczny w Katowicach	Jak włączyć koszty użytkowania mieszkań przez właścicieli do koszyka inflacyjnego?
Natalia Pawelec, Uniwersytet Łódzki	Zastosowanie tranzytowych indyktorów Montgomery'ego do analizy danych skanowanych
Jacek Białek, Uniwersytet Łódzki	
Adam Juszcak, Uniwersytet Łódzki	Zastosowanie danych skanowanych w pomiarze dynamiki cen
Natalia Pawelec, Uniwersytet Łódzki	Porównanie indyktorów Benneta i Montgomery'ego z wykorzystaniem danych skanowanych
Sesja VIB przewodniczący: Józef Dziechciarz, Uniwersytet Ekonomiczny we Wrocławiu	
Maciej Kot, Politechnika Gdańska	Aversion to inequality and to rank inequality in European Union countries
Sławomir Śmiech, Uniwersytet Ekonomiczny w Krakowie	Multiple faces of poverty. Exploring housing-costs-induced energy poverty in Central and Eastern Europe
Lilia Karpińska, Uniwersytet Ekonomiczny w Krakowie	
Svitlana Chugaievska, Krakowska Akademia im. Andrzeja Frycza Modrzewskiego	Integration of forced Ukrainian migrants into new European spaces: results of the conducted survey
Krzysztof Dmytrów, Uniwersytet Szczeciński	Comparison of results for selected multi-criteria methods of supporting location selection during order picking
Sesja VII plenarna (wykład przygotowany na zaproszenie) przewodniczący: Mirosław Szreder, Uniwersytet Gdański	
Francesca Greselin, University of Milano-Bicocca	Measuring income inequality via percentile relativeities

Zestawienie 2. Referaty wygłoszone podczas drugiego dnia konferencji (dok.)

Autor, afiliacja	Tytuł
Sesja VIIIA	
przewodniczący: Marek Walesiak, Uniwersytet Ekonomiczny we Wrocławiu	
Andrzej Sokołowski, Collegium Humanum – Szkoła Główna Menedżerska Małgorzata Markowska, Uniwersytet Ekonomiczny we Wrocławiu Katarzyna Budny, Uniwersytet Ekonomiczny w Krakowie	Modalna median
Piotr Sulewski, Uniwersytet Pomorski w Słupsku	O wykrywalności rozkładów asymetrycznych odbiegających od normalności z powodu małej skończoności
Grzegorz Sitek, Uniwersytet Ekonomiczny w Katowicach	Rozkład sumy rozkładu normalnego i rozkładu gamma
Kamila Hasilova, University of Defence Ivana Horova, Masaryk University David Valis, University of Defence Stanislav Zamecnik, Masaryk University	Kompleksowa eksploracja zupełnej metody krzyżowego uwiarygodnienia dla danych okrągłych
Sesja VIIIB	
przewodniczący: Tomasz Żądło, Uniwersytet Ekonomiczny w Katowicach	
Józef Pociecha, Uniwersytet Ekonomiczny w Krakowie Paweł Lula, Uniwersytet Ekonomiczny w Krakowie	Problemy modelowania regresyjnego narzędziami wnioskowania statystycznego oraz narzędziami uczenia maszynowego
Grzegorz Kończak, Uniwersytet Ekonomiczny w Katowicach Martyna Kosińska, Uniwersytet Ekonomiczny w Katowicach	O wnioskowaniu statystycznym dla danych w wielowymiarowych tablicach wielodzielczych
Piotr Bolesław Nowak, Uniwersytet Wrocławski	Estymacja w modelu Coxa z pogrupowanymi czasami życia
Dominik Mielczarek, Centrum Badawczo-Rozwojowe OPEGIEKA Krzysztof Najman, Uniwersytet Gdański	Zastosowanie konwolucyjnych sieci neuronowych w szacowaniu szkód pohuraganowych na przykładzie nadleśnictwa Stare Jabłonki

Atrakcją tego dnia przygotowaną przez organizatorów dla uczestników konferencji była wizyta w Monopolis – zabytkowym budynku w centrum Łodzi.

Trzeci dzień konferencji rozpoczął się szkoleniem *Zaawansowana wizualizacja danych w języku R i pakiecie ggplot2* poprowadzonym przez Piotra Ćwiakowskiego z firmy LabMasters.

Następnie odbyły się trzy sesje równoległe (zestawienie 3).

Zestawienie 3. Referaty wygłoszone podczas trzeciego dnia konferencji

Autor, afiliacja	Tytuł
Sesja IXA	
przewodnicząca: Danuta Strahl, Uniwersytet Ekonomiczny we Wrocławiu, filia w Jeleniej Górze	
Radosław Pietrzyk, Uniwersytet Ekonomiczny we Wrocławiu, Empirica	Zależność miar płynności od poziomu cen i stóp zwrotu na rynku kryptowalut
Paweł Rokita, Uniwersytet Ekonomiczny we Wrocławiu, Empirica	
Aneta Ptak-Chmielewska, Szkoła Główna Handlowa w Warszawie	Wpływ ryzyka klimatycznego na ryzyko kredytowe
Paweł Kopciuszewski, Szkoła Główna Handlowa w Warszawie, Vistula University	
Wioletta Grzenda, Szkoła Główna Handlowa w Warszawie	Zalety metody Shapley additive explanations w analizie czynników determinujących spłatę stałych zobowiązań finansowych wobec nieplanowanych zdarzeń
Olga Momot, Szkoła Główna Handlowa w Warszawie	
Sesja IXB	
przewodnicząca: Elżbieta Gołata, Uniwersytet Ekonomiczny w Poznaniu	
Michał Bernard Pietrzak, Politechnika Gdańska	Zastosowanie modelu ilorazu potencjałów na poziomie regionalnym z wykorzystaniem odległości ekonomicznej
Ewa Feder-Sempach, Uniwersytet Łódzki	Zmiany strukturalne CAPM w przekroju krajów oraz branż. Wpływ pandemii COVID-19 na firmy europejskie
Piotr Szczepocki, Uniwersytet Łódzki	
Maciej Oesterreich, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie	Wykluczenie cyfrowe a realizacja wybranych celów zrównoważonego rozwoju w krajach UE w 2021 roku
Anna Majdzińska, Uniwersytet Łódzki	Ocena potencjału demograficznego obszarów Polski
Sesja IXC	
przewodniczący: Andrzej Sokołowski, Collegium Humanum – Szkoła Główna Menedżerska	
Józef Dziechciarz, Uniwersytet Ekonomiczny we Wrocławiu	Peoples' with disabilities households' material situation. Subjective assessment
Marta Dziechciarz, Uniwersytet Ekonomiczny we Wrocławiu	
Błażej Kochański, Politechnika Gdańska	The shape of an ROC curve in the evaluation of credit scoring models
Ryszard Szupiluk, Szkoła Główna Handlowa w Warszawie	Clustering financial time series using latent component analysis
Paweł Rubach, Szkoła Główna Handlowa w Warszawie	
Jerzy Korzeniewski, Uniwersytet Łódzki	An algorithm for determining the sentiment of Polish-language texts using translation into English
Adam Idczak, Uniwersytet Łódzki	
Sesja X plenarna (wykłady przygotowane na zaproszenie)	
przewodniczący: Mirosław Krzyśko, Uniwersytet im. Adama Mickiewicza w Poznaniu	
Dominik Rozkrut, Główny Urząd Statystyczny	Academic partnerships – the case of Statistics Poland
Tadeusz Trzaskalik, Uniwersytet Ekonomiczny w Katowicach	Reference points in multiobjective, multistage decision making

Na zakończenie Czesław Domański i Alina Jędrzejczak podsumowali obrady oraz złożyli podziękowania uczestnikom, współorganizatorom i partnerom konferencji.

Kolejna konferencja MSA odbędzie się w dniach 4–6 listopada 2024 r.

Elżbieta Zalewska

Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny, Polska /
University of Lodz, Faculty of Economics and Sociology, Poland

Kamila Trzcńska

Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny, Polska /
University of Lodz, Faculty of Economics and Sociology, Poland

NOWOŚCI WYDAWNICZE
W ZBIORACH CENTRALNEJ BIBLIOTEKI STATYSTYCZNEJ
NEW PUBLICATIONS
IN THE CENTRAL STATISTICAL LIBRARY RESOURCES

W zbiorach Centralnej Biblioteki Statystycznej im. Stefana Szulca dostępne są następujące, warte polecenia publikacje:

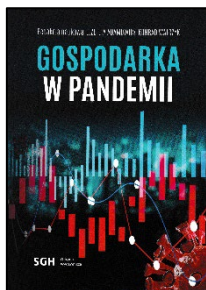
The resources of the Stefan Szulc Central Statistical Library offer the following, highly recommendable publications:

Redakcja naukowa: Elżbieta Adamowicz, Konrad Walczyk

Gospodarka w pandemii

Economy in a pandemic

Publikacja poświęcona procesom gospodarczym zachodzącym w latach 2020–2021, ze szczególnym uwzględnieniem badań koniunktury w poszczególnych sektorach gospodarki.



Język: polski

Language: Polish

Wydawnictwo/Publisher: Oficyna Wydawnicza SGH

Miejsce i rok wydania / Place and year of publication: Warszawa 2022

Liczba stron / Number of pages: 203

Wybuch pandemii COVID-19 istotnie wpłynął na poziom dochodów i wydatków sektora finansów publicznych. Z jednej strony nastąpił wzrost nakładów na ochronę zdrowia, policję, służby sanitarne, obsługującą działania antykryzysowe administrację publiczną, a także przystosowanie instytucji do pracy zdalnej. Z drugiej strony w okresie pandemii zauważono ograniczenie wydatków związanych ze zużyciem wody, energii elektrycznej i artykułów biurowych. Co ważne, odnotowane w tym czasie zmiany procesów gospodarczych miały nietypowy charakter i odbiegały od ustalonego dla polskiej gospodarki wzorca zmian sezonowych i cyklicznych.

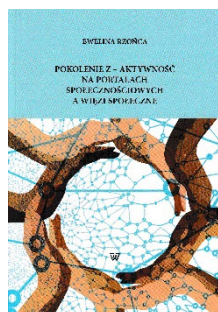
Autorzy opracowania przedstawili przebieg zdarzeń gospodarczych za pomocą danych jakościowych i ilościowych, korzystając z raportów z badań koniunktury prowadzonych przez Instytut Rozwoju Gospodarczego Szkoły Głównej Handlowej i informacji Głównego Urzędu Statystycznego, a także z innych źródeł, m.in. sprawozdań z wykonania ustawy budżetowej za 2020 r., raportu zarządu z działalności Grupy Kapitałowej Banku Gospodarstwa Krajowego, publikacji Zakładu Ubezpieczeń Społecznych i Polskiego Instytutu Ekonomicznego oraz komunikatów ministerstw.

Monografia składa się z dziewięciu rozdziałów. W pierwszym omówiono politykę antykryzysową realizowaną w Polsce w okresie pandemii. Zamieszczono kalendarium najważniejszych wydarzeń oraz wykaz restrykcji społecznych i gospodarczych wprowadzonych przez władze centralne. Zaprezentowano także szacunki dotyczące wsparcia antykryzysowego w 2020 r. W drugim przeanalizowano zmiany koniunktury gospodarczej w Polsce w okresie poprzedzającym wybuch pandemii, w jej trakcie i w czasie stabilizacji gospodarki. Rozdział trzeci dotyczy wpływu pandemii na polski przemysł przetwórczy. Podane przez autorów wskaźniki koniunktury wskazują na znaczne ograniczenia podaży i popytowe w tym sektorze, a także problemy ze znalezieniem pracowników oraz z zaopatrzeniem w surowce i materiały. W rozdziale czwartym scharakteryzowano branżę budowlaną w okresie pandemii, zarówno całościowo, jak i w podziale na sektor prywatny i publiczny, ze szczególnym uwzględnieniem mikrofirm. W piątym przedstawiono zmiany koniunktury w polskim handlu. W szóstym określono, głównie na podstawie danych GUS, grupy podmiotów i rodzaje działalności usługowej, które najbardziej ucierpiały wskutek kryzysu spowodowanego wybuchem pandemii, a także omówiono zastosowane środki zaradcze. Rozdział siódmy poświęcono sektorowi bankowemu w Polsce w latach 2020–2021. Zwrócono uwagę na zachowania klientów instytucji finansowych, a zwłaszcza poruszono kwestie depozytów i preferencji w zaciąganiu kredytów. W rozdziale ósmym przedstawiono m.in. zmiany wydatków gospodarstw domowych w okresie pandemii, a w dziewiątym – metodę modelowania miękkiego jako podstawę analizy procesów kształtujących koniunkturę w polskim przemyśle przetwórczym.

Ewelina Rzońca

Pokolenie Z – aktywność na portalach społecznościowych a więzi społeczne ***Generation Z – activity on social networking sites and social bonds***

Monografia zawierająca wyniki badań nad korzystaniem z portali społecznościowych przez młodzież licealną, a także budowaniem przez nią więzi z rodziną, rówieśnikami z klasy oraz członkami grup pozaszkolnych.



Język: polski

Language: Polish

Wydawnictwo/Publisher: Wydawnictwo Naukowe Uniwersytetu Kardynała Stefana Wyszyńskiego

Miejsce i rok wydania / Place and year of publication: Warszawa 2023

Liczba stron / Number of pages: 141

Internet umożliwia szybką komunikację i wymianę informacji na odległość i w czasie rzeczywistym. Korzystanie z portali społecznościowych to sposób na auto-prezentację, zawieranie nowych znajomości, podtrzymywanie relacji społecznych

i śledzenie życia innych. W przedstawionym badaniu postawiono pytanie o to, jak aktywność młodzieży w wirtualnym świecie może wpływać na budowanie więzi społecznych i jak te więzi zmieniają się w cyfrowej rzeczywistości. Grupę badawczą stanowili przedstawiciele generacji Z, czyli osoby urodzone po 1995 r., ale nie później niż w 2010 r., które właściwie od dziecka miały możliwość korzystania z Internetu. Dla wielu z nich aktywność w mediach społecznościowych stała się naturalnym elementem egzystencji.

Opracowanie składa się z trzech rozdziałów. W pierwszym autorka poruszyła zagadnienia socjalizacji, wychowania i edukacji w dobie nowych technologii oraz scharakteryzowała generację Z. Ponadto omówiła historię i rodzaje portali społecznościowych, a także dotychczasowe badania nad korzystaniem z Internetu, w tym z serwisów społecznościowych, w kontekście relacji międzyludzkich. W drugim rozdziale przedstawiła założenia metodologiczne swojego badania, a w trzecim zaprezentowała jego rezultaty. Dotyczyły one obecności młodzieży licealnej w serwisach społecznościowych oraz więzi społecznych nastolatków w rodzinie, klasie i pozaszkolnej grupie zrzeszeniowej (stowarzyszenia, fundacje, drużyny sportowe). Wyniki badania zostały zaprezentowane z podziałem na płeć, miejsce zamieszkania i wiek respondentów oraz ich aktywność na portalach społecznościowych z uwzględnieniem intensywności prowadzenia profili i ich treści, a więzi społeczne – ze względu na typ i siłę.

Publikacja zawiera ponadto kwestionariusz ankiety wykorzystany w badaniu.

Dorota Kierska

Centralna Biblioteka Statystyczna, Polska / Central Statistical Library, Poland

Centralna Biblioteka Statystyczna im. Stefana Szulca

Biblioteka została założona w 1918 r. Gromadzi i udostępnia polskie i zagraniczne wydawnictwa statystyczne, bieżące i archiwalne (od początku XIX w.). Katalog CBS jest dostępny online na stronie <http://cbs.stat.gov.pl>. Biblioteka posiada zbiory cyfrowe: zeskanowane książki i czasopisma statystyczne z okresu międzywojennego, cymelia statystyczno-demograficzne z końca XIX i początku XX w. oraz najważniejsze publikacje GUS wydane po II wojnie światowej. Znajdują się w nich także wszystkie numery „WS”.

Zapotrzebowanie na kwerendy oraz zamówienia na odbitki kserograficzne i skany można zgłaszać pod adresem: zapytajCBS@stat.gov.pl.

The Stefan Szulc Central Statistical Library

The library was founded in 1918. It collects and provides access to Polish and foreign statistical publications, both current and archival (from the beginning of the 19th century). The catalogue of the book collection is available online at <http://cbs.stat.gov.pl>. The library offers digital resources: scanned statistical books and journals from the interwar period, rare statistical-demographic publications from the late 19th and early 20th centuries and the most important publications of Statistics Poland issued after World War II. These also include all the issues of *WS*.

Requests for queries or photocopies and scans can be submitted to: zapytajCBS@stat.gov.pl.

WYDAWNICTWA GUS. KWIECIEŃ 2024 PUBLICATIONS OF STATISTICS POLAND. APRIL 2024

W ofercie wydawniczej Głównego Urzędu Statystycznego z ubiegłego miesiąca warto zwrócić uwagę na następującą publikację:

Among Statistics Poland's publications from the previous month, we would like to recommend:

Rozwój regionalny Polski – raport analityczny 2023 **Regional development of Poland – analytical report 2023**

Czwarta edycja analitycznego opracowania przygotowywanego na potrzeby monitorowania postępów w realizacji Krajowej Strategii Rozwoju Regionalnego 2030.



Język: polski (przedmowa, spis treści, ważniejsze skróty, synteza, wykresy i mapy również w języku angielskim)

Language: Polish (preface, contents, main abbreviations, executive summary, charts and maps available also in English)

Seria: Analizy statystyczne

Series: Statistical analyses

Dostępne wersje: elektroniczna

Available in: electronic form

W raporcie omówiono zagadnienia z zakresu monitorowania zagrożeń środowiskowych, procesów demograficznych oraz uwarunkowań rozwoju kapitału ludzkiego i społecznego, a także kwestie dotyczące gospodarki, infrastruktury i potencjału administracyjnego samorządów. Prezentowane dane dotyczą roku 2022 lub ostatniego dostępnego. Dla zilustrowania zmian poszczególne informacje ujęto również w retrospekcji od 2011 r. oraz w porównaniu z poprzednim rokiem.

Publikacja składa się z trzech działów dotyczących: pierwszy – regionów NUTS 2 oraz Polski Wschodniej i Śląska na tle całego kraju, drugi – obszarów miejskich, w tym miast tracących funkcje społeczno-gospodarcze, a trzeci – obszarów wiejskich i gmin zagrożonych trwałą marginalizacją. Uzupełnieniem opracowania są dwa załączniki: zestawienie wskaźników służących do monitorowania Krajowej Strategii Rozwoju Regionalnego 2030 oraz mapy przedstawiające delimitacje przestrzenne.

Raport GUS stanowi część *Raportu z monitoringu Krajowej Strategii Rozwoju Regionalnego do 2030 roku* dostępnego na stronie Ministerstwa Funduszy i Polityki Regionalnej.

W kwietniu br. ukazały się ponadto:

- *Aktywność ekonomiczna ludności Polski – 4 kwartał 2023 r.*;
- „Biuletyn statystyczny” nr 3/2024;
- *Ceny robót budowlano-montażowych i obiektów budowlanych (luty 2024 r.)*;
- *Current challenges in survey research: Graham Kalton on probability and nonprobability sampling, Malay Ghosh on small area estimation, with comments*;
- *Działalność przedsiębiorstw posiadających jednostki zagraniczne w 2022 r.*;
- *Koniunktura w przetwórstwie przemysłowym, budownictwie, handlu i usługach 2000–2024 (kwiecień 2024)*;
- *Ludność. Stan i struktura oraz ruch naturalny w przekroju terytorialnym w 2023 r. Stan w dniu 31 grudnia*;
- *Nakłady i wyniki przemysłu w 2023 r.*;
- *Pierwszy biuletyn dotyczący Programu Statystyki Panafrkańskiej II (PAS II)*;
- *Produkcja upraw rolnych i ogrodnich w 2023 r.*;
- *Produkcja ważniejszych wyrobów przemysłowych w marcu 2024 r.*;
- *Rewitalizacja w gminach w latach 2020–2022*;
- *Równoległy oraz wyprzedzający zagregowany wskaźnik koniunktury, zegar koniunktury – szereg do stycznia 2024 r.*;
- *Sytuacja społeczno-gospodarcza kraju w 1 kwartale 2024 r.*;
- *Sytuacja społeczno-gospodarcza województw Nr 4/2023*;
- „Wiadomości Statystyczne. The Polish Statistician” nr 3/2024;
- „Wiadomości Statystyczne. The Polish Statistician” nr 4/2024;
- *Zatrudnienie i wynagrodzenia w gospodarce narodowej w 2023 r.*

Joanna Sadowy

Główny Urząd Statystyczny, Departament Opracowań Statystycznych, Polska
Statistics Poland, Statistical Products Department, Poland

Wszystkie publikacje GUS w wersji elektronicznej są dostępne na stronie stat.gov.pl/publikacje/publikacje-a-z. Wersje drukowane (jeśli zostały wydane) można zamawiać pod adresem: zws-sprzedaz@stat.gov.pl.

All the publications of Statistics Poland available in electronic form can be accessed at stat.gov.pl/en/publications. Printed versions (if available) may be ordered at: zws-sprzedaz@stat.gov.pl.

DLA AUTORÓW FOR THE AUTHORS

(for the English translation of the information given below, please visit ws.stat.gov.pl/ForAuthors)

W „Wiadomościach Statystycznych. The Polish Statistician” („WS”) zamieszczane są artykuły o charakterze naukowym poświęcone teorii i praktyce statystycznej, które prezentują wyniki oryginalnych badań teoretycznych lub analitycznych wykorzystujących metody statystyki matematycznej, opisowej bądź ekonometrii. Ukazują się również artykuły przeglądowe, recenzje publikacji naukowych oraz inne opracowania informacyjne. W czasopiśmie publikowane są prace w języku polskim i angielskim.

Od 2007 r. „WS” znajdują się na liście czasopism naukowych MEiN. Zgodnie z komunikatem Ministra Edukacji i Nauki z dnia 1 grudnia 2021 r. w sprawie wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych „WS” otrzymały 70 punktów.

„Wiadomości Statystyczne. The Polish Statistician” są udostępniane w następujących bazach, repozytoriach, katalogach i wyszukiwarkach: Agro, BazEkon, Biblioteka Nauki, Central and Eastern European Academic Source (CEEAS), Central and Eastern European Online Library (CEEOL), Central European Journal of Social Sciences and Humanities (CEJSH), Directory of Open Access Journals (DOAJ), EBSCO Discovery Service, European Reference Index for the Humanities and Social Sciences (ERIH Plus), Exlibris Primo, Google Scholar, ICI Journals Master List, ICI World of Journals, Norwegian Register for Scientific Journals and Publishers (The Nordic List) oraz Summon.

Za publikację artykułów na łamach „WS” autorzy nie otrzymują honorariów ani nie wnoszą opłat.

1. Zgłaszanie artykułów

Prace przeznaczone do opublikowania w „WS” należy przysyłać za pośrednictwem platformy Editorial System: www.editorialsystem.com/ws.

Zgłaszany artykuł powinien być zanonimizowany, tj. pozbawiony informacji o autorze/autorach (również we właściwościach pliku), podziękowań i informacji o źródłach finansowania, a także innych informacji wskazujących na afiliację lub umożliwiających zidentyfikowanie autora. Jeżeli w pracy występują tablice, wykresy lub mapy, powinny być umieszczone w treści artykułu. Materiały graficzne, razem z danymi do nich, należy ponadto załączyć jako osobny plik / osobne pliki, najlepiej w formacie Excel. **Prosimy o niestosowanie stylów i ograniczenie formatowania do wymogów redakcyjnych.** Więcej informacji w pkt 4 *Wymogi redakcyjne*.

Razem z artykułem należy przesłać skan/zdjęcie oświadczenia o oryginalności pracy i niemożności jej w innym wydawnictwie. **Załączenie oświadczenia jest warunkiem poddania pracy ocenie wstępnej i skierowania do recenzji.**

Zgłoszenie artykułu do opublikowania w „WS” oznacza zgodę na jego udostępnienie na licencji Creative Commons Uznanie autorstwa – Na tych samych warunkach 4.0 (CC BY-SA 4.0).

Autorzy mają prawo do samodzielnego umieszczania w wybranych przez siebie repozytoriach artykułu w wersji zarówno zgłoszonej do „WS”, jak i zaakceptowanej do opublikowania

oraz opublikowanej, z zastrzeżeniem wymogu niezwłocznego podania w repozytorium informacji o numerze „WS”, w którym praca się ukazała, wraz z linkiem do niej (DOI).

2. Przebieg prac redakcyjnych

Zgłoszony artykuł jest oceniany i opracowywany w czteroetapowym procesie:

1. **Ocena wstępna**, dokonywana przez redakcję. Polega na weryfikacji naukowego charakteru artykułu oraz jego struktury i zawartości pod kątem wymogów redakcyjnych, a także zgodności tematyki z profilem czasopisma. Autor uzupełnia i poprawia artykuł stosownie do uwag redakcji, a w przypadku nieuwzględnienia danej uwagi uzasadnia swoje stanowisko. Warunkiem skierowania pracy do recenzji jest potwierdzenie oryginalności tekstu uzyskane za pomocą systemu antyplagiatowego. W przypadku wykrycia znacznego podobieństwa do innych prac artykuł zostanie odrzucony.
2. **Ocena recenzentów**, dokonywana przez specjalistów w danej dziedzinie. Artykuł oceniają dwaj recenzenci spoza jednostki naukowej, przy której afiliowany jest autor; w przypadku pracy w języku angielskim co najmniej jeden recenzent jest afiliowany przy jednostce zagranicznej. W razie sprzecznych opinii dwóch recenzentów powoływany jest trzeci recenzent. Recenzenci kierują się kryteriami oryginalności i jakości opracowania zarówno w odniesieniu do treści, jak i formy artykułu.

Autorzy artykułów, które otrzymały pozytywne oceny, wprowadzają poprawki zalecane przez recenzentów i przesyłają zmodyfikowaną wersję pracy. Jeśli pojawi się różnica zdań dotycząca zasadności proponowanych zmian, autorzy są zobligowani do uzasadnienia swojego stanowiska.

3. **Ocena Kolegium Redakcyjnego (KR)**, decydująca o przyjęciu pracy do publikacji. Jest dokonywana na podstawie recenzji, z uwzględnieniem opinii redaktorów tematycznego i merytorycznego. Polega m.in. na weryfikacji dokonania przez autora zmian w artykule stosownie do uwag recenzentów. KR ocenia artykuł pod względem poprawności i spójności merytorycznej oraz zaleca autorowi wprowadzenie poprawek, jeśli są one konieczne, aby praca spełniała wymogi czasopisma. Autorowi przysługuje prawo do odwołania od decyzji o niepublikowaniu artykułu. W takim przypadku powinien on skontaktować się z redakcją „WS” i przedstawić uzasadnienie. Ostateczna decyzja w tej sprawie należy do redaktora naczelnego.

W „WS” publikowane są wyłącznie te artykuły, które otrzymają pozytywną ocenę na każdym z wymienionych etapów i zostaną poprawione przez autora zgodnie z otrzymanymi uwagami (chyba że autor przedstawi argumenty uzasadniające nieuwzględnienie danej uwagi).

Artykuły przyjęte przez KR do publikacji są zamieszczane na stronie internetowej czasopisma w zakładce Early View, gdzie znajdują się do czasu opublikowania w konkretnym wydaniu „WS”.

4. **Opracowanie redakcyjne, autoryzacja i korekta**. Artykuł zakwalifikowany do druku jest poddawany opracowaniu merytorycznemu i językowemu. Redakcja zastrzega sobie prawo do zmiany tytułu i śródtytułów, modyfikowania tablic, wykresów i innych elementów graficznych oraz przededagowania treści bez naruszenia zasadniczej myśli autora.

Po opracowaniu redakcyjnym artykuł jest przesyłany do autoryzacji. Tekst zatwierdzony przez autora, po składzie i łamaniu, jest poddawany korekcie i rewizji (II korekcje).

Autor dokonuje korekty autorskiej tekstu na etapie rewizji. Wykresy i inne materiały graficzne są opracowywane na podstawie plików i danych przekazanych przez autora i poddawane korekcie i rewizji. Autor dokonuje ich akceptacji na etapie rewizji.

W przypadku odkrycia błędów w opublikowanym artykule zamieszcza się na łamach „WS” sprostowanie, a artykuł w wersji elektronicznej jest poprawiany i umieszczany na stronie internetowej „WS” ze stosownym wyjaśnieniem.

3. Zasady etyki publikacyjnej COPE

Redakcja „WS” podejmuje wszelkie starania w celu utrzymania najwyższych standardów etycznych zgodnie z wytycznymi Komitetu ds. Etyki Publikacyjnej (COPE), dostępnymi na stronie internetowej www.publicationethics.org, oraz wykorzystuje wszystkie możliwe środki mające na celu zapobieżenie nadużyciom i nierzetelności autorskiej. Przyjęte zasady postępowania obowiązują autorów, Radę Naukową, Kolegium Redakcyjne, redakcję, pracowników Wydziału Czasopism Naukowych GUS, recenzentów i wydawcę.

3.1. Odpowiedzialność autorów

1. Artykuły naukowe kierowane do opublikowania w „WS” powinny zawierać precyzyjny opis badanych zjawisk i stosowanych metod oraz autorskie wnioski i sugestie dotyczące rozwoju badań i analiz statystycznych. Autorzy powinni wyraźnie określić cel artykułu oraz jasno przedstawić wyniki przeprowadzonej analizy. Prezentacja efektów badań statystycznych zaprojektowanych i przeprowadzonych przez autorów wymaga opisanie zastosowanej w nich metodologii. W przypadku nowatorskich metod analizy pożądane jest podanie przykładu ilustrującego ich zastosowanie w praktyce statystycznej. Autorzy ponoszą odpowiedzialność za treści prezentowane w artykułach. W razie zgłaszania przez czytelników zastrzeżeń odnoszących się do tych treści autorzy są zobligowani do udzielenia odpowiedzi za pośrednictwem redakcji.
2. Na autorach spoczywa obowiązek zapewnienia pełnej oryginalności przedłożonych prac. Redakcja nie toleruje przejawów nierzetelności naukowej autorów, takich jak:
 - duplikowanie publikacji – ponowne publikowanie własnego utworu lub jego części;
 - plagiat – przywłaszczenie cudzego utworu lub jego fragmentu bez podania informacji o źródle;
 - fabrykowanie danych – oparcie pracy naukowej na nieprawdziwych wynikach badań;
 - autorstwo widmo (*ghost authorship*) – nieujawnianie współautorów, mimo że wnieśli oni istotny wkład w powstanie artykułu;
 - autorstwo gościnne (*guest authorship*) – podawanie jako współautorów osób o znikomym udziale lub niebiorących udziału w opracowywaniu artykułu;
 - autorstwo grzecznościowe (*gift authorship*) – podawanie jako współautorów osób, których wkład jest oparty jedynie na słabym powiązaniu z badaniem.

Autorzy deklarują w stosownym oświadczeniu, że zgłaszany artykuł nie narusza praw autorskich osób trzecich, nie był dotychczas publikowany i nie został złożony w innym wydawnictwie oraz że jest ich oryginalnym dziełem, i określają swój wkład w opracowanie artykułu. Jeżeli doszło do zaprezentowania podobnych materiałów podczas konferencji lub

symposium naukowe, to podczas składania tekstu do publikacji w „WS” autorzy są zobowiązani poinformować o tym redakcję.

3. Autorzy są zobowiązani do podania w treści artykułu wszelkich źródeł finansowania badań będących podstawą pracy.
4. Główną odpowiedzialność za rzetelność przekazanych informacji, łącznie z informacją na temat wkładu poszczególnych współautorów w powstanie artykułu, ponosi zgłaszający artykuł.
5. Autorzy zgłaszający artykuły do publikacji w „WS” biorą udział w procesie recenzji double-blind peer review, dokonywanej przez co najmniej dwóch niezależnych ekspertów z danej dziedziny. Po otrzymaniu pozytywnych recenzji autorzy wprowadzają zalecane przez recenzentów poprawki i dostarczają redakcji zaktualizowaną wersję opracowania wraz z pisemnym poświadczeniem uwzględnienia poprawek. Jeśli pojawi się różnica zdań co do zasadności proponowanych zmian, należy wyjaśnić, które poprawki zostały uwzględnione, a w przypadku ich nieuwzględnienia – uzasadnić swoje stanowisko.
6. Jeżeli autorzy odkryją w swoim maszynopisie lub tekście już opublikowanym błędy, nieścisłości bądź niewłaściwe dane, powinni niezwłocznie poinformować o tym redakcję w celu dokonania korekty, wycofania tekstu lub zamieszczenia sprostowania. W przypadku korekty artykułu już opublikowanego jego nowa wersja jest zamieszczana na stronie internetowej „WS” wraz ze stosownym wyjaśnieniem.

3.2. Odpowiedzialność Rady Naukowej, Kolegium Redakcyjnego i Wydziału Czasopism Naukowych GUS

1. Rada Naukowa (RN) kształtuje profil programowy czasopisma, określa kierunki jego rozwoju i konsultuje jego zakres merytoryczny.
2. Kolegium Redakcyjne (KR) podejmuje decyzję o publikacji danego artykułu z uwzględnieniem ocen recenzentów oraz opinii zespołu redakcyjnego. W swoich rozstrzygnięciach członkowie KR kierują się kryteriami merytorycznej oceny wartości artykułu, jego oryginalności i jasności przekazu, a także ścisłego związku z celem i zakresem tematycznym „WS”. Oceniają artykuły niezależnie od płci, rasy, pochodzenia etnicznego, narodowości, religii, wyznania, światopoglądu, niepełnosprawności, wieku lub orientacji seksualnej ich autorów.
3. Zespół redakcyjny, wyodrębniony z KR, tworzą redaktor naczelny i jego zastępca, redaktorzy tematyczni i redaktor merytoryczny. Członkowie zespołu redakcyjnego weryfikują nadsyłane artykuły pod względem merytorycznym, oceniają ich zgodność z celem i zakresem tematycznym „WS” oraz sprawdzają spełnienie wymogów redakcyjnych i przestrzeganie zasad rzetelności naukowej. Ponadto wybierają recenzentów w taki sposób, aby nie wystąpił konflikt interesów, i dbają o zapewnienie uczciwego, bezstronnego i terminowego procesu recenzowania.
4. Za sprawny przebieg procesu wydawniczego, poinformowanie wszystkich jego uczestników o konieczności przestrzegania obowiązujących zasad i przygotowanie artykułów do publikacji odpowiadają pracownicy Wydziału Czasopism Naukowych (WCN) GUS. W celu uzyskania obiektywnej oceny oryginalności nadsyłanych artykułów przed skierowaniem ich do recenzji WCN wykorzystuje system antyplagiatowy. Informacje dotyczące

artykułu mogą być przekazywane przez WCN wyłącznie autorom, recenzentom, członkom RN i KR oraz wydawcy.

5. Zmiany dokonane w tekście na etapie przygotowania artykułu do publikacji nie mogą naruszać zasadniczej myśli autorów. Wszelkie modyfikacje o charakterze merytorycznym są z nimi konsultowane.
6. W przypadku podjęcia decyzji o niepublikowaniu artykułu nie może on zostać w żaden sposób wykorzystany przez wydawcę lub uczestników procesu wydawniczego bez pisemnej zgody autorów. Autorzy mogą się odwołać od decyzji o niepublikowaniu artykułu. W tym celu powinni się skontaktować z redaktorem naczelnym lub sekretarzem redakcji „WS” i przedstawić stosowną argumentację. Odwołania autorów są rozpatrywane przez redaktora naczelnego.
7. Członkowie RN i KR ani pracownicy WCN nie mogą pozostawać w jakimkolwiek konflikcie interesów w odniesieniu do artykułów zgłaszanych do publikacji. Przez konflikt interesów należy rozumieć sytuację, w której jakiegokolwiek interesy lub zależności (służbowe, finansowe lub inne) mogą mieć wpływ na ocenę artykułu lub decyzję o jego publikacji.
8. W celu przeciwdziałania nierzetelności naukowej wymagane jest złożenie przez autorów oświadczenia, w którym deklarują, że zgłaszany artykuł nie narusza praw autorskich osób trzecich, nie był dotychczas publikowany i jest ich oryginalnym dziełem, a także określają swój wkład w opracowanie artykułu.
9. W celu zapewnienia wysokiej jakości recenzji wymagane jest złożenie przez recenzentów oświadczenia o przestrzeganiu zasad etyki recenzowania COPE i niewystępowaniu konfliktu interesów.
10. W przypadku uzasadnionego podejrzenia na jakimkolwiek etapie procesu wydawniczego, że autorzy dopuścili się nierzetelności naukowej (zob. pkt 3.1. Odpowiedzialność autorów), zespół redakcyjny skrupulatnie zbada sprawę ewentualnego nadużycia. Jeśli nierzetelność autorów zostanie udowodniona, to zgłoszony przez nich artykuł zostanie odrzucony przez KR, a autorzy otrzymają informację o podjętej decyzji wraz z jej uzasadnieniem.
11. Czytelnicy, którzy mają wobec autorów opublikowanego artykułu uzasadnione podejrzenia o nierzetelność naukową, powinni powiadomić o tym redaktora naczelnego lub sekretarza redakcji. Po zbadaniu sprawy ewentualnego nadużycia czytelnicy zostaną poinformowani o rezultacie przeprowadzonego postępowania. W przypadku potwierdzenia nadużycia, na łamach czasopisma zostanie zamieszczona stosowna informacja.

3.3. Odpowiedzialność recenzentów

1. Recenzenci przyjmują artykuł do recenzji tylko wtedy, gdy uznają, że:
 - posiadają odpowiednią wiedzę w określonej dziedzinie, aby rzetelnie ocenić pracę;
 - zgodnie z ich stanem wiedzy nie istnieje konflikt interesów w odniesieniu do autorów, przedstawionych w artykule badań i instytucji je finansujących, co potwierdzają w oświadczeniu;
 - mogą wywiązać się z terminu ustalonego przez redakcję, aby nie opóźnić publikacji.
2. Recenzenci są zobligowani do zachowania obiektywności i poufności oraz powstrzymania się od osobistej krytyki. Zawsze powinni uzasadnić swoją ocenę, przedstawiając stosowną argumentację.

3. Recenzenci powinni wskazać ważne dla wyników badań opublikowane prace, które w ich ocenie powinny zostać przywołane w ocenianym artykule.
4. W razie stwierdzenia wysokiego poziomu zbieżności treści recenzowanej pracy z innymi opublikowanymi materiałami lub podejrzenia innych przejawów nierzetelności naukowej recenzenci są zobowiązani poinformować o tym redakcję.
5. Po ukończeniu recenzji przechowywanie przesłanych przez redakcję materiałów (w jakiegokolwiek formie) oraz posługiwanie się nimi przez recenzentów jest niedozwolone.

3.4. Odpowiedzialność wydawcy

1. Materiały opublikowane w „WS” są chronione prawem autorskim.
2. Wydawca udostępnia pełną treść wszystkich artykułów w internecie w trybie otwartego dostępu, tj. bezpłatnie i bez technicznych ograniczeń, od 1 stycznia 2022 r. na licencji Creative Commons Uznanie autorstwa – Na tych samych warunkach 4.0 (CC BY-SA 4.0). W przypadku artykułów zgłoszonych do „WS” od 2022 r. dozwolone jest dzielenie się artykułem (kopiowanie i rozpowszechnianie go w dowolnym medium i formie) oraz adaptowanie go (w dowolnym celu, także komercyjnym) na warunkach określonych w tej licencji. Z pozostałych artykułów zamieszczonych w czasopiśmie można korzystać w ramach otwartego dostępu, zgodnie z ustawą o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego.
3. Wydawca deklaruje gotowość do opublikowania poprawek, wyjaśnień oraz przeprosin.

4. Wymogi redakcyjne

Zgodnie z wymogami czasopisma omawiany w artykule problem badawczy powinien być jednoznacznie zdefiniowany oraz istotny dla oceny zjawisk społecznych lub gospodarczych. Artykuł powinien zawierać wyraźnie określony cel badania, precyzyjny opis badanych zjawisk i stosowanych metod, uzyskane wyniki przeprowadzonej analizy oraz autorskie wnioski.

4.1. Struktura i zawartość artykułu

Wymagane elementy artykułu:

1. Tytuł.
2. Dane autora: imię/imiona i nazwisko, afiliacja w języku polskim i angielskim, ORCID, wkład w powstanie artykułu, adres e-mail. Wśród autorów artykułu wieloautorskiego należy wskazać autora korespondencyjnego.
3. Streszczenie (zalecana objętość – do 1200 znaków ze spacjami, forma bezosobowa). W przypadku artykułu opisującego badanie empiryczne powinno zawierać: cel, przedmiot, okres i metodę badania, źródła danych i najważniejsze wnioski z badania. W przypadku artykułów o innym charakterze należy podać co najmniej cel pracy, jej przedmiot i najważniejsze wnioski.

Streszczenie to podstawowe źródło informacji o artykule, warunkujące też decyzję czytelnika o zapoznaniu się z całą pracą. Dlatego powinno być przygotowane ze szczególną starannością i dbałością o umieszczenie w nim wszystkich wymaganych elementów.

4. Słowa kluczowe – najistotniejsze pojęcia lub wyrażenia użyte w pracy (nie mniej niż trzy). Powinny być zawarte w streszczeniu i/lub tytule.
 5. Kod/kody z klasyfikacji Journal of Economic Literature (JEL).
 6. Tłumaczenie tytułu, streszczenia i słów kluczowych (na język angielski w przypadku artykułu napisanego w języku polskim, a na język polski w przypadku artykułu napisanego w języku angielskim).
 7. W artykule opisującym badanie empiryczne wymagane są następujące części:
 - wprowadzenie, zawierające syntetyczne przedstawienie zagadnień teoretycznych, uzasadnienie podjęcia danego problemu badawczego, cel badania i krytyczne odniesienie do literatury przedmiotu. W wyjątkowych przypadkach, kiedy istotne dla podjętego tematu jest obszerniejsze przedstawienie dyskusji toczącej się w literaturze, przegląd literatury może stanowić odrębną część artykułu;
 - metoda badania, uwzględniająca przedmiot i okres badania, źródła danych i zastosowane metody badawcze, w tym uzasadnienie ich wyboru;
 - wyniki badania – analiza danych oraz interpretacja wyników i odniesienie ich do rezultatów wcześniejszych badań (dyskusja). W uzasadnionych przypadkach dyskusja może stanowić odrębną część artykułu;
 - podsumowanie, które powinno być zwarte i odzwierciedlać istotę problemu badawczego przedstawionego w artykule, bez podawania danych liczbowych; końcowe wnioski powinny odnosić się do treści artykułu, a w szczególności do celu badania.
- Wszystkie części powinny być opatrzone numerami.
8. Bibliografia, zawierająca pełny wykaz prac i materiałów przywołanych w artykule, przygotowana zgodnie z wymogami czasopisma.

4.2. Przygotowanie artykułu

1. Artykuł powinien być utrzymany w formie bezosobowej.
2. Tekst należy zapisać alfabetem łańciskim. Nazwy własne, tytuły itp. oryginalnie zapisane innym alfabetem powinny być poddane transliteracji.
3. Nie należy stosować stylów; formatowanie należy ograniczyć do wymogów redakcyjnych.
4. Objętość artykułu łącznie ze streszczeniem, słowami kluczowymi, bibliografią, tablicami, wykresami i innymi materiałami graficznymi nie powinna być mniejsza niż 10 stron maszynopisu ani przekraczać 20 stron.
5. Edytor tekstu: Microsoft Word, format *.doc lub *.docx.
6. Krój czcionki:
 - Arial – tytuł, autor, streszczenie, słowa kluczowe, kody JEL, śródtytuły, elementy graficzne (tablice, zestawienia, wykresy, schematy), przypisy;
 - Times New Roman – tekst główny, bibliografia.
7. Wielkość czcionki:
 - 14 pkt – tytuł, autor, śródtytuły wyższego rzędu;
 - 12 pkt – tekst główny, śródtytuły niższego rzędu;
 - 10 pkt – pozostałe elementy.
8. Marginesy – 2,5 cm z każdej strony.

9. Interlinia – 1,5 wiersza; tablice i przypisy – 1 wiersz; przed tytułami rozdziałów i podrozdziałów oraz po nich – pusty wiersz.
10. Wcięcie akapitowe – 0,4 cm; bibliografia – bez wcięcia, wysunięcie 0,4 cm.
11. Przy wycienieniach należy posłużyć się listą punktowaną z punktarami w postaci kropek (wysunięcie 0,4 cm, wcięcie 0 cm); wiersze (oprócz ostatniego) zakończone średnikiem.
12. Strony ponumerowane automatycznie.
13. Tablice i elementy graficzne (wykresy, mapy, schematy) muszą być przywołane w tekście.
14. Wykresy, mapy i schematy należy zamieścić w tekście głównym. Wykresy powinny być edytowalne (optymalnie wykonane w programie Excel; w przypadku wykonania w programie graficznym powinny mieć postać wektorową). Wykresy i inne materiały graficzne należy przekazać osobno, najlepiej w pliku programu Excel lub innym edytowalnym w pakiecie Microsoft Office.
15. Tablice muszą być edytowalne. Nie należy stosować rastrów, cieniowania, pogrubiania czy też podwójnych linii itp.
16. Wskazówki dotyczące opracowywania map znajdują się w publikacji *Mapy statystyczne. Opracowanie i prezentacja danych*, dostępnej na stronie internetowej GUS.
17. Pod tablicami i każdym elementem graficznym należy podać źródło opracowania, a także objaśnić użyte w nich skróty i symbole.
18. Literowe symbole liczb i innych wielkości niezłożonych należy zapisywać małą lub dużą literą i pismem pochyłym (np. a , A , $y(x)$, a_i); wektorów – pismem pochyłym i pogrubionym (np. \mathbf{a} , \mathbf{A} , \mathbf{w} , $\mathbf{y}(x)$, \mathbf{w}_i); macierzy – pismem prostym i pogrubionym (np. \mathbf{A} , \mathbf{a} , \mathbf{M} , $\mathbf{Y}(x)$, \mathbf{M}_i).
19. Objasnienia znaków umownych i zapisów w tablicach: kreska (–) – zjawisko nie wystąpiło; zero (0) – zjawisko istniało w wielkości mniejszej od 0,5; (0,0) – zjawisko istniało w wielkości mniejszej od 0,05; kropka (.) – brak informacji, konieczność zachowania tajemnicy statystycznej, wypełnienie pozycji jest niemożliwe lub niecelowe; „w tym” – oznacza, że nie podaje się wszystkich składników sumy.
20. Stosowane są następujące skróty: tablica – tabl., wykres – wykr.
21. Wszystkie zawarte w artykule informacje, dane i stwierdzenia wykraczające poza wiedzę powszechną – np. wyniki badań innych autorów, zarówno o charakterze empirycznym, jak i koncepcyjnym – muszą być opatrzone przypisem bibliograficznym. Przez wiedzę powszechną należy rozumieć informacje ogólnie znane i niebudzące wątpliwości ani kontrowersji w danej grupie społecznej, np. utworzenie GUS w 1918 r. lub powstanie UE w 1993 r. na podstawie traktatu z Maastricht. Natomiast dane statystyczne udostępniane lub publikowane np. przez GUS lub Eurostat nie należą do takich informacji. Charakteru wiedzy powszechnej nie mają również stwierdzenia odnoszące się do idei, zjawisk i procesów społecznych, politycznych czy gospodarczych. Nawet pozornie zdroworozsądkowe idee zmieniają bowiem swój sens w zależności od kultury, języka lub dyscypliny naukowej, a także bywają w rozmaity sposób konceptualizowane, jak np. pojęcie poznania w naukach społecznych.

Podanie źródła jest konieczne niezależnie od tego, czy informacje lub stwierdzenia są ujęte w ramy cytatu, czy przedstawione bez dosłownego przytoczenia, np. w formie parafrazy. Jeżeli stwierdzenie może budzić jakiegokolwiek wątpliwości odbiorców, autor powinien wskazać stosowne źródło podawanej informacji.

22. Przypisy rzeczowe, słownikowe lub informacyjne należy umieszczać na dole strony. Przypisy bibliograficzne, zgodnie ze standardem APA (American Psychological Association), należy podawać w tekście głównym.
23. Bibliografię należy przygotować zgodnie ze standardem APA.

4.3. Zasady przywoływania publikacji w treści artykułu

Wyszczególnienie	Przykład przywołania	
	w odsyłaczu	w treści zdania
Autor indywidualny		
Jeden autor	(Iksiński, 2001)	Iksiński (2001)
Dwóch autorów	(Iksiński i Nowak, 1999)	Iksiński i Nowak (1999)
Trzech autorów lub więcej	(Jankiewicz i in., 2003)	Jankiewicz i in. (2003)
Autor instytucjonalny		
Nazwa funkcjonuje jako powszechnie znany skrótowiec: pierwsze przywołanie w tekście	(International Labour Organization [ILO], 2020)	International Labour Organization (ILO, 2020)
kolejne przywołanie	(ILO, 2020)	ILO (2020)
Pełna nazwa	(Stanford University, 1995)	Stanford University (1995)
Typ publikacji		
Publikacja bez ustalonego autorstwa	(<i>Skrócony tytuł ...</i> , 2015)	<i>Pełny tytuł</i> (2015)
Publikacja bez roku wydania	(Iksiński, b.r.)	Iksiński (b.r.)
Akt prawny	(Pełny tytuł)	Pełny tytuł
Strona internetowa / Zbiór danych: znana data publikacji	(Iksiński, 2020) / (Nazwa instytucji, 2020)	Iksiński (2020) / Nazwa instytucji (2020)
nieznana data publikacji	(Iksiński, b.r.) / (Nazwa instytucji, b.r.)	Iksiński (b.r.) / Nazwa instytucji (b.r.)
Rodzaj przywołania		
Przywoływanie kilku prac (porządek prac – chronologiczny, porządek autorów – alfabetyczny)	(Iksiński, 1997, 1999, 2004a, 2004b; Nowak, 2002)	Iksiński (1997, 1999, 2004a, 2004b) i Nowak (2002)
Przywoływanie publikacji za innym autorem (uwaga: w bibliografii należy wymienić tylko pracę czytaną)	(Nowakowski, 1990, za: Zienniecka, 2007)	Nowakowski (1990, za: Zienniecka, 2007)

Źródło: opracowanie na podstawie: American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th edition). <https://doi.org/10.1037/0000165-000>.

4.4. Przykłady opisu bibliograficznego

Bibliografia powinna być zamieszczona na końcu opracowania. Prace należy uszeregować alfabetycznie według nazwiska pierwszego autora. W przypadku dwóch lub więcej prac tego samego autora / tych samych autorów trzeba je uporządkować chronologicznie według roku publikacji. Jeśli kilka prac tego samego autora / tych samych autorów zostało opublikowanych w tym samym roku, należy podać je w kolejności alfabetycznej według tytułu i odpowiednio oznaczyć literami a, b, c itd.

Typ publikacji	Przykład opisu bibliograficznego
Artykuł w czasopiśmie	
W wersji drukowanej	Nazwisko, X. (rok). Tytuł artykułu. <i>Tytuł czasopisma, rocznik (zeszyt)</i> , strona początku–strona końca.
Dostępny w internecie, z DOI	Nazwisko, X., Nazwisko 2, Y. (rok). Tytuł artykułu. <i>Tytuł czasopisma, rocznik(zeszyt)</i> , strona początku–strona końca. https://doi.org/xxx .
Dostępny w internecie, bez DOI	Nazwisko, X., Nazwisko 2, Y., Nazwisko 3, Z. (rok). Tytuł artykułu. <i>Tytuł czasopisma, rocznik(zeszyt)</i> , strona początku–strona końca. https://xxx .
Maszynopis	
Niepublikowany / przygotowywany przez autora / zgłoszony do publikacji, ale jeszcze niezaakceptowany	Nazwisko, X. (rok). <i>Tytuł [maszynopis niepublikowany / w przygotowaniu / zgłoszony do publikacji]</i> .
Zaakceptowany do publikacji	Nazwisko, X. (w druku). Tytuł artykułu. <i>Tytuł czasopisma</i> .
Opublikowany nieformalnie (np. na stronie internetowej autora)	Nazwisko, X., Nazwisko 2, Y. (rok). <i>Tytuł artykułu</i> . https://xxx .
Opublikowany w trybie online first (przed włączeniem do zeszytu)	Nazwisko, X. (rok). Tytuł artykułu. <i>Tytuł czasopisma</i> . Online first. https://xxx .
Książka	
W wersji drukowanej	Nazwisko, X. (rok). <i>Tytuł książki</i> . Wydawnictwo.
Dostępna w internecie, z DOI	Nazwisko, X. (rok). <i>Tytuł książki</i> . Wydawnictwo. https://doi.org/xxx .
Dostępna w internecie, bez DOI	Nazwisko, X. (rok). <i>Tytuł książki</i> . Wydawnictwo. https://xxx .
W przekładzie	Nazwisko, X. (rok). <i>Tytuł książki</i> (tłum. Y. Nazwisko). Wydawnictwo.
Wydanie wielotomowe: tom zatytułowany	Nazwisko, X. (rok). <i>Tytuł książki: nr tomu. Tytuł tomu</i> . Wydawnictwo.
tom niezatytułowany	Nazwisko, X. (rok). <i>Tytuł książki (nr tomu)</i> . Wydawnictwo.
Kolejne wydanie	Nazwisko, X. (rok). <i>Tytuł książki (nr wydania)</i> . Wydawnictwo.
Pod redakcją: w języku polskim	Nazwisko, X. (red.). (rok). <i>Tytuł książki</i> . Wydawnictwo.
w języku angielskim	Nazwisko, X. (Ed.). (rok). <i>Tytuł książki</i> . Wydawnictwo.
Rozdział w pracy zbiorowej	Nazwisko, X. (rok). Tytuł rozdziału. W: Y. Nazwisko, Z. Nazwisko 2 (red.), <i>Tytuł książki (s. strona początku–strona końca)</i> . Wydawnictwo. https://doi.org/xxx lub https://xxx .
Inne prace	
Raport: autor indywidualny	Nazwisko, X. (rok). <i>Tytuł raportu</i> . Wydawnictwo.
autor instytucjonalny	Nazwa instytucji. (rok). <i>Tytuł raportu</i> . Wydawnictwo.
Working Papers	Nazwisko, X. (rok). <i>Tytuł pracy</i> (nazwa serii i numer). https://doi.org/xxx lub https://xxx .
Sesja konferencyjna / prezentacja / referat	Nazwisko, X. (rok, dzień i miesiąc). <i>Tytuł pracy</i> [typ wystąpienia, np. referat]. Nazwa konferencji, miejsce konferencji.
Rozprawa doktorska: nieopublikowana	Nazwisko, X. (rok). <i>Tytuł pracy</i> [nieopublikowana rozprawa doktorska]. Nazwa instytucji nadającej tytuł doktorski.
opublikowana	Nazwisko, X. (rok). <i>Tytuł pracy</i> [rozprawa doktorska, nazwa instytucji nadającej tytuł doktorski]. https://xxx .
Akt prawny	Pełny tytuł aktu prawnego wraz z datą publikacji w dzienniku urzędowym.

Typ publikacji	Przykład opisu bibliograficznego
Strona internetowa	
Znana data publikacji, zawartość strony się nie zmienia	Nazwisko, X. (rok, dzień i miesiąc). <i>Tytuł</i> . https://xxx .
Nieznana data publikacji, zawartość strony się zmienia	Nazwa instytucji. (b.r.). <i>Tytuł</i> . Pobrane dzień, miesiąc i rok pobrania z https://xxx .
Zbiór danych	
Surowe dane nieopublikowane	Nazwisko, X. (rok wydania pracy, w której dane są wykorzystywane) [opis danych, np. surowe dane nieopublikowane dotyczące...]. Źródło danych (np. nazwa uniwersytetu).
Dane opublikowane: znana data publikacji, zawartość zbioru się nie zmienia	Nazwisko, X. (rok). <i>Nazwa zbioru danych</i> [zbiór danych]. Wydawca. https://xxx .
nieznana data publikacji, zawartość zbioru się zmienia	Nazwa instytucji. (b.r.). <i>Nazwa zbioru danych</i> [zbiór danych]. Wydawca. Pobrane dzień, miesiąc i rok pobrania z https://xxx .

Źródło: opracowanie na podstawie: American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th edition). <https://doi.org/10.1037/0000165-000>.

Praca przygotowana w sposób niezgodny z powyższymi wskazówkami będzie odesłana do autora z prośbą o dostosowanie formy artykułu do wymogów redakcyjnych.

DZIAŁY „WS” – TEMATYKA ARTYKUŁÓW WS SECTIONS – TOPICS OF THE ARTICLES

Pełny opis zakresu tematycznego działów: ws.stat.gov.pl/AimScope

Description of the topics covered in each section: ws.stat.gov.pl/AimScope

Studia metodologiczne / Methodological studies

- Oryginalne teoretyczne rozwiązania metodologiczne ze wskazaniem ich praktycznej użyteczności
- Prace przeglądowe i porównawcze oraz dotyczące etyki w statystyce, które wnoszą pionierski wkład poznawczy do obecnego stanu wiedzy

Statystyka w praktyce / Statistics in practice

- Nowatorskie zastosowania narzędzi i modeli statystycznych oraz analiza i ocena statystyczna zjawisk społeczno-ekonomicznych i innych, prowadzona w szczególności na danych z zasobów statystyki publicznej
- Wykorzystanie narzędzi informatycznych do uzyskiwania i przetwarzania informacji statystycznych, naliczania danych wynikowych, ich prezentacji i rozpowszechniania
- Projektowanie badań statystycznych, uzyskiwanie, integracja i przetwarzanie danych oraz generowanie wynikowych informacji statystycznych i kontrola ich ujawniania

Studia interdyscyplinarne. Wyzwania badawcze / Interdisciplinary studies. Research challenges

- Wyzwania badawcze wynikające z rosnących potrzeb użytkowników danych statystycznych i wymagające zaangażowania znacznych środków oraz rozwiązań z różnych dziedzin nauki i techniki
- Wykorzystanie technologii informacyjnych i komunikacyjnych, innowacyjność, przetwarzanie i analiza zagadnień związanych z data science i big data
- Wyniki badań prowadzonych przez przedstawicieli dyscyplin innych niż statystyka z wykorzystaniem metod statystycznych

Spisy powszechne – problemy i wyzwania / Issues and challenges in census taking

- Propozycje rozwiązań – zarówno organizacyjnych, jak i metodologicznych – możliwych do zastosowania w spisach oraz rezultaty analiz danych spisowych
- Praktyczne aspekty związane z gromadzeniem i udostępnianiem danych ze spisów, w tym dotyczące obciążenia odpowiedzi i ochrony tajemnicy statystycznej

Edukacja statystyczna / Statistical education

- Metody i efekty nauczania statystyki oraz popularyzacja myślenia statystycznego i rzetelnego posługiwania się informacjami statystycznymi
- Problemy związane z kształceniem w zakresie umiejętności stosowania statystyki na wszystkich poziomach edukacji, a także dotyczące wykorzystywania nowoczesnych koncepcji i metod dydaktycznych oraz pomocy naukowych w nauczaniu statystyki

Z dziejów statystyki / From the history of statistics

- Historia prowadzenia obserwacji statystycznych oraz rozwoju ich metodologii i narzędzi
- Życie i osiągnięcia zawodowe wybitnych statystyków, jak również działalność najważniejszych instytucji i organizacji statystycznych w Polsce i za granicą

In memoriam

- Nekrologi i artykuły wspomnieniowe

Informacje. Recenzje. Dyskusje / Discussions. Reviews. Information

- Teksty nierecenzowane i niemające charakteru artykułów naukowych: sprawozdania z konferencji naukowych i innych wydarzeń dotyczących statystyki, recenzje książek, omówienia nowości wydawniczych GUS, polemiki i dyskusje