Cena 15,00 zł
(VAT 8%)

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

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

LIPIEC / JULY
ROCZNIK / VOLUME 69

2024 | 7

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

POLSKIE TOWARZYSTWO STATYSTYCZNE
POLISH STATISTICAL ASSOCIATION

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

LIPIEC / JULY
ROCZNIK / VOLUME 69

2024 | 7 (758)

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 (Szkoła 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 Rady Naukowej / Scientific Council 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 Cierpiał-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. UK (Uniwersytet Kaliski 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)

sekretarz Kolegium Redakcyjnego / Editorial Board Secretary: Małgorzata Zygmont, Główny Urząd Statystyczny, Polska

Redakcja / Editorial Team

redaktor naczelny / Editor-in-Chief: Marek Cierpiał-Wolan
zastępca redaktora naczelnego / Deputy Editor-in-Chief: Andrzej Młodak
redaktorzy tematyczni / Thematic Editors: Andrzej Młodak, Małgorzata Tarczyńska-Łuniewska, Wioletta Wrzaszcz, Agnieszka Zgierska
sekretarz redakcji / Editorial Secretary: Małgorzata Zygmont

ADRES REDAKCJI I KONTAKT / EDITORIAL OFFICE ADDRESS AND CONTACT

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

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

Redakcja techniczna, skład i łamanie, opracowanie materiałów graficznych i korekta:
Zakład Wydawnictw Statystycznych – zespół pod kierunkiem Macieja Adamowicza
Technical editing, typesetting, preparation of graphic materials and proofreading:
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
e-mail: zws-sprzedaz@stat.gov.pl, tel./phone +48 22 608 32 10, +48 22 608 38 10

Zam. 241/2024 – nakład 275 egz.

SPIS TREŚCI
CONTENTS

Od redakcji IV
From the editorial team

Statystyka w praktyce
Statistics in practice

Barbara Dańska-Borsiak
Ocena adekwatności PKB per capita jako miary poziomu życia w powiatach 1
Assessment of the adequacy of GDP *per capita* in measuring the standard of living in Polish powiats

Magdalena Kawecka
The impact of sociodemographic factors on unemployment of young people in Poland 22
Wpływ czynników społeczno-demograficznych na bezrobocie młodych ludzi w Polsce

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

Tomasz Klimanek, Tomasz Józefowski, Andrzej Młodak, Amelia Wardzińska-Sharif
Udostępnianie danych spisowych w przekroju siatki kilometrowej – wyzwania związane z ochroną tajemnicy statystycznej 42
Data dissemination in the kilometre grid: challenges connected to the protection of statistical confidentiality

In memoriam
Dr Kazimierz Kruszka (1940–2024) 59

Dyskusje. Recenzje. Informacje
Discussions. Reviews. Information

Mirosław Szreder
Recenzja książki Józefa Pociechy *Filozoficzne i metodologiczne podstawy współczesnych analiz statystycznych* 61
Review of Józef Pociecha's book *Philosophical and methodological foundations of modern statistical analyses*

Joanna Sadowy
Wydawnictwa GUS. Czerwiec 2024 65
Publications of Statistics Poland. June 2024

Dla autorów 67
For the authors

Stałe działy „WS" – zakres tematyczny 80
Permanent sections of *WS* – thematic scope

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(7), IV–V
Od redakcji / From the editorial team

OD REDAKCJI

W lipcowym numerze „Wiadomości Statystycznych. The Polish Statistician" proponujemy Państwu lekturę artykułów z działów Statystyka w praktyce oraz Spisy powszechne – problemy i wyzwania.

Dr hab. Barbara Dańska-Borsiak, prof. UŁ, autorka artykułu *Ocena adekwatności PKB per capita jako miary poziomu życia w powiatach*, stawia pytanie, w jakim stopniu PKB per capita można uznać za wystarczający wskaźnik poziomu życia w powiatach. Aby na nie odpowiedzieć, najpierw szacuje PKB per capita dla powiatów, wykorzystując dane z Banku Danych Lokalnych GUS za lata 2019 i 2020 i stosując zmodyfikowane przez siebie metody proponowane w literaturze, a następnie zestawia wartości tego wskaźnika z wartościami syntetycznej miary poziomu życia (MPZ) skonstruowanej za pomocą metody Hellwiga i porównuje ich rozkłady przestrzenne. Z zależności między analizowanymi cechami wynika, że ocena poziomu życia ludności tylko na podstawie PKB per capita byłaby nieadekwatna. Dodatkowo autorka bada wpływ pandemii COVID-19 na sytuację w powiatach ocenianą na podstawie PKB per capita i MPZ. Stwierdza, że w 2020 r. w porównaniu z 2019 r. nie nastąpiły istotne zmiany w rozkładach przestrzennych obu zmiennych ani znaczące przesunięcia w rankingach powiatów.

W pracy *The impact of sociodemographic factors on unemployment of young people in Poland* mgr Magdalena Kawecka identyfikuje czynniki wpływające na poziom bezrobocia młodych ludzi, tj. będących w wieku od 18 do 30 lat. Analiza, przeprowadzona przez autorkę z użyciem modelu logistycznego, opiera się na danych z Bilansu Kapitału Ludzkiego za 2019 r. realizowanego przez Polską Agencję Rozwoju Przedsiębiorczości we współpracy z Uniwersytetem Jagiellońskim w Krakowie. Z badania wynika, że na ryzyko bezrobocia istotny wpływ mają płeć i poziom wykształcenia: kobiety są znacznie bardziej narażone na bezrobocie niż mężczyźni, a wyższy poziom wykształcenia zwiększa szanse na zatrudnienie.

Udostępnianie danych spisowych w przekroju siatki kilometrowej – wyzwania związane z ochroną tajemnicy statystycznej to temat artykułu dr. hab. Tomasza Klimanka, prof. UEP, dr. Tomasza Józefowskiego, dr. hab. Andrzeja Młodaka, prof. UK, i mgr Amelii Wardzińskiej-Sharif. Autorzy omawiają najważniejsze działania dotyczące kontroli ujawniania danych prezentowanych w przekroju siatki kilometrowej na przykładzie wyników Narodowego Spisu Powszechnego Ludności i Mieszkań 2021. Zwracają uwagę na najistotniejsze problemy związane z podjęciem tego rodzaju działań ochronnych w zależności od zakresu udostępnianych informacji i wskazują możliwe do zastosowania metody i narzędzia kontroli, w szczególności niezakłóceniowe, czyli prowadzące do ukrywania danych wrażliwych.

Ponadto w tym numerze żegnamy zmarłego niedawno dr. Kazimierza Kruszkę, byłego dyrektora Urzędu Statystycznego w Poznaniu, cieszącego się wielkim uznaniem i sympatią nauczyciela akademickiego, eksperta i popularyzatora statystyki.

Zamieszczamy również recenzję książki Józefa Pociechy *Filozoficzne i metodologiczne podstawy współczesnych analiz statystycznych*. W opinii recenzenta – prof. dr. hab. Mirosława Szredera – monografia może zainteresować nie tylko ekonomistów.

W numerze znajdą Państwo także omówienie najnowszych publikacji GUS przygotowane przez Joannę Sadowy.

Życzymy miłej lektury.

© Główny Urząd Statystyczny
Artykuł udostępniony na licencji CC BY-SA 4.0 / Article available under the CC BY-SA 4.0 licence

FROM THE EDITORIAL TEAM

The July issue of *Wiadomości Statystyczne. The Polish Statistician* features articles in the Statistics in practice and Issues and challenges in census taking sections.

In her article, entitled *Assessment of the adequacy of GDP* per capita *in measuring the standard of living in Polish powiats*, Barbara Dańska-Borsiak, PhD, DSc, professor at the University of Lodz, poses a question as to what extent can GDP *per capita* be considered a sufficient indicator of the standard of living in powiats (second-level administrative division in Poland). In order to answer it, the author first estimates GDP *per capita* for the powiats using data from the Local Data Bank of Statistics Poland for the years 2019 and 2020 and applies methods proposed in the literature, but in her modified version. The values of this indicator are then compared with the values of the synthetic measure of the standard of living, constructed using Hellwig's method, and their spatial distributions are compared. The relationship between the analysed characteristics shows that basing the assessment of the population's standard of living only on GDP *per capita* would be inadequate. In addition, the author examines the impact of the COVID-19 pandemic on the situation in powiats assessed using GDP *per capita* and the synthetic measure of the standard of living. The findings show that no significant changes occurred in the spatial distributions of the two variables in 2020 compared to 2019, nor did any significant shifts in the powiat rankings appear.

Magdalena Kawecka, MSc, in her article, entitled *The impact of sociodemographic factors on unemployment of young people in Poland*, identifies the factors affecting the level of unemployment of young people, i.e. those aged between 18 and 30. The analysis conducted with the use of a logistic model is based on data from the Human Capital Study for 2019. The survey was carried out by the Polish Agency for Enterprise Development in cooperation with the Jagiellonian University in Kraków. The findings show that the risk of unemployment is greatly influenced by gender and level of education. Consequently, women are significantly more likely to face unemployment than men, while a higher level of education increases the chances of finding a job.

Tomasz Klimanek, PhD, DSc, professor at the Poznań University of Economics and Business, Tomasz Józefowski, PhD, Andrzej Młodak, PhD, DSc, professor at the University of Kalisz, and Amelia Wardzińska-Sharif, MSc, in their article focus on *Data dissemination in the kilometre grid: challenges connected to the protection of statistical confidentiality*. The paper describes the most important measures designed to control the disclosure of data presented in a kilometre grid on the example of the results of the National Population and Housing Census 2021. The authors indicate the most significant problems related to such protective measures depending on the scope of the shared information. The authors also present methods and tools of control that can be applied, especially non-perturbative ones, i.e. leading to the suppression of sensitive data.

In this issue, we say our final goodbyes to Dr Kazimierz Kruszka. He was the former director of the Statistical Office in Poznań, a highly respected and popular university teacher, an expert in statistics and its active promoter.

We also include a review of a book by Józef Pociecha, entitled *Philosophical and methodological foundations of modern statistical analyse.* According to the reviewer, Mirosław Szreder, PhD, DSc, ProfTit, the monograph may be of interest not only to economists.

This month's issue also includes an overview of the latest publications of Statistics Poland, prepared by Joanna Sadowy.

We wish you a pleasant reading.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(7), 1–21 https://doi.org/10.59139/ws.2024.07.1
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 24.02.2023, poprawiony/revised: 2.02.2024, zaakceptowany/accepted: 25.03.2024

Ocena adekwatności PKB per capita jako miary poziomu życia w powiatach

Barbara Dańska-Borsiak[a]

Streszczenie. Rozwój społeczny i poprawa warunków materialnych spowodowały, że dla określenia poziomu życia coraz mniej istotny jest produkt krajowy brutto (PKB), a coraz większego znaczenia nabierają aspekty niematerialne, nieuwzględniane w tym mierniku. Głównym celem badania omawianego w artykule jest ocena, w jakim stopniu PKB per capita można uznać za wystarczający wskaźnik poziomu życia w powiatach. Dodatkowe zamierzenia to stworzenie rankingów i wskazanie powiatów o najwyższych wartościach obu analizowanych cech oraz sprawdzenie, czy wpływ pandemii COVID-19 był zróżnicowany w powiatach i czy pandemia zmieniła relację zachodzącą między badanymi cechami.

Przeprowadzone analizy obejmowały oszacowanie PKB per capita dla powiatów, z zastosowaniem autorskiej modyfikacji metod proponowanych w literaturze, zestawienie wartości tego wskaźnika z wartościami syntetycznej miary poziomu życia (MPZ) skonstruowanej za pomocą metody Hellwiga i porównanie ich rozkładów przestrzennych. Wykorzystano dane dla 380 powiatów zaczerpnięte z Banku Danych Lokalnych GUS. Do analizy wybrano dane za lata 2019 i 2020.

Na podstawie wartości współczynnika korelacji Spearmana stwierdzono, że w większości powiatów wysoki/niski poziom PKB per capita współwystępuje z wysokim/niskim poziomem życia. Istnieją jednak, głównie w Polsce południowo-wschodniej, miasta na prawach powiatu o bardzo wysokich wartościach MPZ i PKB per capita zawierających się w trzeciej grupie kwartylowej, otoczone powiatami, w których wartości MPZ i PKB per capita są o wiele niższe. Zależność między badanymi cechami jest umiarkowana i porównywalna w analizowanych latach, a zatem ocena poziomu życia ludności tylko na podstawie PKB per capita byłaby nieadekwatna. W 2020 r. w porównaniu z 2019 r. nie nastąpiły istotne zmiany w rozkładach przestrzennych obu zmiennych ani znaczące przesunięcia w rankingach powiatów.

Słowa kluczowe: PKB per capita, poziom życia, rozwój społeczny, powiat, miara syntetyczna Hellwiga

JEL: I31, O40, R11, R12

Assessment of the adequacy of GDP *per capita* in measuring the standard of living in Polish powiats

Abstract. Social development and the improving material living conditions have caused GDP to become increasingly less important in determining the standard of living, while intangible aspects, which are not included in the measure, are becoming dominant in this respect. The main aim of the study described in the article is to assess the extent to which GDP *per capita* may be considered a sufficient indicator of the standard of living in Polish powiats (second-level administrative division in Poland). Additional objectives are: to create rankings and identify the powiats with the highest values of the two analysed characteristics, to determine whether

[a] Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny, Polska / University of Lodz, Faculty of Economics and Sociology, Poland. ORCID: https://orcid.org/0000-0002-3264-3954. E-mail: barbara.danska@uni.lodz.pl.

© Barbara Dańska-Borsiak
Artykuł udostępniony na licencji CC BY-SA 4.0 / Article available under the CC BY-SA 4.0 licence

the impact of the COVID-19 pandemic varied across powiats and to assess whether the pandemic changed the relationship between the studied characteristics.

The conducted analyses included the estimation of powiats' GDP *per capita* using the author's modification of the methods proposed in the literature. Moreover, the values of this indicator were compared with the values of the synthetic measure of the standard of living based on Hellwig's composite measure, as were their spatial distributions. Data concerning 380 powiats for the years 2019 and 2020 were extracted from the Local Data Bank of Statistics Poland.

Based on the value of Spearman's correlation coefficient, it was found that in most powiats a high/low GDP *per capita* co-occurred with a high/low standard of living. However, there are cities with powiat status (mainly in south-eastern Poland) with a very high standard of living and values of GDP *per capita* in the third quartile group, which are surrounded by powiats with a much lower GDP and standard of living. The correlation between the studied characteristics is moderate and comparable in the analysed years, therefore an assessment of the population's standard of living based on GDP *per capita* only would prove inadequate. In 2020, compared to 2019, no significant changes occurred in the spatial distributions of the two variables nor were any significant shifts in rankings of powiats observed.

Keywords: GDP *per capita*, standard of living, social development, powiat, Hellwig's composite measure

1. Wprowadzenie

Tradycyjnie stosowanym wskaźnikiem kondycji gospodarki – który określa również poziom życia – jest produkt krajowy brutto (PKB). Wraz z poprawą warunków materialnych i rozwojem społecznym coraz większego znaczenia dla oceny rzeczywistości społeczno-gospodarczej nabierają jednak czynniki nieuwzględniane w PKB, takie jak sytuacja mieszkaniowa, stan opieki zdrowotnej i socjalnej, dostęp do edukacji i kultury oraz wiele innych. Krytycznej ocenie PKB i PKB per capita jako wielkościom odzwierciedlającym poziom życia towarzyszą próby wskazania innych mierników dobrobytu i jakości życia.

Kwestie precyzyjnego zdefiniowania pojęć i pomiaru poziomu i jakości życia są przedmiotem zainteresowania krajowych i międzynarodowych organizacji oraz instytucji statystyki publicznej. Podmioty te prowadzą badania i podejmują inicjatywy mające na celu stworzenie systemu wskaźników poziomu i jakości życia. Dotyczy to skali kraju. Ponieważ jednak regiony w krajach różnią się pod względem rozwoju społeczno-gospodarczego, w tym poziomu życia, równie ważne jest prowadzenie badań w przekroju regionalnym.

Głównym celem badania omawianego w artykule jest ocena, w jakim stopniu PKB per capita można uznać za wystarczający wskaźnik poziomu życia w powiatach. Dodatkowe zamierzenia to stworzenie rankingów i wskazanie powiatów o najwyższych wartościach obu analizowanych cech oraz sprawdzenie, czy wpływ pandemii COVID-19 był zróżnicowany w powiatach i czy pandemia zmieniła relację zachodzącą między badanymi cechami.

2. Przegląd literatury

Nie negując opinii wyrażonej m.in. przez Samuelsona i Nordhausa (2010, s. 370), że PKB jest najbardziej wszechstronnym i najdokładniejszym miernikiem produkcji całkowitej w gospodarce, trzeba jednakże uznać argumenty Coyle (2014). Ekonomistka stwierdza, że PKB nie jest adekwatną miarą gospodarek XXI w., których siłą napędową są innowacje, usługi i dobra niematerialne, choć przyznaje, że – szczególnie w XX w. – dobrze odzwierciedlał on warunki stwarzane przez gospodarkę rynkową. Sytuacja w tym zakresie zaczęła się zmieniać na przełomie XX w. i XXI w., kiedy przyspieszenie wzrostu gospodarczego pociągnęło za sobą wzrost aspiracji społeczeństw i przypisywanie coraz większego znaczenia jakości życia. Jak zauważają badacze, m.in. Chmielewska (2022) i Panek (2016), adekwatna ocena poziomu zaspokojenia potrzeb ludności musi zawierać informacje o dysproporcjach między grupami społecznymi związanych z podziałem dochodu. Dokonanie takiej oceny nie jest możliwe wyłącznie na podstawie PKB per capita. Panek (2016) podkreśla, że w krajach rozwiniętych wzrost PKB może wywierać negatywny wpływ na poziom życia, głównie z powodu zanieczyszczenia środowiska naturalnego, które wiąże się z rozwojem gospodarczym. Ta teza jest zbieżna z koncepcją środowiskowej krzywej Kuznetsa (Grossman i Krueger, 1995), zgodnie z którą po osiągnięciu pewnego poziomu dochodów negatywny wpływ rozwoju gospodarczego na środowisko zanika, co ma związek z gotowością wysoko rozwiniętych gospodarek do ponoszenia nakładów na ochronę środowiska. Przegląd współczesnych badań weryfikujących tę koncepcję przedstawiła np. Genstwa (2020).

Istnieje konsensus co do tego, że poziom zaspokojenia potrzeb można ocenić jedynie z uwzględnieniem jego wymiaru obiektywnego i subiektywnego, jednak stworzenie uniwersalnego wskaźnika wydaje się nierealne. Ponadto terminologia dotycząca kategorii poziomu i jakości życia stosowana w literaturze jest nieprecyzyjna. Mimo że w rzeczywistości istnieją istotne różnice między pojęciami *dobrobyt*, *poziom życia* i *jakość życia*, bywają one używane zamiennie przez uczestników życia społeczno-politycznego. Obszernie omówili te zagadnienia m.in. Chmielewska (2022), Dąbrowska (2017), Panek (2016) i Słaby (2017).

Organizacje krajowe i międzynarodowe oraz podmioty statystyki publicznej prowadzą badania dotyczące poziomu i jakości życia. Na gruncie polskim jednym z najszerzej zakrojonych badań tego typu była Diagnoza Społeczna – projekt badawczy realizowany przez Radę Monitoringu Społecznego. W latach 2000–2015 odbyło się osiem edycji panelowego badania warunków i jakości życia w Polsce[1]. Od 2006 r. Główny Urząd Statystyczny (GUS) przeprowadza co roku Europejskie Badanie Warunków Życia Ludności (EU-SILC), z zastosowaniem metodologii uspójnionej dla

[1] Raporty, kwestionariusze i instrukcje oraz bazy danych są dostępne na stronie http://www.diagnoza.com.

obszaru UE. Raporty z tego badania[2] zawierają dane na temat zróżnicowania sytuacji w Polsce: dochodów, zdrowia osób w wieku 16 lat i więcej, warunków mieszkaniowych, wyposażenia gospodarstw domowych w dobra trwałego użytkowania, ocen respondentów dotyczących wybranych materialnych i niematerialnych aspektów jakości życia, a także wybrane wskaźniki spójności społecznej. Dane prezentowane są w przekrojach regionów NUTS 2, klas miejscowości i grup społeczno-ekonomicznych[3]. Innym badaniem cyklicznie prowadzonym przez GUS (od 2011 r.) jest wieloaspektowe Badanie spójności społecznej (GUS, 2019, 2020).

Na gruncie europejskim prace nad stworzeniem koncepcji pomiaru jakości i poziomu życia podjął Eurostat. Dostrzeżona konieczność opracowania odpowiednich wskaźników, aby sprostać globalnym wyzwaniom, takim jak zmiany klimatu, ubóstwo, wyczerpywanie się zasobów, zdrowie i jakość życia, zaowocowała inicjatywą Beyond GDP, w ramach której powstały liczne publikacje, m.in. raport Stiglitza i in. (2009) oraz – wydana również w polskiej wersji językowej – książka Stiglitza i in. (2019). W 2008 r. powołano Komisję ds. Wydajności Ekonomicznej i Postępu Społecznego (Commission on the Measurement of Economic Performance and Social Progress), zwaną też komisją Stiglitza-Sena-Fitoussiego. Celem jej działań było określenie ograniczeń PKB jako wskaźnika efektywności gospodarczej i postępu społecznego. W raporcie Stiglitza i in. (2009) potwierdzono, że PKB jest niewystarczającą miarą do pomiaru dobrobytu i jego zmian w czasie, zwłaszcza w wymiarach gospodarczym, środowiskowym i społecznym ujmowanych łącznie. Autorzy zasugerowali też pomiar postępu społecznego i niektórych aspektów jakości życia za pomocą wskaźników niepieniężnych, mogących uzupełnić informacje, których dostarczają konwencjonalne wskaźniki ekonomiczne. Zwrócili przy tym uwagę, że cele rozwojowe są zróżnicowane i względne oraz zależą od systemu wartości wyznawanych przez społeczeństwa i ich potrzeb. Uznawszy niemożność stworzenia uniwersalnego wskaźnika do pomiaru stopnia realizacji celów gospodarczych, komisja zaproponowała, aby pomiar dobrobytu oprzeć na ośmiu domenach, a mianowicie: *warunki materialne*, *aktywność ekonomiczna i warunki pracy*, *służba zdrowia*, *edukacja*, *czas wolny i relacje społeczne*, *jakość środowiska przyrodniczego*, *bezpieczeństwo* oraz *państwo i podstawowe prawa obywateli*, które powinny być rozważane łącznie (Stiglitz i in., 2009, s. 14–15).

Zgodnie z tymi rekomendacjami w 2011 r. Organizacja Współpracy Gospodarczej i Rozwoju stworzyła interaktywne narzędzie OECD Better Life Initiative, dostępne na platformie https://www.oecdbetterlifeindex.org, które dostarcza danych dotyczących

[2] Raporty, publikowane od 2006 r., są dostępne na stronie https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/dochody-i-warunki-zycia-ludnosci-polski-raport-z-badania-eu-silc-2022,6,16.html.

[3] Klasy miejscowości to: miasta o liczbie ludności (w tys.): 500 i więcej, 200–499, 100–199, 20–99, poniżej 20 oraz wieś. Grupy społeczno-ekonomiczne to gospodarstwa domowe: pracowników, rolników, pracujących na własny rachunek, emerytów, rencistów oraz utrzymujących się ze źródeł niezarobkowych.

11 obszarów uznanych za istotne wyznaczniki warunków materialnych i jakości życia. Za jego pomocą – poprzez rangowanie poszczególnych mierników zgodnie z przypisywanym im indywidualnie znaczeniem – można stworzyć własny wskaźnik Better Life Index.

Opisane powyżej próby stworzenia systemów wskaźników poziomu i jakości życia, tak jak i inne inicjatywy instytucji publicznych były podejmowane w skali krajów. Jeśli chodzi o badania poziomu i jakości życia w przekroju regionalnym, to w Polsce analizy takie prowadziło wielu badaczy, m.in. Majecka i Nowak (2019), Panek i Zwierzchowski (2019) oraz Polak (2016). Mimo znacznie mniejszej dostępności danych statystycznych dotyczących powiatów podejmowano też próby konstrukcji miar poziomu życia w tych jednostkach – opisano je m.in. w publikacjach: Dańskiej-Borsiak (2022a, 2022c), Kusterki-Jefmańskiej i Jefmańskiego (2017), Sobolewskiego i in. (2014), Szaban (2019) oraz Wolaka (2020).

3. Metoda badania

Zrealizowanie celów badania omawianego w niniejszym artykule wymagało oszacowania PKB w powiatach, ponieważ dane dla tych jednostek nie są dostępne. Na podstawie danych pierwotnych, pochodzących z informacji źródłowych, GUS udostępnia dane dotyczące PKB według województw. Dane te, zaczerpnięte z Banku Danych Lokalnych (BDL) GUS[4], posłużyły jako baza do rozszacowania wartości PKB na powiaty leżące w obrębie danego województwa – łącznie dla 380 jednostek. Wartości te zestawiono z wartościami syntetycznej miary poziomu życia (MPZ) skonstruowanej za pomocą metody Hellwiga na podstawie danych GUS i porównano ich rozkłady przestrzenne.

Analizą objęto lata 2019 i 2020. Rok 2020 to czas najostrzejszych restrykcji związanych z pandemią COVID-19, które spowodowały spadek PKB w Polsce. Zestawienie wartości PKB per capita w tych latach pozwoliło ocenić, czy i w jaki sposób pandemia wpłynęła na gospodarkę powiatów, a porównanie wartości MPZ i siły zależności między PKB per capita a MPZ – czy pandemia zmieniła relację zachodzącą między obiema cechami.

3.1. Syntetyczna miara poziomu życia

Poziom życia powinien być oceniany w różnych aspektach, na podstawie wielu zmiennych. Z punktu widzenia wielowymiarowej analizy statystycznej jest to zjawisko złożone, którego wielkość w badanych obiektach (np. regionach) można ocenić

[4] https://bdl.stat.gov.pl/bdl.

na podstawie zmiennej syntetycznej, będącej wypadkową zmiennych determinujących zjawisko, zwanych *zmiennymi diagnostycznymi*.

W literaturze przedmiotu proponowane są wzorcowe i bezwzorcowe metody konstrukcji zmiennych syntetycznych. Do najczęściej stosowanych metod wzorcowych należą miara rozwoju zaproponowana przez Hellwiga (1968) i metoda TOPSIS (Technique for Order Preference by Similarity to Ideal Solution; Hwang i Yoon, 1981). Pierwsza opiera się na porównaniu każdego obiektu z wzorcem rozwojowym, czyli sztucznym obiektem reprezentowanym przez wektor najlepszych – z punktu widzenia rozpatrywanego zjawiska złożonego – wartości każdej zmiennej diagnostycznej, a druga wykorzystuje również porównanie z antywzorcem, czyli obiektem opisanym przez wektor o najgorszych wartościach tych zmiennych. Metody bezwzorcowe opierają się natomiast na obliczeniu sum albo średnich ze znormalizowanych wartości zmiennych diagnostycznych. Różnice polegają na wyborze sposobu normalizacji i obliczania średniej (prostej lub ważonej i doborze wag). Dokładniejszy opis metod normalizacji zmiennych i metod konstrukcji zmiennych syntetycznych można znaleźć m.in. w pracy Balickiego (2013, s. 318–329).

W prezentowanym badaniu syntetyczną miarę odzwierciedlającą poziom życia obliczono na podstawie siedmiu wskaźników cząstkowych opisujących wymiary poziomu życia (domeny), które zostały wymienione w raporcie Stiglitza i in. (2009, s. 14–15). Pominięto domenę *państwo i podstawowe prawa obywateli,* ponieważ nie ma różnic między powiatami pod tym względem.

Poniżej wymieniono zmienne charakteryzujące kolejne domeny. Wszystkie zostały poddane wstępnej weryfikacji statystycznej – wyeliminowano zmienne zbyt nisko zróżnicowane, o współczynniku zmienności poniżej 10%, i zbyt silnie ze sobą skorelowane.

Zestawienie. Zmienne charakteryzujące domeny uwzględnione w konstrukcji MPZ

Symbol zmiennej	Nazwa i jednostka miary	Typ
	Warunki materialne	
X1	przeciętne miesięczne wynagrodzenie brutto w zł	stymulanta
X2	przeciętna powierzchnia użytkowa mieszkania w m^2 na osobę	stymulanta
X3	liczba mieszkań na 1000 mieszkańców	stymulanta
X4	odsetek ludności korzystającej z sieci wodociągowej	stymulanta
X5	beneficjenci środowiskowej pomocy społecznej na 10 000 ludności	destymulanta

Zestawienie. Zmienne charakteryzujące domeny uwzględnione w konstrukcji MPZ (dok.)

Symbol zmiennej	Nazwa i jednostka miary	Typ
	Aktywność ekonomiczna i warunki pracy	
X6	pracujący na 1000 ludności w wieku produkcyjnym	stymulanta
X7	stopa bezrobocia rejestrowanego w %	destymulanta
X8	udział bezrobotnych zarejestrowanych pozostających bez pracy dłużej niż rok w ogólnej liczbie bezrobotnych w %	destymulanta
X9	poszkodowani w wypadkach przy pracy na 1000 pracujących	destymulanta
	Służba zdrowia	
X10	liczba porad lekarskich na 1000 mieszkańców	stymulanta
X11	lekarze (personel pracujący) na 10 000 ludności	stymulanta
X12	pielęgniarki i położne na 10 000 ludności	stymulanta
	Edukacja	
X13	dzieci w placówkach wychowania przedszkolnego na 1000 dzieci w wieku 3–5 lat	stymulanta
X14	uczniowie szkół ponadgimnazjalnych w stosunku do ludności w wieku 15–18 lat w %	stymulanta
X15	studenci szkół wyższych na 1000 ludności w wieku 20–24 lat	stymulanta
	Czas wolny i relacje społeczne	
X16	rozwody i separacje na 10 000 ludności	destymulanta
X17	małżeństwa zawarte na 1000 ludności	stymulanta
X18	czytelnicy bibliotek publicznych na 1000 ludności	stymulanta
X19	udział członków kół, klubów i sekcji działających w centrach, domach kultury i świetlicach w liczbie ludności ogółem w %	stymulanta
X20	udział ćwiczących w klubach sportowych w liczbie ludności ogółem w %	stymulanta
	Bezpieczeństwo	
X21	wydatki w Dziale 754 – Bezpieczeństwo publiczne i ochrona przeciwpożarowa na mieszkańca w zł	stymulanta
X22	przestępstwa stwierdzone przez Policję ogółem na 1000 mieszkańców	destymulanta
X23	wskaźnik wykrywalności sprawców przestępstw stwierdzonych przez Policję w %	stymulanta
X24	liczba prób (zachowań) samobójczych na 100 000 ludności	destymulanta
	Jakość środowiska przyrodniczego	
X25	udział ludności korzystającej z oczyszczalni ścieków w %	stymulanta
X26	emisja zanieczyszczeń gazowych i pyłowych w tonach na 1 km^2 powierzchni	destymulanta
X27	udział odpadów zebranych selektywnie w %	stymulanta
X28	udział wydatków w Dziale 900 – Gospodarka komunalna i ochrona środowiska w wydatkach ogółem w %	stymulanta

Źródło: opracowanie własne na podstawie danych z BDL.

Każdy wskaźnik cząstkowy dla danej domeny został wyznaczony jako średnia arytmetyczna zunitaryzowanych zmiennych diagnostycznych, a następnie wskaźniki zostały zagregowane w celu utworzenia MPZ. Zastosowano cztery metody: Hellwiga, TOPSIS, porządkowania według pierwszej głównej składowej i metodę sum standaryzowanych, a wyboru najlepszej z nich dokonano na podstawie agregatowego miernika oceny poprawności metod porządkowania liniowego Q^* (Kądziołka, 2021). Miernik Q^* jest pierwiastkiem z sumy kwadratów mierników cząstkowych opisujących zgodność odwzorowania (bazującą na porównaniu odległości euklidesowej między każdą parą obiektów w jednowymiarowej przestrzeni zmiennej syntetycznej i w wielowymiarowej przestrzeni zmiennych diagnostycznych), korelację rang między zmienną syntetyczną a zmiennymi diagnostycznymi, zmienność i koncentrację zmiennej syntetycznej oraz podobieństwo rankingów uzyskanych za pomocą rozważanych zmiennych syntetycznych. Najpoprawniejsza w tym ujęciu okazała się metoda Hellwiga, zatem dalszą analizę oparto na wynikach uzyskanych tą metodą. Jako że jest to metoda wzorcowa, pierwszy krok polega na skonstruowaniu wzorca. Jest to wektor $[z_{01}, \dots, z_{0m}]$, którego j-a współrzędna $z_{0j} = \max_i z_{ij}$, jeśli j-a zmienna jest stymulantą, albo $z_{0j} = \min_i z_{ij}$ w przypadku destymulanty (z_{ij} to wartość standaryzowanej j-ej zmiennej w i-tym obiekcie, $j = 1, \dots, m$, $i = 1, \dots, n$). Następnie wylicza się odległość każdego obiektu od wzorca: $d_{i0} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(z_{ij} - z_{0j}\right)^2}$. Wartości zmiennej syntetycznej wyznacza się według wzoru:

$$MH_i = 1 - \frac{d_{i0}}{d_0}, \tag{1}$$

gdzie $d_0 = \bar{d}_0 + 2s_d$, $\bar{d}_0$ i s_d to odpowiednio średnia arytmetyczna i odchylenie standardowe odległości obiektów od wzorca d_{i0}.

Do badania zależności między wartościami danej cechy w powiatach w latach 2019 i 2020 (dotyczy to PKB per capita i MPZ), a także różnych cech w danym roku zastosowano współczynnik korelacji rang Spearmana.

3.2. PKB wytworzony w powiatach

W literaturze proponowane są różne metody rozszacowania wojewódzkich wartości PKB na powiaty. Tokarski (2013) estymował PKB w powiatach na podstawie modeli regresji objaśniających PKB w województwach w zależności od produkcji sprzedanej, wartości brutto środków trwałych i inwestycji. Z tych równań wyliczył wstępne oszacowania PKB w powiatach, podstawiając za zmienne niezależne wartości powiatowe produkcji sprzedanej, wartości brutto środków trwałych i inwestycji. Jako ostateczne

wartości PKB dla powiatów przyjął średnią geometryczną oszacowań wynikających z każdego równania. Tę propozycję zmodyfikowali Dykas i Misiak (2018), którzy wyestymowali dodatkowo model PKB w zależności od funduszu płac w województwie i włączyli uzyskane oszacowanie do średniej. Dwie alternatywne propozycje zaprezentowała Ciołek (2017), która wyznaczyła PKB w powiatach proporcjonalnie do ich dochodów budżetowych z tytułu podatku dochodowego od osób fizycznych (PIT) i z podatku rolnego albo proporcjonalnie do udziałów powiatów w wojewódzkim funduszu płac. Ostatecznie badaczka uznała, że wyznaczenie PKB proporcjonalnie do dochodów podatkowych lepiej odzwierciedla rzeczywisty rozkład PKB w powiatach.

W prezentowanym badaniu zastosowano autorską procedurę, będącą modyfikacją podejścia zaproponowanego przez Ciołek (2017). Wydaje się, że autorka tej pracy zbyt pochopnie odrzuciła oszacowanie PKB na podstawie funduszu płac jako nieodzwierciedlające procesu tworzenia produkcji i niebiorące pod uwagę rachunku rozdysponowania. Nie polemizując z tymi argumentami, uznano, że uwzględnienie poziomu wynagrodzeń może pomóc zmniejszyć błąd szacunku PKB na podstawie PIT wynikający z odprowadzania podatku poza miejscem pracy. Warto bowiem zauważyć, że wiele osób z mniejszych miejscowości dojeżdża do pracy do większych ośrodków, więc wytwarzają one produkcję w innych miejscach, niż płacą podatki. Omawiana autorska metoda polega na wyznaczeniu średniej geometrycznej z wartości PKB uzyskanych na podstawie PIT i podatku rolnego oraz na podstawie funduszu płac. Została ona przetestowana w badaniu pilotażowym powiatów woj. łódzkiego, którego wyniki, wraz ze szczegółowym opisem procedury, przedstawiono w pracach Dańskiej-Borsiak (2022b, 2022c).

Rozszacowania PKB wytworzonego w każdym województwie na powiaty na podstawie udziałów powiatów w dochodach z PIT i podatku rolnego dokonano zgodnie z formułą:

$$PKB_{p,tax} = uPIT_{\frac{p}{w}} \cdot \left(1 - uWDB_{agr}\right) \cdot PKB_w + \\ + uPR_{p/w} \cdot uWDB_{agr} \cdot PKB_w, \tag{2}$$

gdzie:

$uPIT_{p/w}$ – stosunek dochodów budżetu powiatu p w dochodach województwa w z tytułu podatku dochodowego od osób fizycznych,

$uPR_{p/w}$ – stosunek dochodów budżetu powiatu p w dochodach województwa w z tytułu podatku rolnego,

$uWDB_{agr}$ – udział wartości dodanej brutto województwa w wytworzonej w sekcji A (rolnictwo, leśnictwo, łowiectwo i rybactwo).

Rozszacowania PKB wytworzonego w każdym województwie proporcjonalnie do udziału powiatowych funduszów płac w funduszu wojewódzkim dokonano zgodnie z formułą:

$$PKB_{p,wage} = \frac{W_p \cdot LP_p}{W_w \cdot LP_W} \cdot PKB_w, \tag{3}$$

gdzie:

W_p, W_w – przeciętne miesięczne wynagrodzenie brutto odpowiednio w powiecie p i w województwie w,

LP_p, LP_w – liczba pracujących odpowiednio w powiecie p i w województwie w.

Ostateczne wartości PKB w powiecie p wyznaczono jako średnią geometryczną:

$$PKB_p = \sqrt{PKB_{p,tax} \cdot PKB_{p,wage}}\,. \tag{4}$$

Argumentem przemawiającym za zastosowaniem takiego podejścia jest porównanie oszacowań powiatowych wartości PKB z odpowiednimi danymi publikowanymi przez GUS dla podregionów. W Polsce istnieje sześć podregionów tożsamych z właściwymi powiatami grodzkimi (miasta: Warszawa, Kraków, Łódź, Poznań, Wrocław i Szczecin). Przeciętny względny błąd uzyskanych wyników w stosunku do oficjalnych danych GUS za 2019 r. dla tych podregionów wyniósł 8,2%, a przy porównaniu oszacowań na podstawie PIT i podatku rolnego – 9,8%.

4. Wyniki badania

4.1. Poziom życia w powiatach mierzony za pomocą MPZ

Wartości MPZ mieszczą się w przedziale $[0, 1]$. W tabl. 1 zamieszczono wartości statystyk opisowych tej miary. W szczytowym roku pandemii można zauważyć wzrost wartości średniej, a także maksimum i minimum w stosunku do roku przed pandemią. Prawie wcale nie zmieniło się natomiast zróżnicowanie mierzone współczynnikiem zmienności. Widoczna jest też silna asymetria rozkładu, która pogłębiła się w 2020 r. Powiatów, w których poziom życia przewyższa średnią, jest znacznie mniej niż tych o poziomie życia poniżej średniej.

Tabl. 1. Podstawowe statystyki opisowe MPZ w powiatach

Lata	Średnia	Odchylenie stand.	Współczynik zmienności w %	Min.	Maks.	Mediana	Skośność[a]
2019	0,234	0,117	49,884	0,002	0,659	0,223	0,899
2020	0,247	0,123	49,899	0,003	0,705	0,234	0,997

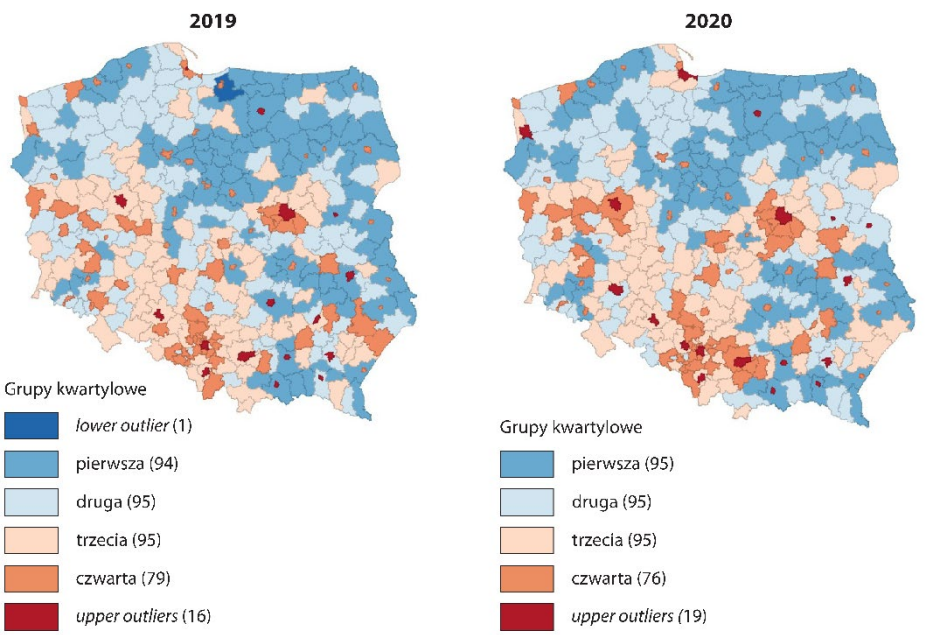
a Współczynnik skośności obliczono na podstawie wzoru $As = \mu_3/D^3(X)$, gdzie μ_3 jest trzecim momentem centralnym, a $D(X)$ to odchylenie standardowe.

Źródło: obliczenia własne na podstawie danych z BDL.

Rozkład przestrzenny MPZ w powiatach w latach 2019 i 2020 przedstawiono na mapie 1. Porównanie obu kartogramów wskazuje, że był on bardzo stabilny, co potwierdza wartość współczynnika korelacji rang Spearmana mierników dla obu lat, wynosząca 0,967.

Uwagę zwraca duża koncentracja powiatów z najniższej grupy kwartylowej w północno-wschodniej Polsce i w rejonie Kujaw. W 2020 r. w województwach warmińsko-mazurskim i podlaskim znajdowało się 24,2% powiatów z tej grupy, a w woj. kujawsko-pomorskim – 14,8%. Widoczne są też lokalne centra w południowo-wschodniej Polsce, w których poziom życia jest znacznie wyższy niż w powiatach sąsiednich. Różnice takie nie występują w zachodniej części kraju, w której przeważają powiaty z dwóch najwyższych grup kwartylowych. W 2020 r. jedynie w 2,1% powiatów województw śląskiego i opolskiego wartość MPZ była poniżej mediany.

Mapa 1. Poziom życia w powiatach mierzony za pomocą MPZ



Uwaga. Zakres grup w %: pierwsza < 25, druga [25–50), trzecia [50–75), czwarta ≥ 75.
Źródło: opracowanie własne.

W 2020 r. wśród 50 powiatów o najwyższych wartościach MPZ było tylko 5 niebędących miastami na prawach powiatu: piaseczyński, warszawski zachodni, otwocki, grodziski i pruszkowski. Wszystkie one sąsiadują z Warszawą – bezpośrednio lub przez inny powiat. Pięć miast na prawach powiatu: Bytom (97. miejsce w rankingu), Świętochłowice (130.), Piekary Śląskie (132.), Mysłowice (146.) i Włocławek (161.)

znalazło się w drugiej grupie kwartylowej, a Grudziądz (209.) – w trzeciej. Wymienione miasta śląskie miały relatywnie niskie wartości wskaźnika dla domen *czas wolny i relacje społeczne* oraz *aktywność ekonomiczna i warunki pracy*. Znajdują się one jednak w obrębie konurbacji górnośląskiej, co ułatwia mieszkańcom dostęp do obiektów rekreacyjnych i atrakcyjnych miejsc pracy w sąsiednich powiatach. Bardziej wnikliwą analizę zróżnicowania poziomu życia między powiatami w 2020 r. oraz omówienie sytuacji w domenach wymienionych w tabl. 1 zaprezentowano w pracy Dańskiej-Borsiak (2022a).

W 2019 r. wartość MPZ w przypadku jednego powiatu była niższa od pierwszego kwartyla o co najmniej półtorakrotność rozstępu kwartylowego. Był to powiat elbląski, oznaczony na mapie 1 jako *lower outlier*. W kolejnym roku poziom życia w tym powiecie podniósł się na tyle, że nie zaliczono go już do tej kategorii, jednak nadal charakteryzował się najniższą wartością MPZ. Wzrosła również liczba powiatów, w których wartość miary przekraczała trzeci kwartyl o co najmniej półtorakrotność rozstępu kwartylowego, oznaczonych na mapie jako *upper outliers* – w 2019 r. było ich 16, a w 2020 r. – 19 (tabl. 2). W 2020 r. w skład grupy *upper outliers* weszły cztery miasta na prawach powiatu: Nowy Sącz, Szczecin, Tarnów i Biała Podlaska, które rok wcześniej – na podstawie wartości miernika – znajdowały się w trzeciej grupie kwartylowej. Świadczy to o znacznym wzroście poziomu życia ich mieszkańców. Duże pogorszenie poziomu życia odnotowano natomiast w Kielcach – powiat ten spadł w rankingu o siedem miejsc i w 2020 r. nie zaliczał się do grupy *upper outliers*.

Tabl. 2. Powiaty o najwyższych wartościach MPZ

Pozycja	2019		2020		
	powiaty	MPZ	powiaty	MPZ	zmiana w stosunku do 2019
1	m. Kraków	0,659	m. Kraków	0,708	.
2	m. Katowice	0,648	m. Katowice	0,687	.
3	m. Rzeszów	0,613	m. Rzeszów	0,635	.
4	m. Krosno	0,598	m.st. Warszawa	0,634	↑2
5	m. Olsztyn	0,593	m. Sopot	0,612	↑7
6	m.st. Warszawa	0,571	m. Poznań	0,607	↑3
7	m. Lublin	0,553	m. Wrocław	0,590	↑7
8	m. Opole	0,552	m. Olsztyn	0,590	↓3
9	m. Poznań	0,548	m. Lublin	0,589	↓2
10	m. Bielsko-Biała	0,546	m. Bielsko-Biała	0,575	.
11	m. Siedlce	0,542	m. Opole	0,565	↓3
12	m. Sopot	0,534	m. Krosno	0,555	↓8
13	m. Gdańsk	0,523	m. Nowy Sącz	0,551	↑4
14	m. Wrocław	0,511	m. Siedlce	0,549	↓3
15	m. Gliwice	0,498	m. Gdańsk	0,545	↓2
16	m. Kielce[a]	0,498	m. Szczecin	0,533	↑3

Tabl. 2. Powiaty o najwyższych wartościach MPZ (dok.)

Pozycja	2019		2020		
	powiaty	MPZ	powiaty	MPZ	zmiana w stosunku do 2019
17	m. Nowy Sącz	0,492	m. Biała Podlaska	0,516	↑3
18	m. Tarnów	0,489	m. Tarnów	0,503	.
19	m. Szczecin	0,472	m. Gliwice	0,502	↓4

a W 2020 r. spadek o 7 pozycji w rankingu (23. pozycja).
Źródło: obliczenia własne na podstawie danych z BDL.

Z tabl. 2 wynika, że najwyższy poziom życia utrzymał się w tych samych powiatach, choć doszło do pewnych przesunięć rankingowych. Miasto Krosno spadło w klasyfikacji o osiem pozycji, a Wrocław i Sopot przesunęły się do góry o siedem pozycji. Wszystkie wymienione jednostki są miastami na prawach powiatu. Są to zarówno największe aglomeracje (z wyjątkiem Łodzi, która jednak w 2020 r. awansowała aż o 15 pozycji i zajęła 25. miejsce w rankingu), jak i mniejsze miasta będące lokalnymi centrami. Wyniki badania poziomu życia i sytuacji w poszczególnych jego wymiarach przedstawione w pracy Dańskiej-Borsiak (2022a) wskazują, że powiaty grupy *upper outlier*s charakteryzują się najwyższymi wartościami wskaźników cząstkowych dla wszystkich domen, co przekłada się na najwyższe wartości miary syntetycznej. Przewaga tej grupy nad pozostałymi powiatami jest widoczna przede wszystkim w domenie *edukacja*. Wszystkie opisujące ją zmienne diagnostyczne osiągają tu wysokie wartości, ale dotyczy to zwłaszcza liczby studentów szkół wyższych i dzieci w placówkach wychowania przedszkolnego. Ponadto zaliczone do *upper outliers* miasta na prawach powiatu położone we wschodniej Polsce charakteryzują się liczbą uczniów szkół ponadgimnazjalnych przewyższającą populację osób w wieku 15–18 lat, co wynika z tego, że do tych szkół uczęszczają również uczniowie z pobliskich powiatów.

4.2. PKB w powiatach

Jak wynika z danych GUS, PKB w cenach stałych wytworzony w Polsce w 2020 r. stanowił 98% PKB wytworzonego w 2019 r. W 2021 r. dynamika wróciła na ścieżkę wzrostu i PKB osiągnął 106,8% wartości z roku poprzedniego. Przyczyn jednookresowego spadku PKB upatruje się w wybuchu pandemii COVID-19 i wprowadzeniu lock downu. Przełożyło się to na sytuację województw, które również doświadczyły spadk u PKB, przy czym jego siła była zróżnicowana – wartość tego wskaźnika w 2020 r. w stosunku do 2019 r. wynosiła od 95,4% w woj. podkarpackim do 99,5% w woj. kujawsko-pomorskim. Podobnie kształtował się PKB per capita – w 2020 r. stanowił 98,1% wartości ogólnopolskiej z 2019 r., a na poziomie województw – od 95,5% w woj. podkarpackim do 99,8% w woj. warmińsko-mazurskim.

Spadek PKB i PKB per capita był zauważalny również na poziomie powiatów. Podstawowe statystyki z tego zakresu zebrano w tabl. 3. Średnia spadła o 860 zł, a mediana o prawie 700 zł na mieszkańca. Znacznie bardziej, o prawie 6000 zł, spadła wartość maksymalna (powiat m.st. Warszawa), przy wzroście minimum (powiaty piski w 2020 r. i wałbrzyski w 2019 r.) o prawie 1400 zł. Występuje również prawostronna asymetria rozkładu PKB per capita, która jest dużo silniejsza niż w przypadku MPZ.

Tabl. 3. Podstawowe statystyki opisowe PKB per capita na poziomie powiatów

Lata	Średnia	Odchylenie stand.	Współczynnik zmienności w %	Min.	Maks.	Mediana	Skośność[a]
	w tys. zł			w tys. zł			
2019	47,477	16,641	35,05	26,034	168,525	43,415	2,383
2020	46,617	16,069	34,47	27,406	162,599	42,723	2,392

a Jak przy tabl. 1.
Źródło: obliczenia własne.

Statystyki zawarte w tabl. 3 wskazują na zmniejszenie zróżnicowania wartości PKB per capita w 2020 r. – współczynnik zmienności spadł o 0,58 p.proc. Rozkład terytorialny był bardzo stabilny, na co wskazuje porównanie kartogramów na mapie 2. Współczynnik korelacji rang Spearmana PKB per capita dla analizowanych lat wynosi 0,995. Pozycje powiatów zajmujących czołowe miejsca w rankingu zmieniły się tylko nieznacznie, co pokazuje tabl. 4. Zamieszczono w niej 26 powiatów o najwyższych wartościach PKB per capita. W 2019 r. należały one do grupy *upper outliers* zaznaczonej na mapie 2, co oznacza, że PKB per capita przekracza wartość trzeciego kwartyla o co najmniej półtorakrotność rozstępu kwartylowego. W kolejnym roku grupa ta zmniejszyła liczebność do 24. Powiaty: m. Dąbrowa Górnicza i m. Kalisz znalazły się poza nią.

Tabl. 4. Powiaty o najwyższych wartościach PKB per capita

Pozycja	2019		2020		
	powiaty	PKB per capita w tys. zł	powiaty	PKB per capita w tys. zł	zmiana w stosunku do 2019
1	m.st. Warszawa	168,525	m.st. Warszawa	162,599	.
2	m. Katowice	116,168	m. Katowice	114,233	.
3	m. Poznań	113,618	m. Poznań	110,354	.
4	m. Sopot	110,707	m. Sopot	109,743	.
5	m. Wrocław	110,242	m. Wrocław	107,434	.
6	warszawski zach.	103,750	warszawski zach.	102,058	.
7	m. Kraków	98,779	m. Kraków	96,899	.

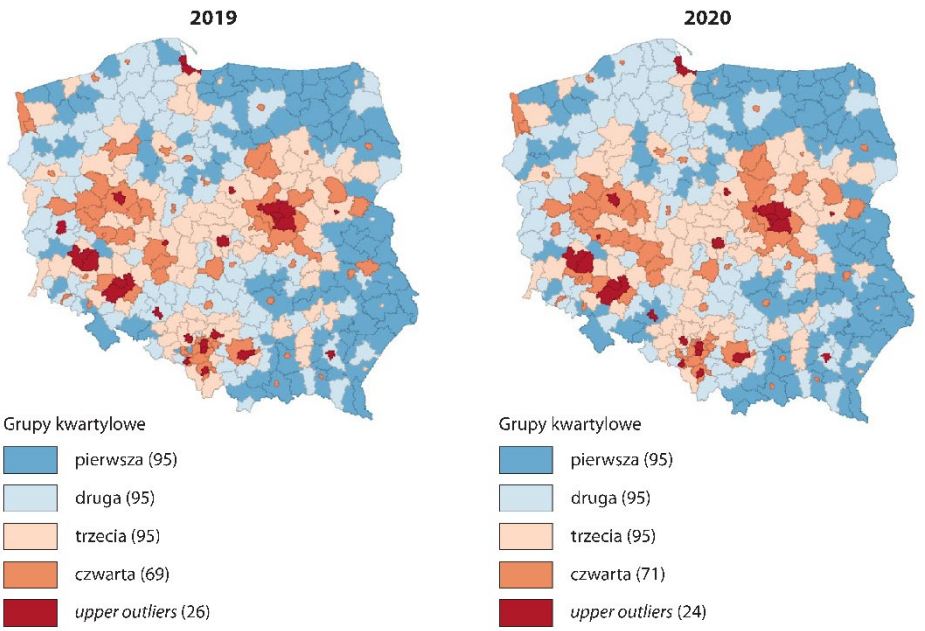
Tabl. 4. Powiaty o najwyższych wartościach PKB per capita (dok.)

Pozycja	2019		2020		
	powiaty	PKB per capita w tys. zł	powiaty	PKB per capita w tys. zł	zmiana w stosunku do 2019
8	m. Gliwice	96,747	m. Gliwice	96,119	.
9	m. Gdańsk	94,689	m. Gdańsk	92,746	.
10	m. Płock	93,729	m. Płock	91,953	.
11	m. Opole	91,137	pruszkowski	88,038	↑1
12	pruszkowski	90,517	m. Bielsko-Biała	88,031	↑1
13	m. Bielsko-Biała	89,860	m. Opole	87,829	↓2
14	piaseczyński	85,856	wrocławski	86,827	↑2
15	polkowicki	84,023	polkowicki	79,892	15
16	wrocławski	83,473	piaseczyński	79,452	↓2
17	m. Gdynia	80,623	m. Tychy	78,728	↑4
18	lubiński	80,587	m. Jastrzębie-Zdrój	77,356	↑1
19	m. Jastrzębie-Zdrój	80,457	lubiński	76,530	↓1
20	m. Rzeszów	78,771	m. Gdynia	76,164	↓3
21	m. Tychy	78,475	m. Rzeszów	75,663	↓1
22	m. Kalisz	76,992	m. Leszno	74,813	↑6
23	m. Zielona Góra[a]	76,755	m. Siedlce	74,806	↑2
24	m. Dąbrowa Górnicza	76,726	m. Łódź	74,394	↑2
25	m. Siedlce	76,598	m. Dąbrowa Górnicza	74,171	↓1
26	m. Łódź	76,586	m. Kalisz	73,692	↓4

a W 2020 r. spadek o 5 pozycji w rankingu (28. pozycja).
Źródło: obliczenia własne na podstawie danych z BDL.

Powiaty, które zajmowały pierwsze 10 miejsc w rankingu w 2019 r. (tabl. 4), pozostały na tych samych pozycjach również w 2020 r., przy czym uwagę zwraca znaczna różnica między m.st. Warszawą a znajdującymi się na drugim miejscu Katowicami. Zmiany pozycji rankingowych w większości przypadków nie przekraczają dwóch, jedynie m. Leszno awansowało o sześć pozycji, wchodząc do grupy *upper outliers* na miejsce m. Zielona Góra, które spadło o pięć pozycji.

Mapa 2. PKB per capita w powiatach



Uwaga. Jak przy mapie 1.
Źródło: opracowanie własne na podstawie danych z BDL.

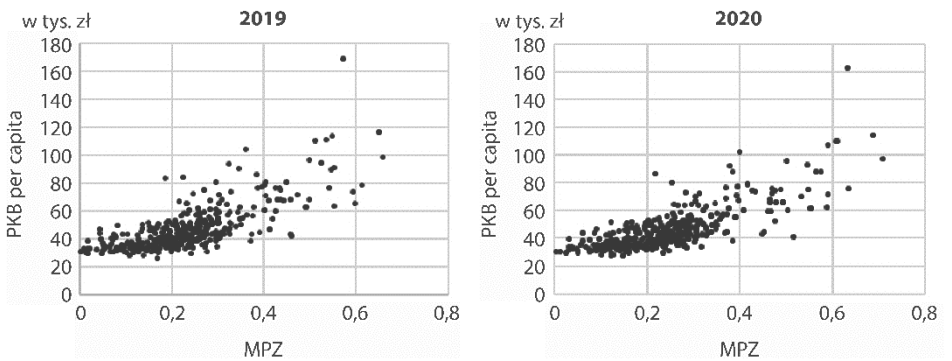
W rozkładzie przestrzennym PKB per capita pokazanym na mapie 2 widoczne jest, że powiaty o wartościach zmiennej mniejszych niż pierwszy kwartyl dominują na wschodzie i północnym wschodzie kraju. W pasie centralnym przeważają powiaty, w których PKB per capita zawiera się pomiędzy pierwszym a trzecim kwartylem. Najwyższe wartości zmiennej, przekraczające trzeci kwartyl, dominują w otoczeniu Warszawy oraz na terenach województw wielkopolskiego, dolnośląskiego i śląskiego, a największe miasta na prawach powiatu charakteryzują się PKB per capita przewyższającym trzeci kwartyl o co najmniej półtorakrotność rozstępu kwartylowego. Z porównania kartogramów wynika, że w 2020 r. podział powiatów na grupy kwartylowe w zasadzie nie zmienił się w stosunku do 2019 r., a zatem skutki pandemii dotknęły regiony z porównywalną siłą. Nawet jeśli w pierwszych miesiącach 2020 r., czyli w okresie najostrzejszego lockdownu w pewnych regionach – szczególnie tych generujących duże dochody z turystyki, rekreacji i gastronomii – wystąpiły większe spadki wartości PKB, to po poluzowaniu obostrzeń zostały one wyrównane i w skali roku wahania te nie są widoczne.

4.3. PKB per capita a poziom życia

Główne pytanie postawione w omawianym badaniu dotyczyło tego, w jakim stopniu PKB per capita można uznać za adekwatne odzwierciedlenie poziomu życia. Próbę odpowiedzi na to pytanie podjęto na podstawie zestawienia wartości tego miernika z wartościami MPZ w analizowanych latach.

Współczynnik korelacji Spearmana PKB per capita i MPZ wyniósł 0,715 w 2019 r. i 0,736 w 2020 r., co świadczy o silnej zależności między obiema zmiennymi. Jest to uwidocznione również na wykresie.

Wykres. Zależność PKB per capita i MPZ



Źródło: opracowanie własne na podstawie danych z BDL i obliczeń własnych.

Zwraca uwagę prawostronna asymetria rozkładu obu zmiennych (por. wykres z tabl. 1 i 3). Na bardzo silnej asymetrii PKB per capita zaważył poziom zmiennej w Warszawie, chociaż warto też zwrócić uwagę, że są powiaty charakteryzujące się bardzo dobrym poziomem życia przy umiarkowanych wartościach PKB per capita, jak np. Rzeszów (MPZ – 0,635, PKB per capita – 75,663 tys. zł), Lublin (odpowiednio: 0,588 i 62,086 tys. zł) i Olsztyn (odpowiednio: 0,59 i 71,277 tys. zł). Spośród wymiarów poziomu życia wymienionych w zestawieniu najsilniejsze powiązania z poziomem PKB per capita wykazują *aktywność ekonomiczna i warunki pracy* oraz *warunki materialne*. Jednak o poziomie życia w dużym stopniu decydują inne czynniki, niezwiązane bezpośrednio z wymienionymi wyżej domenami. Uzyskane wyniki świadczą o istnieniu pewnej ogólnej zależności: im wyższy PKB per capita w powiecie, tym wyższy poziom życia, na co wskazuje stosunkowo wysoka korelacja obu miar i zbliżony rozkład przestrzenny. Niemniej można zauważyć wiele odstępstw od tej reguły, zwłaszcza w powiatach, w których niższy poziom materialny jest rekompensowany lepszą sytuacją w domenach *jakość środowiska przyrodniczego*, *czas wolny i relacje*

społeczne oraz *bezpieczeństwo*. Wynika z tego, że PKB per capita w niewystarczającym stopniu odzwierciedla poziom życia w powiatach.

5. Podsumowanie

W badaniu omówionym w artykule wyznaczono syntetyczną miarę poziomu życia (MPZ) w powiatach i oszacowano PKB na mieszkańca powiatu. Zestawienie tych mierników pozwoliło stwierdzić, w jakim stopniu PKB per capita można uznać za adekwatne odzwierciedlenie poziomu życia. Do analizy wybrano rok 2019, który bezpośrednio poprzedzał wybuch pandemii COVID-19, i rok 2020, czyli czas najostrzejszych restrykcji. Okazało się, że pandemia wywarła negatywny wpływ na gospodarkę powiatów w zbliżonym stopniu. Różnice mogłyby się uwidocznić przy wykorzystaniu danych o mniejszym stopniu agregacji, np. miesięcznych; w skali roku uległy zatarciu.

Aby obliczyć PKB dla powiatów (dane niedostępne w bazach GUS), zastosowano autorską metodę szacowania tych wielkości na podstawie danych wojewódzkich udostępnianych przez GUS. Wartości MPZ wyznaczono za pomocą metody Hellwiga.

Najwyższymi wartościami MPZ i PKB per capita charakteryzowały się miasta na prawach powiatu, szczególnie największe aglomeracje. Trzeba jednak zwrócić uwagę, że istnieje ogromna różnica między pierwszą w rankingu pod względem PKB per capita Warszawą a zajmującymi kolejną pozycję Katowicami, wynosząca w 2020 r. niemal 45,5 tys. zł, co stanowiło ok. 42% PKB na mieszkańca Katowic.

Rozkład przestrzenny PKB per capita i MPZ jest podobny. Na czołowych miejscach w rankingach znajdują się przede wszystkim największe miasta, powiaty otaczające Warszawę, a także powiaty województw śląskiego, dolnośląskiego i wielkopolskiego. Jednak w przypadku MPZ wśród liderów znajdują się też powiaty będące lokalnymi ośrodkami: Rzeszów, Krosno, Biała Podlaska, Siedlce i Tarnów, w otoczeniu jednostek, w których wartość tej miary jest o wiele niższa. Należy to interpretować następująco: mniejsze ośrodki, niekoniecznie odznaczające się wysokim PKB per capita, mogą stanowić centra kulturalne, edukacyjne, charakteryzujące się dobrymi relacjami społecznymi, jakością środowiska i warunkami mieszkaniowymi.

Rozkłady przestrzenne MPZ i PKB per capita w powiatach w latach 2019–2020 były stabilne. Silna okazała się też zależność między obu zmiennymi, mimo że w niektórych powiatach obserwuje się znaczne różnice rang nadanych rosnąco obu miarom, przekraczające nawet 200 pozycji. Wśród 10 powiatów, w których przewaga rangi MPZ i rangi PKB per capita jest największa, aż 6 należy do woj. lubelskiego, 3 do woj. małopolskiego i 1 do woj. podkarpackiego. Pogłębiona analiza rozkładów przestrzennych MPZ przedstawiona w pracy Dańskiej-Borsiak (2022a) pokazała, że są to powiaty o wysokich wskaźnikach dla domen *czas wolny i relacje społeczne* oraz *bezpieczeństwo*. Z kolei powiaty o największej przewadze rangi PKB nad rangą MPZ znajdują się

przede wszystkim na terenie województw mazowieckiego (siedem powiatów), dolnośląskiego i łódzkiego. Te powiaty charakteryzuje niski wskaźnik domen *służba zdrowia* i *bezpieczeństwo*.

Wydaje się, że na podstawie przeprowadzonych analiz można sformułować odpowiedź na główne pytanie postawione w niniejszym badaniu – w jakim stopniu można ocenić poziom życia na podstawie PKB. Brzmi ona: w niewystarczającym. Ocena stopnia zaspokojenia potrzeb ludności wyłącznie na podstawie PKB per capita byłaby niepełna, ponieważ nie uwzględniałaby możliwości rekompensowania gorszych warunków materialnych lepszą sytuacją w innych obszarach życia.

Poziom życia jest zjawiskiem wielowymiarowym, trudno mierzalnym i w dużym stopniu subiektywnym. Zdaniem komisji Stiglitza-Sena-Fitoussiego, jego pomiar powinien uwzględniać osiem domen, a do opisu każdej z nich można dobierać różne wskaźniki cząstkowe (zmienne diagnostyczne), co będzie skutkowało innymi wartościami wynikowymi; możliwe jest także zastosowanie różnych metod konstrukcji miary agregatowej. Poza tym trzeba brać pod uwagę problemy związane z dostępnością zmiennych diagnostycznych do opisania kolejnych domen. W związku z tym PKB per capita może być atrakcyjnym miernikiem, ponieważ pozwala na wstępne wskazanie regionów o potencjalnie wysokim bądź niskim poziomie życia.

Warto zauważyć, że powiatów, w których wartości MPZ i PKB per capita są poniżej średniej, jest znacznie więcej niż tych, w których te miary przewyższają średnią. Jest to zjawisko niekorzystne, świadczące o istnieniu dużych obszarów wymagających wsparcia w rozwoju.

Bibliografia

Balicki, A. (2013). *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne* (wyd. 2). Wydawnictwo Uniwersytetu Gdańskiego.

Chmielewska, B. (2022). Degree and scope of the fulfillment of needs as indicators of living standard and life quality. *Zagadnienia Ekonomiki Rolnej. Problems of Agricultural Economics*, *370*(1), 107–125. https://doi.org/10.30858/zer/145776.

Ciołek, D. (2017). Oszacowanie wartości produktu krajowego brutto w polskich powiatach. *Gospodarka Narodowa*, *289*(3), 55–87. https://doi.org/10.33119/GN/100738.

Coyle, D. (2014). *GDP: A Brief but Affectionate History*. Princeton University Press. https://doi.org/10.2307/j.ctvc77mfx.

Dańska-Borsiak, B. (2022a). GDP and TFP in poviats of the Łódzkie voivodeship. Estimation and analysis of differentiation. *Econometrics. Ekonometria. Advances in Applied Data Analysis*, *26*(1), 16–30. https://doi.org/10.15611/eada.2022.1.02.

Dańska-Borsiak, B. (2022b). PKB per capita a poziom życia mieszkańców powiatów województwa łódzkiego. *Rozwój Regionalny i Polityka Regionalna*, (59), 15–28. https://doi.org/10.14746/rrpr.2022.59.03.

Dańska-Borsiak, B. (2022c). The Standard of Living and Its Dimensions in NUTS–4 Districts in Poland. An Analysis of Diversification. *Acta Universitatis Lodziensis. Folia Oeconomica*, (1), 59–79. https://doi.org/10.18778/0208-6018.358.04.

Dąbrowska, A. (2017). Jakość życia – aspekty definicyjno-badawcze. W: K. Błoński, A. Burlita, J. Witek (red.), *Pomiar jakości życia na poziomie lokalnym (na przykładzie powiatu wałeckiego)* (s. 11–18). Wydawnictwo Naukowe Uniwersytetu Szczecińskiego. http://www.wneiz.pl/katedry/market/blonski_pomiar_jakosci_zycia_poziom_lokalny.pdf.

Dykas, P., Misiak, T. (2018). Przestrzenne zróżnicowanie łącznej produktywności czynników produkcji w grupach powiatów. *Studia Prawno-Ekonomiczne*, *109*, 205–223. https://doi.org/10.26485/SPE/2018/109/13.

Genstwa, N. (2020). Środowiskowa krzywa Kuznetsa: przegląd teoretyczno-metodyczny. *Rozwój Regionalny i Polityka Regionalna*, (49), 39–50. https://doi.org/10.14746/rrpr.2020.49.04.

Główny Urząd Statystyczny. (2019). *Regionalne zróżnicowanie jakości życia w Polsce w 2018 r. Wyniki Badania spójności społecznej 2018*. https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/regionalne-zroznicowanie-jakosci-zycia-w-polsce-w-2018-roku-wyniki-badania-spojnosci-spolecznej-2018,31,1.html.

Główny Urząd Statystyczny. (2020). *Jakość życia i kapitał społeczny w Polsce. Wyniki Badania spójności społecznej 2018*. https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/jakosc-zycia-i-kapital-spoleczny-w-polsce-wyniki-badania-spojnosci-spolecznej-2018,4,3.html.

Grossman, G. M., Krueger, A. B. (1995). Economic growth and the environment. *The Quarterly Journal of Economics*, *110*(2), 353–377. https://doi.org/10.2307/2118443.

Hellwig, Z. (1968). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny*, *15*(4), 307–327.

Hwang, C. L., Yoon, K. (1981). Methods for Multiple Attribute Decision Making. W: M. Beckmann, H. P. Künzi (red.), *Multiple Attribute Decision Making. Methods and Applications A State-of-the-Art Survey* (s. 58–191). Springer-Verlag. https://doi.org/10.1007/978-3-642-48318-9_3.

Kądziołka, K. (2021). Propozycja miernika oceny poprawności metod porządkowania liniowego. *Zeszyty Naukowe Polskiego Towarzystwa Ekonomicznego w Zielonej Górze*, (15), 19–33. https://doi.org/10.26366/PTE.ZG.2021.201.

Kusterka-Jefmańska, M., Jefmański, B. (2017). Jakość życia mieszkańców powiatu wałeckiego w świetle wyników Indeksu Jakości Życia. W: K. Błoński, A. Burlita, J. Witek (red.), *Pomiar jakości życia na poziomie lokalnym (na przykładzie powiatu wałeckiego)* (s. 83–98). Wydawnictwo Naukowe Uniwersytetu Szczecińskiego. http://www.wneiz.pl/katedry/market/blonski_pomiar_jakosci_zycia_poziom_lokalny.pdf.

Majecka, A., Nowak, P. (2019). Uwarunkowania jakości życia w polskich województwach. *Nierówności Społeczne a Wzrost Gospodarczy, 59*(3), 149–161. https://doi.org/10.15584/nsawg.2019.3.10.

Panek, T. (2016). *Jakość życia – od koncepcji do pomiaru*. Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie.

Panek, T., Zwierzchowski, J. (2019). *Analiza porównawcza jakości życia w Polsce w układzie wojewódzkim w ramach podejścia możliwości* (ISiD Working Papers, nr 51). https://cor.sgh.waw.pl/bitstream/haddle/20.500.12182/1285/ISID%20_WP_51_2019.pdf?sequence=2&isAllowed=y.

Polak, E. (2016). Jakość życia w polskich województwach – analiza porównawcza wybranych regionów. *Nierówności Społeczne a Wzrost Gospodarczy*, (48), 66–89. https://doi.org/10.15584/nsawg.2016.4.6.

Samuelson, P. A., Nordhaus, W. D. (2010). *Economics* (wyd. 19). McGraw-Hill/Irwin.

Słaby, T. (2017). Przegląd koncepcji pomiaru jakości życia ludności. W: K. Błoński, A. Burlita, J. Witek (red.), *Pomiar jakości życia na poziomie lokalnym (na przykładzie powiatu wałeckiego)* (s. 19–30). Wydawnictwo Naukowe Uniwersytetu Szczecińskiego. http://www.wneiz.pl/katedry/market/blonski_pomiar_jakosci_zycia_poziom_lokalny.pdf.

Sobolewski, M., Migała-Warchoł, A., Mentel, G. (2014). Ranking poziomu życia w powiatach w latach 2003–2012 z uwzględnieniem korelacji przestrzennych. *Acta Universitatis Lodziensis. Folia Oeconomica*, *6*(308), 159–172. https://czasopisma.uni.lodz.pl/foe/article/view/272/255.

Stiglitz, J. E., Fitoussi, J. P., Durand, M. (2019). *Poza PKB. Mierzmy to, co ma znaczenie dla rozwoju społeczno-gospodarczego* (tłum. Z. Matkowski). OECD Publishing, Polskie Towarzystwo Ekonomiczne. https://doi.org/10.1787/e2b4d21b-pl.

Stiglitz, J. E., Sen, A., Fitoussi, J. P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress*. https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf.

Szaban, D. (2019). Poziom życia w powiatach województwa lubuskiego. *Wiadomości Statystyczne. The Polish Statistician*, *64*(11), 25–41. https://doi.org/10.5604/01.3001.0013.7584.

Tokarski, T. (2013). Zróżnicowanie podstawowych zmiennych makroekonomicznych w powiatach. W: M. Trojak, T. Tokarski (red.), *Statystyczna analiza przestrzennego zróżnicowania rozwoju ekonomicznego i społecznego Polski* (s. 91–138). Wydawnictwo Uniwersytetu Jagiellońskiego.

Wolak, J. (2020). Analiza zmian poziomu życia w powiatach województwa małopolskiego. *Metody Ilościowe w Badaniach Ekonomicznych*, *21*(1), 30–38. https://doi.org/10.22630/MIBE.2020.21.1.4.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(7), 22–41 https://doi.org/10.59139/ws.2024.07.2
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 16.12.2022, poprawiony/revised: 6.12.2023, zaakceptowany/accepted: 18.04.2024

The impact of sociodemographic factors on unemployment of young people in Poland[1]

Magdalena Kawecka[a]

Abstract. The labour market is an area of a dynamic structure, sensitive to both external and internal factors. The aim of the research discussed in the article is to identify the sociodemographic factors that affect the level of unemployment of young people and to deepen our knowledge about their situation in the labour market in Poland. The study is based on a sample of 387 respondents aged 18–30 and uses data provided by the 2019 Human Capital Study, which was conducted by the Polish Agency for Enterprise Development in collaboration with the Jagiellonian University in Kraków. The objective of the survey was to monitor skills demand in the labour market. A logistic model was used to analyse the impact of selected variables on the employment of young people. The findings demonstrate that the risk of unemployment highly depends on gender and level of education. The study shows that women are substantially more likely to experience unemployment than men, while a higher level of education increases the chances of employment. Additionally, owning a driver's licence, which improves mobility, allows easier access to the labour market.

Keywords: labour market, young people, unemployment, logistic regression model, logit model
JEL: C35, E24, J21, J24, J64

Wpływ czynników społeczno-demograficznych na bezrobocie młodych ludzi w Polsce

Streszczenie. Rynek pracy to obszar o dynamicznej strukturze, który wykazuje wrażliwość na czynniki zarówno zewnętrzne, jak i wewnętrzne. Celem badania omawianego w artykule jest identyfikacja czynników społeczno-demograficznych wpływających na poziom bezrobocia młodych ludzi, a dzięki temu – pogłębienie wiedzy na temat sytuacji młodych ludzi na rynku pracy w Polsce. Badanie zostało przeprowadzone na próbie 387 respondentów w wieku od 18 do 30 lat. Wykorzystano dane z Bilansu Kapitału Ludzkiego za 2019 r. realizowanego przez Polską Agencję Rozwoju Przedsiębiorczości we współpracy z Uniwer-sytetem Jagiellońskim w Krakowie, którego celem jest monitorowanie zapotrzebowania na kompetencje na rynku pracy. Do analizy wpływu wybranych zmiennych na zatrudnienie wśród młodych osób wykorzystano model logistyczny. Z uzyskanych rezultatów wynika, że na ryzyko bezrobocia istotny wpływ mają płeć i poziom wykształcenia: kobiety są znacznie bardziej zagrożone bezrobociem niż mężczyźni, a wyższy poziom wykształcenia zwiększa szanse na zatrudnienie. Dodatkowo posiadanie prawa jazdy, poprawiające mobilność, ułatwia dostęp do rynku pracy.

Słowa kluczowe: rynek pracy, młodzi ludzie, bezrobocie, model regresji logistycznej, model logitowy

---

[1] Artykuł został opracowany na podstawie referatu wygłoszonego na 40th International Conference MSA'2022 joined with MASEP, która odbyła się w dniach 7–9 listopada 2022 r. w Łodzi / The article is based on a paper delivered at the 40th International Conference MSA'2022 joined with MASEP, held on 7–9 November 2022 in Łódź, Poland.

[a] Uniwersytet Ekonomiczny w Katowicach, Wydział Informatyki i Komunikacji, Polska / University of Economics in Katowice, Faculty of Informatics and Communication, Poland. ORCID: https://orcid.org/0000-0001-7671-8340. E-mail: magdalena.kawecka@edu.uekat.pl.

© Magdalena Kawecka
Artykuł udostępniony na licencji CC BY-SA 4.0 / Article available under the CC BY-SA 4.0 licence

1. Introduction

This article focuses on the unemployment of young people (aged 18–30) and the factors that significantly affect its level. This issue is very important from the scientific point of view, but it also provides an opportunity to gain comprehensive knowledge about a particular age group, which, in turn, might impact the employment policy. Finding optimal solutions that would enable young people to gain work experience and enhance their skills and qualifications would likely lead to an increase in the employment within this age group.

The reason behind choosing such a research problem was a noticeable gap in studies focusing on young people in the face of the dynamic structure of the labour market. Theoretical discussions and empirical research are essential to understand the market structure and to identify the factors influencing young people's employment situation. This study will help clarify how such factors as education, work experience and access to training affect their job opportunities. Focusing on these issues will also reveal the barriers young people face in accessing the labour market, which is crucial for developing effective policies to improve their employment prospects. This research will lead to a better understanding of the most effective strategies for integrating young people into the labour market, and will show how the above-mentioned factors may improve the overall market conditions.

The purpose of this study is to gain in-depth knowledge about the situation of young people on the Polish labour market, particularly by identifying the sociodemographic factors that influence unemployment levels within this group. To this end, the following research question was formulated: does education in the 18–30 age group, irrespective of its form, have any impact on the professional activity of young people? The research hypothesis, based on this research question, reads: in the 18–30 age group, the participation in education, regardless of its form, affects the unemployment levels.

2. Literature review

The apparent trend over the past two decades has been a noticeable decline in young people's participation in the labour force. One of the main reasons for this has been their increased engagement in formal, non-formal and informal education.[2] In many

[2] Formal education: 'provided by universities and other entities of the higher education system, within the framework of programs that lead to full qualifications, post-graduate qualifications or qualifications in professions.' Non-formal education: 'education and training provided as part of programmes that do not lead to full qualifications or qualifications referred to above. Non-formal education is learning organised institutionally, within the framework of programmes that are not part of formal education. Learning outcomes acquired through non-formal education can be validated, accumulated and transferred in the course of acquiring 'school' and 'academic' qualifications.' Informal education: 'the acquisition of learning

countries, the outcome of this trend has been positive, i.e. the labour force has become highly-qualified. However, this development also resulted in increased requirements for the employment, e.g., higher salaries demanded by employees, and more duties assigned by employers along with longer working hours. Young people aged 18–30 are more likely to be unemployed than older adults, which is partly due to their limited work experience; however there also seem to be structural barriers preventing their free entry into the labour market.

The mismatch between the education system and the needs of the labour market is a challenge faced by many countries worldwide. This is a problem that affects both developed and developing countries. Its numerous aspects can influence the effectiveness and sustainability of the education system, with changing labour market needs being one of the most prominent factors. In the era of globalisation and technological advancement, new technologies and economic sectors are emerging, and the labour market is evolving rapidly; as a consequence, some traditional professions lose their significance (Antoninis, 2020; Stasiak, 2011; Ziomek & Szubert, 2020). Education systems often struggle to keep up with these changes, and there is a mismatch between the curriculum and the current trends in the job market. This often results in a skills gap between graduates and employer expectations.

Stasiak (2022) rightly observes that education is one of the strongest determinants defining the chances of finding and retaining employment. The author describes phenomena in some aspects opposite to the above-mentioned developments, focusing on educational reforms in Poland that started in 1999. These reforms have led to significant changes in the labour market, influenced by both demographic and educational factors, and have been extensively analysed in the context of the human capital theory. In the current phase of economic development, a typical jobless economic growth is noticeable. It is linked to the growing labour productivity resulting from the fact that the technical equipment is becoming increasingly advanced. This process contributes to growing competition in the labour market and is responsible for the fact that in developed countries, secondary or even higher education has become a minimum requirement for an individual to enter the job market (Jedlińska, 2021).

Research by Żyra (2022) shows that in the long-term, education in engineering and technology benefits regional growth and reduces unemployment. However, short-term effects for both fields are negative, indicating insufficient quality of investments in human capital. The author also points out a significant decrease in the interest in these fields over the last decade. This decline contrasts sharply with the 2000s, when

outcomes through various activities outside formal education and non-formal education, e.g. self-directed learning (e.g. foreign language, computer skills).' (Stęchły, 2021).

both Poland and other countries experienced a marked increase in the number of students pursuing engineering and technology degrees. This shift may be attributed to various socio-economic factors, including changes in industry demands, economic incentives, and evolving career preferences among young people.

In his work, Mikiewicz (2021) addresses the issue of educationalisation, which signifies the increasing importance of formal-education institutions in areas previously unrelated to them. The author asserts that these processes are common in developed countries and result from the economic and technological development (including the era of Industry 4.0) and cultural transformations that require people to have a wide range of competencies and to apply sophisticated knowledge not only to professional situations. The author also mentions 'lifelong learning', a slogan which is gaining increasingly more importance. Education becomes a tool, a task, and an outcome defining the logic of contemporary individuals' actions. Constant skills development is crucial for functioning in a modern society.

In the study on adult education conducted by Statistics Poland in 2022 and in the analysis of human capital in Poland in the years 2018–2022, the focus was on the learning processes of adults. The findings indicated that 54% of individuals aged 18–69 actively pursued knowledge through various forms of education. The study also analysed the participation rates of young adults in formal and non-formal education. In the age group of 18–24, 21.2% of respondents participated in non-formal, and 65.6% in formal education. For the age group of 25–29, these rates were 27.2% and 10.8%, respectively. These data highlight differences in the engagement in formal and non-formal education based on age group, which could have implications for educational policies and labour market dynamics (Główny Urząd Statystyczny [GUS], 2023, 2024).

Kotliński (2017) astutely observes that young individuals find themselves in a particularly unfavourable situation during economic recessions, as their employment is more susceptible to business cycle fluctuations than that of other age groups. This is due, among other factors, to the fact that their jobs are predominantly concentrated in industries sensitive to cyclical events and that they engage more frequently in part-time or temporary employment (Wysocka, 2021). The significance of knowledge and skills has been emphasised for decades in the field of economics, where growth theories highlight the role of human capital (Wierzbicka, 2021). In the theory of endogenous growth, it is especially the scientific and technical knowledge that serves as a source of long-term economic growth (Lucas, 1988). Human capital is not only associated with formal education and skills (Cabrita et al., 2015), but also with informal and para-formal education, as well as practical skills, which ultimately influence the increasing role of the knowledge-based economy (Wierzbicka, 2021).

Meanwhile, educational policy is no longer associated only with formal education at the secondary or higher education levels. It begins to encompass aspects and stages of life previously linked to social or economic policy. The key medium of this expansion is the concept of lifelong learning, empowering educational policy officials to venture into new areas previously reserved for specialists from other fields. In Poland, these changes are being implemented through systemic solutions, such as the Integrated Skills Strategy 2030, with the Ministry of Education playing a pivotal role in creating development and skills utilisation strategies (Kula & Pękowska, 2022; Ministerstwo Edukacji Narodowej, 2020).

The European Union considers smart development as one of its most important goals. Consequently, Poland is implementing a systemic solution aimed at raising the level of human capital by describing, organising and collecting various qualifications in a single registry called the Integrated Qualifications System. Based on a detailed analysis of general recommendations, the development of the Polish Qualifications Framework was initiated, which covers the entirety of educational activities. These projects integrate both the spheres of formal and non-formal education, as well as aspects related to the informal learning process, ensuring a holistic approach to qualification development in a national context (Morańska & Ostrowska, 2021; Trawińska-Konador, 2020).

The EU also emphasises the importance of skills relevant to the job market. The European Commission invests in training, aligning skills with market needs, ecological and digital transformation, and increasing mobility and recognition of professional qualifications. Flaszyńska (2023) asserts that the key element in balancing supply and demand in the labour market is skills development. The author points to the deepening problem of skills mismatching the job market, which leads to a situation where vacancies are unable to absorb the majority of the registered unemployed, despite the growing level of formal education in the society. At the European level, there is an increasing emphasis on the need for the development of lifelong learning, considered one of the main indicators within the action plan for the implementation of the European Pillar of Social Rights (to be achieved by member states by 2030; Flaszyńska, 2023).

In Poland in 2019,[3] approximately 6.5 million of the total 'potential labour force' was constituted by young people aged 15–29 (while there were about 6 million people in the group of 18–30 year-olds). This group includes people who are ready to start in

[3] The study was conducted in 2022 and was based on the most recent data available at the time. The data were extracted from the Human Balance Survey conducted by the Polish Agency for Enterprise Development jointly with the Jagiellonian University in Kraków (Polska Agencja Rozwoju Przedsiębiorczości & Uniwersytet Jagielloński w Krakowie, 2019). It is for this reason that the data relate to 2019. The literature review on the subject has been supplemented with more recent items.

a job immediately, but cannot do it as they face a barrier. Another sub-group are people who are ready for employment, but do not actively seek it for personal reasons, e.g. discouragement after previous unsuccessful attempts, disability or the fact that they already work in the 'shadow economy' (Eurostat, n.d. a, n.d. c; Krause, 2016).

An undoubtedly negative condition that affects young people is NEET (Not in Education, Employment or Training). This term refers to young people who are unemployed and at the same time not in education or vocational training. In 2019, among people aged 18–24, the NEET rate totalled 11.1% (0.6 p.p. down from the previous year, and 5.3 p.p. down from 2013), among those aged 25–30 it was 17.7% (up by 0.5 p.p. compared to the previous year, down by 5 p.p. compared to 2013), and in the 30–34 age group it reached 16.5% (0.8 p.p. down from 2018 and 4 p.p. down from 2013).

What could also be observed was that young women in most cases were as much as twice or three times more likely to have a NEET status than their male counterparts. In 2019, in the 18–24 age group, the NEET rate among females equalled 13.1% (9.3% among men), in the group of women aged 25–29 it was 27.1% (8.7% among men), and among 30–34-year-old females it reached 26% (7.4% among men). These results are not surprising, bearing in mind that women continue to experience obstacles to employment (despite the EU's gender equality policy), and the fact that they still have to often play traditional social roles (e.g. mothers, carers; Eurostat, n.d. d; Saczyńska-Sokół & Łojko, 2016; Szcześniak & Rondón, 2011). Any form of stalling a young person's professional development and career in the early stages can significantly worsen the employment and earning prospects of this person in the future.

Gender inequality is still evident in the labour market. Despite the increase in the number of women employed on the Polish market, economic discrimination of this sex can still be observed. The working-age employment rate among women is lower than that of men (70% and 79% respectively), and the unemployment rate for women in the same age group is higher than that of men (4% and 3% respectively). Women are more likely than men to withdraw completely from the labour market, mostly due to caregiving responsibilities. Among women aged 18–30, 23% are economically inactive – fewer than in 2010 (28%), but still more than men (11%). The reason for women's professional inactivity in 64% of cases is caring for children, seriously ill adults or other personal or family obligations (among professionally inactive men, it is only 16% of cases). In Poland, the percentage of economically inactive women due to caring responsibilities is one of the highest in the EU, and shows an upward trend. On average, working women are paid 9% less than men (GUS, 2020). Women are better educated than men. In 2019, 25.8% of women had higher education, which was 11.2 p.p. more than men. Women are also less likely than men to drop out of school early in their education. Educational inequalities in Poland based on gender remain

high. The issue of premature school dropout affects both genders. In 2019, the dropout rate for women aged 18–24 was 3.6%, while for men from the same age group it was 6.7% (Eurostat, n.d. b).

Low levels of social activity is mentioned as one of the factors that have a significant impact on the increase in unemployment (Centrum Badania Opinii Społecznej, 2018). In the view of this, it can be inferred that the concept of the social economy is still not properly promoted (Reichel et al., 2021). The authors also highlight that significant challenges include effectively engaging young people by emphasising the importance of social economy concepts and developing comprehensive, widely accessible educational forms and tools. Simola (2018) shows that contemporary employment of young people in Europe is characterised by precarity, regardless of the young people's education level. This age group are most often offered jobs not matching their qualifications and competencies, which reinforces the negative effects associated with job insecurity (Ross, 2008). One factor, Grabowska (2019) notes, is that 'the impact of migration on social competence may not be apparent immediately. It can be apparent after some time, but it can also blur in the course of other experiences, which is why it is so difficult to determine the direct impact of migration on a person's life and human capital.'

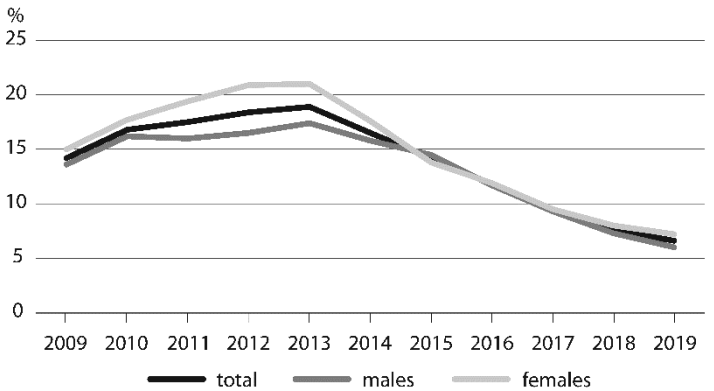
An important factor integrating and dynamising the lives of today's young people is personal development. This is dependent on paradigm shifts in the functioning of education and the perception of higher education which is considered the only opportunity for prosperity, despite, as Sukarieh and Tannock (2015, pp. 120) argue, a degree no longer guarantees a high standard of living. Also, young people spending increasingly long periods in formal, non-formal and informal education, and the need for constant further training and supplementation of competencies mean that the transition from education to employment (transition from school to work) is no longer linear and irreversible. This phenomenon is reinforced by the instability in the labour market and the lack of continuity of employment (Messyasz, 2021).

In 2019, the employment rate for young people (in the age group of 15–29) equalled 50.8%, which marked an increase by 0.6 p.p. year to year (Eurostat, n.d. e). The unemployment rate was 6.6% (7.2% for females and 6% for males), as demonstrated in the Figure. Over the 2009–2019 decade, in the unemployment rate decreased by as much as 7.6 p.p.[4] Despite this positive trend, the unemployment rate was still relatively high among young people. This was the emergence of several barriers in the labour market, which also negatively affected their health mentally and physically, and especially in

[4] In 2020, the unemployment rate among young people rose to 7.1% (up by 0.5 p.p. year to year), while 2021 saw an increase to 7.2% year to year. This growth was caused by the outbreak of the COVID-19 pandemic, which began in Poland in March 2020 (GUS, 2020). On 24th February 2022, Russia attacked Ukraine. The pandemic and the outbreak of the war had an impact on the relations in the Polish labour market.

terms of social integration because they suffers from material deprivation and worse, poverty (Trzpiot & Kawecka, 2021a, 2021b).

Figure. The unemployment rate of young people aged 15–29 in Poland



Source: author's work based on data from Eurostat (n.d. e).

3. Research method

The study uses micro-data from the Human Capital Study survey conducted in 2019 by the Polish Agency for Enterprise Development jointly with the Jagiellonian University in Kraków (Polska Agencja Rozwoju Przedsiębiorczości & Uniwersytet Jagielloński w Krakowie, 2019). The surveys are designed to monitor the demand for competencies in the labour market. The research features cross-sectional surveys that provide general information on the labour market situation, as well as industry-specific surveys.

Before proceeding with the analysis and model building, the collected data was evaluated for completeness and relevance of the given factors. The Human Capital Study included a total of 2,534 observations for the 18–70 age group. The participants were selected to provide a representative cross-section of young people in Poland, ensuring that the study captured a wide range of educational backgrounds and professional statuses. This comprehensive approach aimed to assess how different sociodemographic factors influenced the unemployment rate of young adults in the Polish dynamic labour market. The study focused on young people under 30.[5] It took into account the Act on Promotion of Employment and Labour Market Institutions (Pol. Ustawa z dnia 20 kwietnia 2004 r. o promocji zatrudnienia i instytucjach rynku

[5] Taking into account the merits and definitions that are used by both Statistics Poland and Eurostat, a group presenting the widest possible range of young people was selected. The data was subjected to a frequency analysis, which made it possible to select a representative group.

pracy), where particular emphasis is placed on assisting people under 30 years of age in entering the labour market and then remaining in employment). The results from 387 observations for people in the 18–30 age group qualified for the study. Taking into account the substantive rationale and statistical criteria, the study analysed variables describing young people in terms of education as well as categorising variables, such as age and gender (a total of 13 independent variables; see Table 1).

The Human Capital Study uses a demographically-comprehensive sample, 21.8% of which is constituted by young people aged 18–30. In terms of gender distribution, women represent 42.12%, and men 57.64% of this group. 7% of the participants do not know any foreign language. As regards skills and education, 78.9% of the respondents have a valid driver's licence. In terms of educational activities over the last 12 months, 39.7% participated in formal, 71.6% in non-formal, and 85.6% in informal education. As far as the level of education is considered, 14.7% lower secondary primary and incomplete primary or lower, 11.5% basic vocational, 48.4% secondary, and 25.2% tertiary education. To sum up, the sample was varied in terms of age, gender, degree of foreign language proficiency, ability to drive, participation in various forms of education, and the level of education. This group represents a broad section of the society, which can be useful in multiple research contexts.

In order to classify the status of a person labour market (employed or unemployed), the definitions used in the Labour Force Survey (LFS; GUS, 2022) were applied. We used a logit model, so for the unemployment, the dependent variable is represented by a dichotomous variable, where 1 means the respondent (a young person) is unemployed, whereas 0 denotes an employed person (according to the LFS methodology).

In the study, the rule of 10 events per variable was applied, providing a solid foundation for estimating parameters in a logistic regression model with 13 variables. The rule of 10 events per variable, applied in logistic regression analysis, refers to the minimum number of cases (observations) required for each independent variable in the model to enable reliable estimation. Accordingly, the number of observations (units) in the model should be at least ten times greater than the number of independent variables; in other words, there should be at least 10 cases (events) for each independent variable to evaluate its effect on the dependent variable. The rule does not say that each variable must take on at least 10 different values, but rather that there should be a sufficient number of observations for each independent variable to ensure that the estimations are statistically robust. Adhering to this rule is particularly important when the model includes a relatively large number of variables compared to the available observations. It helps avoid issues such as overfitting which can lead to incorrect conclusions and poor generalisability of the model with new data (Szaleniec, 2008). The 387 observations we used was even more than 10 events per variable, ensuring enough data for a robust and reliable analysis.

In the domain of statistical modelling, logistic regression is employed as a robust method for estimating the probabilities of a binary outcome based on a set of independent variables, which can be either categorical or continuous in nature. This model is particularly adept at handling situations where the dependent variable is dichotomous, thus facilitating the exploration of relationships between various predictor variables and the likelihood of a particular event occurring. Within this framework, the model under discussion incorporates both binary and continuous variables. The inclusion of a continuous variable, such as age, is a strategic choice, serving to capture the effect of incremental changes in this predictor on the log odds of the outcome. This approach allows a more granular analysis of how variations in age, treated as a continuous spectrum, systematically influence the probability of the occurrence of an event (here: becoming unemployed), as delineated in Hosmer et al. (2013).

The variables used for the study are presented in Table 1.

Table 1. Variables used in the survey

Variable	Details of the variables
Employment	0 – employed, 1 – unemployed
Age	continuous, quantitative variable
Gender	0 – men, 1 – women
Higher education	0 – does not have, 1 – has
Secondary education	0 – does not have, 1 – has
Basic and vocational education	0 – does not have, 1 – has
Participation in formal education (including but not limited to lower secondary primary and incomplete primary, upper secondary, post-secondary, tertiary) in the past 12 months	0 – no, 1 – yes
Participation in non-formal education (including adult schools, residential/internet courses and training, internships, apprenticeships) in the past 12 months	0 – no, 1 – yes
Participation in informal education (self-study) in the past 12 months	0 – no, 1 – yes
Frequency of Internet or e-mail use (for personal purposes, including further training)	1 – daily 2 – several times a week 3 – several times a month 4 – less frequently than several times a month
Formal confirmation of one's skills in the form of a certificate, diploma or licence	0 – no, 1 – yes
Valid driver's licence	0 – no, 1 – yes
Knowledge of at least one foreign language	0 – no, 1 – yes
Willing to develop skills and knowledge in the next 12 months	0 – no, 1 – yes

Source: author's work based on Polska Agencja Rozwoju Przedsiębiorczości and Uniwersytet Jagielloński w Krakowie (2019).

Considering the nature of the used variables, it is important to acknowledge certain limitations that might mean the data does not fully reflect current trends in the labour market, especially given the dynamic changes in the economic and social environment. The study's concentration on the 18–30 age group restricts the extrapolation of its findings to a wider demographic group. Moreover, the results may not comprehensively represent the entire young population in Poland. The author, aware of the limitations associated with the data set, would like to stress the necessity to acknowledge these constraints in order to interpret the results and formulate conclusions accurately. Despite some drawbacks, however, the study holds scientific relevance, as it offers insights into young people's conscious choices for their professional development and employment pursuits.

The logit model in statistics is also called a logit regression or logistic regression model. This model is used in explaining a dichotomous qualitative variable Y depending on the level of exogenous variables $X_1, X_2, \dots, X_k$ (qualitative or quantitative). The explained variable then has a zero-one form (Agresti, 2018; Borkowski et al., 2003; Hosmer et al., 2013):

$$Y = \begin{cases} 1 \text{ the incident occurred} \\ 0 \text{ the incident did not occur} \end{cases} \tag{1}$$

The logistic model function has the following form:

$$Y_t = \frac{e^{Z_t}}{1+e^{Z_t}}, \tag{2}$$

where $Z_t = a + bt,$ a and b are the parameters, while the saturation level is 1. Variable t means that we still have a growth curve.

The logit model has the following form:

$$P(Y = 1|X_1, X_2, \dots, X_k) = \frac{e^{\alpha+\sum_{i=1}^{k}\beta_i X_i}}{1+e^{\alpha+\sum_{i=1}^{k}\beta_i X_i}}, \tag{3}$$

where $P(Y = 1|X_1, X_2, \dots, X_k)$ means the conditional probability of the explanatory (dependent) variable to reach the distinguished value (2) under the condition of specific values of the explanatory (independent) variables $X_1, X_2, \dots, X_k$. Then α is the logistic regression constant, β_i is the logistic regression coefficient for the i-th independent variable, while X_k is the k-th independent variable.

In statistics, the coefficient of determination, denoted as R-square (R^2), is commonly used to assess the fit of a linear regression model by quantifying how well it explains the variability of the dependent variable. However, in the context of logistic regression, applying traditional R^2 measures is less straightforward due to the binary nature of the

dependent variable. As a result, adaptations such as Cox and Snell's *R*-square and Nagelkerke's *R*-square are frequently employed. Cox and Snell's *R*-square is sensitive to the number of variables in the model and cannot reach the upper limit of 1, a restriction stemming from the inherent characteristics of logistic regression models. To overcome the limitations of Cox and Snell's *R*-square, Nagelkerke proposed a modified version that scales the *R*-square to the maximum of 1, making it more interpretable and useful in logistic regression analysis.

Additionally, the Hosmer and Lemeshow test is an important measure used in logistic regression models to assess the model's goodness-of-fit. This test evaluates the fit of the model by comparing the observed probabilities with the predicted probabilities. A non-significant result is desired, as it indicates that there is no significant difference between the observed and predicted values, and so that the model fits the data well (Agresti, 2018; Borkowski et al., 2003; Hosmer et al., 2013).

This method makes it possible to determine the strength, but also the direction of the relationship between a qualitative factor (class type) or quantitative factor (discrete or continuous type) and a dichotomous explanatory variable. The model has to meet the following criteria: the randomness of the sample, numerous observations (the number of observations should be greater than $10 \text{ x } (k + 1)$, where k is the number of independent variables; Stanisz, 2000), no collinearity of explanatory variables, and the independence of observations.

On the basis of the odds ratio, inference is carried out, which answers the question of the effect of independent variables on the dependent variable. This indicator expresses how many times the probability of the studied event increases or decreases if there is a change in the independent variable (at established values of independent variables). Thus, it determines the change in the probability of the occurrence of a specific value of one of the dependent variable as the independent variable increases by one unit, assuming that the other independent variables remain constant (Hosmer et al., 2013; O'Loughlin et al., 2012).

4. Results

The labour market is one of the most popular areas of scientific research and the object of politicians' interest. It is also one of the main elements of the economy. As observing the indicators and trends in the labour market is very important, researchers and other interested actors monitor them on a daily basis. The labour market is affected by many domestic economic factors as well as by global ones.

The author conducted a reliability quotient test to verify the estimation of the logistic model. It should be noted that in the case of large numbers, the test statistics

has a chi-square distribution (χ^2) with the number of degrees of freedom equal to (corresponding to) the number of variables in the estimated model (Table 2). Based on the results, it should be inferred that all variables in the model are statistically significant.

Table 2. Test results and the value of the log-likelihood and pseudo-R^2 for the model

Test	χ^2	Degrees of freedom	Significance
Chi-square	78,958	13	<0.01
Hosmer and Lemeshow	13,299	8	0.102

Note. Significance at *p*-value = 0.05. Model summary: –2 logarithm of credibility = 396,446, R^2 Cox and Snell = 0.514, R^2 Nagelkerke = 0.243.
Source: author's work based on Polska Agencja Rozwoju Przedsiębiorczości and Uniwersytet Jagielloński w Krakowie (2019).

In the model, Nagelkerke's *R*-square is 0.243, indicating that the model explains approximately 24.3% of the variance in the group membership of the dependent variable. The Hosmer and Lemeshow test yields a statistic of 0.102, indicating no significant difference between the observed and predicted values, which demonstrates a good fit of the model to the data.

Table 3 juxtaposes the observed values with the values predicted by the estimated model. The results demonstrate that the logistic function correctly predicts belonging to one of the two groups for about 81.5% of young people.

Table 3. Classification table for the Logistic Regression Model of unemployment

Specification	Observed	Predicted in the model	Percentage of correct classifications
Employed	375	361	96,2
Unemployed	96	23	23,8
Total	471	384	81,5

Note. The division point is 500.
Source: author's work based on Polska Agencja Rozwoju Przedsiębiorczości and Uniwersytet Jagielloński w Krakowie (2019).

From Table 4, which presents a logistic model for selected socioeconomic and demographic variables, we can see that only four variables are statistically significant, assuming that the *p*-value is smaller than or equal to 0.05. These variables have been highlighted in bold in the table, emphasising their importance in the analysed model.

Table 4. Estimated values of the parameters of the logistic regression model

Specification	β	Std. error	Wald test	Significance	Exp(β)	95% CI for exp(β)	
						lower	upper
Age	–0.071	0.054	1.740	0.187	0.931	0.838	1.035
Gender	**0.755**	**0.272**	**7.700**	**0.006**	**2.127**	**1.248**	**3.624**
Tertiary education	–0.761	0.482	2.491	0.115	0.467	0.182	1.202
Secondary education	**–0.979**	**0.355**	**7.582**	**0.006**	**0.376**	**0.187**	**0.754**
Basic and vocational education	**–2.526**	**0.774**	**10.651**	**0.001**	**0.080**	**0.018**	**0.365**
Participation in formal education (including but not limited to lower secondary primary and incomplete primary, upper secondary, post-secondary, tertiary) in the past 12 months	0.617	0.334	3.416	0.065	1.853	0.963	3.565
Participation in non-formal education (including adult schools, residential/internet courses and training, internships, apprenticeships) in the past 12 months	–0.255	0.362	0.494	0.482	0.775	0.381	1.577
Participation in informal education (self-study) in the past 12 months	0.136	0.515	0.069	0.792	1.145	0.417	3.145
Frequency of Internet or e-mail use (for personal purposes, including further training)	0.092	0.364	0.064	0.801	1.096	0.537	2.239
Formal confirmation of one's skills by obtaining a certificate, diploma or a licence	0.590	0.376	2.460	0.117	1.804	0.863	3.773
Valid driver's licence	**–0.740**	**0.317**	**5.450**	**0.020**	**0.477**	**0.256**	**0.888**
Knowledge of at least one foreign language	–0.276	0.608	0.206	0.650	0.759	0.230	2.500
Willing to develop skills and knowledge in the next 12 months	0.126	0.160	0.617	0.432	1.134	0.829	1.551
Constant	0.708	1.730	0.168	0.682	2.030	.	.

Note. Degrees of freedom = 1. Statistically significant variables are marked in bold (significance ≤ 0.05).
Source: author's work based on Polska Agencja Rozwoju Przedsiębiorczości and Uniwersytet Jagielloński w Krakowie (2019).

The first significant variable in the model is the gender of the respondents, coded as 0 for male and 1 for female. The obtained result (significance = 0.006) indicates that this variable is statistically significant. More specifically, the odds ratio (exp(β) = 2.127) with a 95% confidence interval ranging from 1.248 to 3.624 shows that women

face more than double the odds of being unemployed compared to the situation of men. This result highlights a significant gender disparity in the likelihood of the unemployment within the studied population.

The second significant variable identified in the model is secondary education, compared to the reference category of lower secondary primary and incomplete primary or lower. The findings indicate considerable differences in the unemployment likelihood depending on the education level. For people with secondary education, the likelihood of being unemployed is 62.4% lower ($\exp(\beta) = 0.376$, with a 95% confidence interval of 0.187 to 0.754) than for those with lower secondary primary and incomplete primary or lower education. This demonstrates that secondary education significantly reduces the risk of unemployment. Also, individuals with vocational and basic education are up to 92% less likely to be unemployed ($\exp(\beta) = 0.080$, with a 95% confidence interval of 0.018 to 0.365) than the reference group. This shows that having vocational and basic education reduces the unemployment risk even more pronouncedly, proving that the qualifications earned through this type of education are highly valued in the labour market and equip individuals with more marketable skills.

The last statistically significant factor in the model is the respondent's ability to drive. The analysis shows that people holding a driver's licence are 52.3% less likely to be unemployed ($\exp(\beta) = 0.477$, with a 95% confidence interval of 0.256 to 0.888) than those who do not. The influence of this variable is considerable and aligns with contemporary expectations and demands in the labour market. The ability to drive is increasingly seen as an asset, particularly for young people in the age of mobility. In many cases, it has become a necessary or even mandatory requirement for many types of jobs, especially in places where public transportation is poor or in jobs that require travelling or transportation of goods. Therefore, being able to drive broadens the employment possibilities and enhances job prospects.

5. Conclusions

The study focuses on a group of young people who are either entering the labour market or are already present there. The objective of the article was to examine the influence of socioeconomic and demographic factors on the unemployment among young people in Poland using a logistic model. This study successfully meets this objective, concentrating on the analysis of how different education levels impact the unemployment risk for individuals aged 18–30. The research shows a statistically significant correlation between the education level and the employment status of young people: the higher the education level, the lower the risk of unemployment.

The study primarily dealt with education-related factors affecting the unemployment of young people. The results indicate that a young person's gender plays a significant role in this context, with women being more likely to be unemployed – a trend reflecting the persisting disparities in wages and employment conditions between two genders. Higher level of education correlates strongly with the reduced unemployment risk among young people, which emphasises the importance of academic achievement in enhancing one's job prospects. Another important skill boosting the employment opportunities is the ability to drive, which highlights the significance of mobility in the current job market. Lack of a driver's licence might be a substantial obstacle for young people in finding employment.

In response to the research question, 'does education in the 18–30 age group, irrespective of its form, have any impact on the professional activity of young people?', the analysis of data from the 2019 Human Capital Study demonstrates that the level of education significantly affects the professional activity of the 18–30 age group. The findings show that as young people's education level increases, their chances for employment grow. More specifically, having tertiary, secondary, basic or vocational education reduces the unemployment risk among young people. Variables connected to secondary, basic and vocational education were statistically significant.

Data from the 2019 Human Capital Study reveals that gender inequality has persisted for decades despite the efforts to promote equality. The 2020 'Poland on the Path of Sustainable Development' Statistics Poland report indicates that one of the main goals for 2030 is to achieve gender equality and empower women. In Poland, despite higher educational attainment, women continue to face significant labour market disparities – their employment rate is lower than that of men (70% and 79%, respectively), and their unemployment rate is higher (4% and 3%, respectively). Additionally, 23% of women aged 25-59 are economically inactive, primarily due to caregiving responsibilities, which is one of the highest rates in the EU. Women earn on average 9% less than men, with the largest disparities in the financial, insurance, and ICT sectors. Despite an increase in women's political representation due to legal regulations, their participation remains below the EU average.

International investments in education significantly influence the global development by fostering economic growth, reducing poverty, and promoting socio-political stability. Education, as a catalyst for development, is essential to encouraging innovation and economic competitiveness, and it plays a pivotal role in sustainable development. This study could serve as an encouragement for further research into the impact of socioeconomic and demographic factors on the unemployment of young people, which is a global issue. The author would like to broaden the scope of her future research to encompass a global perspective in order to receive a more comprehensive analysis of how international trends and worldwide socioeconomic

and demographic factors influence the unemployment rates among young people. Such a comprehensive analysis will require adopting several research methodologies that would address the complexity of the issue.

In conclusion, it is extremely important and relevant to observe a country's economy, but also to study it in order to be able to identify the determinants of the increase in the unemployment. Young people are a much varied group, as they might participate in several forms of education and at the same time be employed, sometimes on less favourable terms (e.g., civil law contracts, part-time contracts). They might also belong to the NEET category. There are several possibilities of the employment status of people in this age group. However, it should be remembered that these are the people who predominantly are just entering the labour market. It is important, therefore, to expand the employment policy to cover this group and offer other forms of support to them that will result in their increased employment.

References

Agresti, A. (2018). *An Introduction to Categorical Data Analysis* (3rd edition). Wiley.

Antoninis, M. (2020). *Global Education Monitoring Report Summary. Inclusion and education: all means all*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000373721/PDF/373721eng.pdf.multi.

Borkowski, B., Dudek, H., & Szczesny, W. (2003). *Ekonometria. Wybrane zagadnienia*. Wydawnictwo Naukowe PWN.

Cabrita, M. R., Cabrita, C., Matos, F., & del Pilar Muñoz Dueñas, M. (2015). Entrepreneurship Capital and Regional Development: A Perspective Based on Intellectual Capital. In R. Baptista, & J. Leitão (Eds.), *Entrepre, Human Capital, and Regional Development* (pp. 15–28). Springer. https://doi.org/10.1007/978-3-319-12871-9_2.

Centrum Badania Opinii Społecznej. (2018). *Aktywność Polaków w organizacjach obywatelskich. Komunikat z badań nr 29/2018*. https://www.cbos.pl/SPISKOM.POL/2018/K_029_18.PDF.

Eurostat. (n.d. a). *Child and youth population on 1 January by sex and age* [Data set]. https://doi.org/10.2908/YTH_DEMO_010.

Eurostat. (n.d. b). *Children and youth*. https://ec.europa.eu/eurostat/web/children-youth.

Eurostat. (n.d. c). *Population on 1 January by age and sex* [Data set]. https://doi.org/10.2908/DEMO_PJAN.

Eurostat. (n.d. d). *Young people neither in employment nor in education and training by sex, age and degree of urbanisation (NEET rates)* [Data set]. https://doi.org/10.2908/EDAT_LFSE_29.

Eurostat. (n.d. e). *Youth employment rate by sex, age and country of birth* [Data set]. https://doi.org/10.2908/YTH_EMPL_020.

Flaszyńska, E. (2023). Umiejętności na zmieniającym się rynku pracy XXI wieku. *Rynek Pracy*, *185*(2), 71–89.

Główny Urząd Statystyczny. (2020). Cel 5. Osiągnąć równość płci oraz wzmocnić pozycję kobiet i dziewcząt. In *Raport 2020. Polska na drodze zrównoważonego rozwoju*. https://raportsdg.stat.gov.pl/2020/cel5.html.

Główny Urząd Statystyczny. (2022). *Badanie Aktywności Ekonomicznej Ludności (BAEL)*. https://stat.gov.pl/badania-statystyczne/badania-ankietowe/badania-spoleczne/badanie-aktywnosci-ekonomicznej-ludnosci-bael/.

Główny Urząd Statystyczny. (2023). *Kapitał ludzki w Polsce w latach 2018–2022*. https://stat.gov.pl/obszary-tematyczne/inne-opracowania/inne-opracowania-zbiorcze/kapital-ludzki-w-polsce-w-latach-2018-2022,8,9.html.

Główny Urząd Statystyczny. (2024). *Uczenie się osób dorosłych w 2022 r*. https://stat.gov.pl/obszary-tematyczne/edukacja/edukacja/uczenie-sie-osob-doroslych-w-2022-roku,22,1.html.

Grabowska, I. (2019). *Otwierając głowy. Migracje i kompetencje społeczne*. Wydawnictwo Naukowe Scholar.

Hosmer, D. W. Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd edition). John Wiley & Sons. https://doi.org/10.1002/9781118548387.ch1.

Jedlińska, R. (2021). Edukacja i rynek pracy w dobie globalizacji. *Studia Pedagogiczne. Problemy społeczne, edukacyjne i artystyczne*, *38*, 33–54. https://doi.org/10.25951/4664.

Kotliński, K. (2017). Zróżnicowanie rynku pracy młodzieży w państwach członkowskich Unii Europejskiej. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (465), 79–86. https://doi.org/10.15611/pn.2017.465.08.

Krause, E. (2016). Zjawisko NEET, czyli o młodzieży trzy razy nic. *Problemy Profesjologii*, (2), 67–81.

Kula, E., & Pękowska, M. (2022). Kształcenie całożyciowe w perspektywie Europejskiego programu na rzecz umiejętności. *Edukacja Dorosłych*, *87*(2), 45–60. https://doi.org/10.12775/ED.2022.013.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*(1), 3–42.

Messyasz, K. (2021). Pokolenie Z na rynku pracy – strukturalne uwarunkowania i oczekiwania. *Acta Universitatis Lodziensis. Folia Sociologica*, (76), 97–114. https://doi.org/10.18778/0208-600X.76.06.

Mikiewicz, P. (2021). Edukacjonalizacja i jej znaczenie we współczesnym świecie. *Youth in Central and Eastern Europe*, *8*(12), 43–51. https://doi.org/10.24917/ycee.2021.12.43-51.

Ministerstwo Edukacji Narodowej. (2020). *Zintegrowana Strategia Umiejętności 2030 (część szczegółowa). Polityka na rzecz rozwijania umiejętności zgodnie z ideą uczenia się przez całe życie*. https://www.gov.pl/web/edukacja-i-nauka/zintegrowana-strategia-umiejetnosci-2030-czesc-szczegolowa-dokument-przyjety-przez-rade-ministrow.

Morańska, D., & Ostrowska, B. (2021). Sektorowa mapa kwalifikacji – graficzna reprezentacja sektorowej ramy kwalifikacji na przykładzie sektora IT. *Dydaktyka Informatyki*, *16*, 58–73. https://doi.org/10.15584/di.2021.16.7.

O'Loughlin, E. K., Dugas, E. N., Sabiston, C. M., & O'Loughlin, J. L. (2012). Prevalence and correlates of exergaming in youth. *Pediatrics*, *130*(5), 806–814. https://doi.org/10.1542/peds.2012-0391.

Polska Agencja Rozwoju Przedsiębiorczości & Uniwersytet Jagielloński w Krakowie. (2019) *Baza danych z badania ludności BKL edycja 2019* [Data set]. https://www.study.gov.pl/component/site/site/bilans-kapitalu-ludzkiego#metodologiabadaniabkl.

Reichel, J., Rudnicka, A., & Socha, B. (2021). Gospodarka społeczna – nieodkryty potencjał rynku pracy. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, *83*(1), 277–292. https://doi.org/10.14746/rpeis.2021.83.1.20.

Ross, A. (2008). The New Geography of Work: Power to the Precarious?. *Theory, Culture and Society*, *25*(7–8), 31–49. https://doi.org/10.1177/0263276408097795.

Saczyńska-Sokół, S., & Łojko, M. (2016). Sytuacja młodzieży NEET na rynku pracy. *Zeszyty Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach. Seria Administracja i Zarządzanie*, *35*(108), 97–106. https://czasopisma.uph.edu.pl/znadministracja/article/view/499.

Simola, A. (2018). Lost in Administration: (Re)Producing Precarious Citizenship for Young University-Educated Intra-EU Migrants in Brussels. *Work, Employment and Society*, *32*(3) 458–474. https://doi.org/10.1177/0950017018755653.

Stanisz, A. (2000). *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny: tom 2. Modele liniowe i nieliniowe*. StatSoft Polska.

Stasiak, J. (2011). System edukacji wobec wyzwań rynku pracy w Polsce. *Acta Universitatis Lodziensis. Folia Oeconomica*, (248), 343–357.

Stasiak, J. (2022). Rynek pracy a edukacja w Polsce. In A. Stępniak-Kucharska & M. Kapela (Eds.), *Współczesne problemy gospodarcze. Zrównoważony rozwój* (pp. 97–108). Politechnika Warszawska. https://pw.plock.pl/Media/Files/Wspolczesne-problemy-gospodarcze-Zrownowazony-rozwoj.

Stęchły, W. (2021). *Edukacja formalna wobec edukacji pozaformalnej i uczenia się nieformalnego. Analiza komplementarności instytucjonalnej w kontekście zintegrowanego systemu kwalifikacji*. Instytut Badań Edukacyjnych.

Sukarieh, M., & Tannock, S. (2015). *Youth Rising? The Politics of Youth in the Global Economy*. Taylor & Francis. http://dx.doi.org/10.4324/9781315884660.

Szaleniec, M. (2008). *Sieci neuronowe i regresja wieloraka – czyli jak okiełznać złożoność w badaniach naukowych?*. StatSoft Polska. https://media.statsoft.pl/_old_dnn/downloads/sieci%20neuronowe.pdf.

Szcześniak, M., & Rondón, G. (2011). Pokolenie „ani-ani": o młodzieży, która się nie uczy, nie pracuje i nie dba o samokształcenie. *Psychologia Społeczna*, (3), 241–251.

The official website of Ukraine. (n.d.). *Russia invaded Ukraine*. Retrieved September 20, 2022, from https://war.ukraine.ua/.

Trawińska-Konador, K. (2020). Zintegrowany system kwalifikacji – narzędzie łączące edukację z rynkiem pracy. *Polityka Społeczna*, (10), 8–12. https://doi.org/10.5604/01.3001.0014.4748.

Trzpiot, G., & Kawecka, M. (2021a). Description of the labour market status of young people in selected countries of the European Union – the taxonomic approach. *Econometrics, Ekonometria. Advances in Applied Data Analysis*, *25*(4), 17–39. http://doi.org/10.15611/eada.2021.4.02.

Trzpiot, G., & Kawecka, M. (2021b). Evaluation of the labor market status of young people in selected countries of the European Union – the multiple regression approach. In G. Trzpiot (Ed.), *Modeling of complex data sets and risk analysis* (pp. 46–73). Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.

Ustawa z dnia 20 kwietnia 2004 r. o promocji zatrudnienia i instytucjach rynku pracy (Dz.U. 2022 poz. 690).

Wierzbicka, W. (2021). Differentiation in levels of human capital among small cities in Warminsko-Mazurskie Voivodship. *Entrepreneurship and Sustainability Issues*, *9*(2), 209–222. https://doi.org/10.9770/jesi.2021.9.2(14).

Wysocka, M. (2021). Effect of Demographic Changes on the Labour Supply in Lithuania and Poland. *Inzinerine Ekonomika – Engineering Economics*, *32*(1), 35–47. https://doi.org/10.5755/j01.ee.32.1.25087.

Ziomek, A., & Szubert, T. (2020). Education, Employment and Social Development as Key Measures of Digitalisation Role in Labour Force Development in EU. *Zeszyty Naukowe Politechniki Poznańskiej. Organizacja i Zarządzanie*, (82), 287–305. https://doi.org/10.21008/j.0239-9415.2020.082.17.

Żyra, J. (2022). The Regional Growth and Unemployment Effects of Engineering & Technology and Economics Education in Poland. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu / Research Papers of Wrocław University of Economics*, *66*(4), 186–198. https://doi.org/10.15611/pn.2022.4.13.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(7), 42–58 https://doi.org/10.59139/ws.2024.07.3
Spisy powszechne – problemy i wyzwania / Issues and challenges in census taking
Zgłoszony/submitted: 22.11.2023, poprawiony/revised: 26.03.2024, zaakceptowany/accepted: 23.06.2024

Udostępnianie danych spisowych w przekroju siatki kilometrowej – wyzwania związane z ochroną tajemnicy statystycznej[1]

Tomasz Klimanek[a], Tomasz Józefowski [b], Andrzej Młodak[c], Amelia Wardzińska-Sharif[d]

Streszczenie. Udostępnianie podstawowych danych pochodzących z Narodowego Spisu Powszechnego Ludności i Mieszkań (NSP) w przekroju siatki kilometrowej należy uznać za jeden z najważniejszych kierunków upowszechniania wyników spisowych, a zarazem spełnienie krajowych i międzynarodowych wymogów w tym zakresie. Z uwagi na to, że komórki siatki są relatywnie małe (kwadraty 1 km×1 km), a ryzyko identyfikacji konkretnej osoby i ujawnienia jej danych wrażliwych jest znaczne, niezbędne staje się podjęcie stosownych działań ochronnych. Celem artykułu jest omówienie najważniejszych działań z zakresu kontroli ujawniania danych na przykładzie danych zebranych podczas NSP 2021 oraz zaproponowanie możliwych do zastosowania w tym przypadku metod i narzędzi kontroli. W szczególności dotyczy to podejść niezakłóceniowych, czyli takich, które prowadzą do ukrywania danych wrażliwych. Ponadto w artykule wskazano najistotniejsze problemy i wyzwania związane z podjęciem tego rodzaju działań ochronnych w zależności od zakresu udostępnianych informacji, ponieważ liczba i rodzaj zmiennych mają kluczowe znaczenie dla ryzyka identyfikacji jednostki i doboru narzędzi ochronnych. Przedstawiono także propozycje rozwiązań metodologicznych i technicznych. Analizy wykazały, że ochrona danych stanowi ważne wyzwanie w rozpatrywanym przypadku, zwłaszcza gdy należy kontrolować ujawnianie danych z kilku powiązanych ze sobą baz danych – wówczas konieczne jest uwzględnienie logicznych i matematycznych powiązań między zbiorami. Dodatkowym czynnikiem ryzyka może się okazać gęstość lub hierarchiczność siatki.

[1] Artykuł został opracowany na podstawie referatu wygłoszonego na Konferencji Naukowej „Statystyka w służbie polityki społeczno-gospodarczej: nowe badania, metody i narzędzia", która odbyła się w dniach 23–24 marca 2023 r. w Warszawie / The article is based on a paper delivered at the Scientific Conference 'Statistics in the service of socio-economic policy: new surveys, methods and tools', held in Warsaw on 23–24 March 2023.

[a] Uniwersytet Ekonomiczny w Poznaniu, Instytut Informatyki i Ekonomii Ilościowej; Urząd Statystyczny w Poznaniu, Polska / Poznań University of Economics and Business, Institute of Informatics and Quantitative Economics; Statistical Office in Poznań, Poland. ORCID: https://orcid.org/0000-0001-6411-8145. E-mail: t.klimanek@stat.gov.pl.

[b] Uniwersytet Ekonomiczny w Poznaniu, Instytut Informatyki i Ekonomii Ilościowej; Urząd Statystyczny w Poznaniu, Polska / Poznań University of Economics and Business, Institute of Informatics and Quantitative Economics; Statistical Office in Poznań, Poland. ORCID: https://orcid.org/0000-0001-9485-1946. E-mail: t.jozefowski@stat.gov.pl.

[c] Uniwersytet Kaliski im. Prezydenta Stanisława Wojciechowskiego, Międzywydziałowy Zakład Matematyki i Statystyki; Urząd Statystyczny w Poznaniu, Polska / University of Kalisz, Inter-faculty Department of Mathematics and Statistics; Statistical Office in Poznań, Poland. ORCID: https://orcid.org/0000-0002-6853-9163. Autor korespondencyjny / Corresponding author, e-mail: a.mlodak@stat.gov.pl.

[d] Główny Urząd Statystyczny, Departament Systemów Teleinformatycznych, Geostatystyki i Spisów, Polska / Statistics Poland, ICT Systems, Geostatistics and Census Department, Poland. ORCID: https://orcid.org/0009-0001-8231-8113. E-mail: a.wardzinska@stat.gov.pl.

© Tomasz Klimanek, Tomasz Józefowski, Andrzej Młodak, Amelia Wardzińska-Sharif
Artykuł udostępniony na licencji CC BY-SA 4.0 / Article available under the CC BY-SA 4.0 licence

Słowa kluczowe: siatka kilometrowa, kontrola ujawniania danych, metody niezakłóceniowe, spis powszechny
JEL: C54, C63, C83

Data dissemination in the kilometre grid: challenges connected to the protection of statistical confidentiality

Abstract. Providing basic data from the National Population and Housing Census in a kilometre grid is one of the most important ways of disseminating census results, which at the same time meets the applicable national and international requirements. Due to the fact that the tiles of the kilometre grid are relatively small (squares with one-kilometre-long sides) and thus the risk of identifying a concrete person and disclosing sensitive information about him or her is significant, it is necessary to employ data-protection procedures. The aim of the paper is to discuss the most important directions in the statistical disclosure control on the example of data collected during the National Population and Housing Census 2021, and to propose methods and tools from the aforementioned realm that would be applicable. These will be mainly non-perturbative approaches, i.e. ones that cause suppression of sensitive information. The paper also brings to light the most important issues and challenges dependent on the scope of information disclosed and related to this type of data-protection procedures, as the number and type of variables determine the risk of the identification of individuals and influence the selection of suitable protection tools. The article sets forth proposals for methodological and technical solutions in the field. The analyses demonstrate that data protection poses a significant challenge in the studied case, especially if several mutually-connected databases are to be protected. In such a situation, it is necessary to take into account the logical and mathematical connections between the data sets. An additional risk factor can also be the density or hierarchical character of the grid.
Keywords: kilometre grid, statistical disclosure control, non-perturbative methods, census

1. Wprowadzenie

Współcześnie można zaobserwować rosnące zapotrzebowanie użytkowników informacji statystycznych na coraz bardziej szczegółowe dane. Owa szczegółowość dotyczy nie tylko większej liczby zmiennych opisujących różne aspekty określonych zjawisk, lecz także coraz niższych poziomów agregacji przestrzennej, na których prezentowane są dane. To ważne zwłaszcza w przypadku spisów powszechnych, które z uwagi na zakres uzyskiwanych danych i liczbę badanych respondentów (a w ostatnich latach także skalę wykorzystania źródeł administracyjnych) dostarczają szczególnie użytecznej wiedzy.

W związku z tym – zgodnie z rekomendacjami międzynarodowymi (United Nations, 2017) – najważniejsze dane spisowe są udostępniane w przekroju określonych siatek przestrzennych. W Polsce stosuje się na ogół siatkę kilometrową (ang. *kilometre grid*) o komórkach w rozmiarze 1 km×1 km (powierzchnia 1 km^2). W obszarze wyznaczonym przez daną komórkę nalicza się agregaty określonych zmiennych kategorialnych dla rekordów odpowiadających jednostkom zlokalizowanym w tym

obrębie. Ułatwia to użytkownikom dokonywanie efektywnych analiz zjawisk społeczno-gospodarczych, w tym zróżnicowania ich natężenia przestrzennego[2].

Przed udostępnieniem danych na tak dużym poziomie szczegółowości należy szczególnie zadbać o właściwą ochronę tajemnicy statystycznej. Zgodnie z podstawową zasadą kontroli ujawniania danych (ang. *statistical disclosure control* – SDC) taka ochrona powinna z jednej strony zapewnić minimalizację ryzyka identyfikacji jednostki, a z drugiej – minimalizację straty informacji będącej wynikiem zastosowania odpowiednich mechanizmów ochrony: niezakłóceniowych, czyli prowadzących do ukrycia niektórych danych, lub zakłóceniowych, czyli zniekształcających w przyjęty sposób pewne informacje w celu ochrony poufności (Hundepool i in., 2012). Spośród wielu tego typu narzędzi (np. Hundepool i in., 2012; Młodak i in., 2023) należy wybrać te, które w danej sytuacji okażą się najefektywniejsze.

Celem artykułu jest omówienie najważniejszych działań z zakresu kontroli ujawniania danych na przykładzie danych zebranych podczas Narodowego Spisu Powszechnego Ludności i Mieszkań (NSP) 2021 oraz zaproponowanie możliwych do zastosowania w tym przypadku metod i narzędzi kontroli.

2. Znaczenie danych w przekroju siatki kilometrowej

Rolą statystyki publicznej jest zbieranie, gromadzenie i analizowanie danych statystycznych, a także opracowywanie wyników badań statystycznych i udostępnianie ich w formie wynikowych informacji statystycznych. Z uwagi na rosnące zapotrzebowanie na dane przestrzenne, przed rundą spisów 2010–2011 Główny Urząd Statystyczny (GUS) wdrożył technologię GIS[3] (Geographical Information System), w ramach której utworzono Przestrzenną Bazę Adresową (PBA). W skład bazy wchodzą następujące warstwy:

- punkty adresowe dla budynków, w których znajduje się co najmniej jedno mieszkanie, wraz z lokalizacją (współrzędnymi);
- wektorowe granice podziału statystycznego, tj. rejonów statystycznych i obwodów spisowych.

Warstwy te są na bieżąco aktualizowane przez pracowników statystyki publicznej. Dzięki wprowadzeniu do danych statystycznych punktów adresowych zmieniono

[2] Warto zauważyć, że siatka kilometrowa jest szczególnie przydatna do przeprowadzania porównań w skali regionalnej, kontynentalnej i globalnej, np. analiz zmian klimatycznych (tereny zalewowe a gęstość zaludnienia) lub odległości od domu do szkoły lub miejsca pracy. Nie będzie natomiast użyteczna w przypadku analiz terytorialnych prowadzonych przez samorządy z uwagi na odrębność tego układu od podziału administracyjnego lub kształtu funkcjonalnych obszarów miejskich.

[3] GIS to system pozyskiwania, gromadzenia, weryfikowania, integrowania, analizowania, transferowania i udostępniania danych przestrzennych. W szerokim rozumieniu obejmuje on metody, środki techniczne (w tym sprzęt i oprogramowanie), bazę danych przestrzennych, organizację, zasoby finansowe i ludzi zainteresowanych jego funkcjonowaniem (System informacji przestrzennej, n.d.).

dotychczasowy system identyfikacji przestrzennej z przyporządkowania obszarowego (obwody spisowe) na przyporządkowanie punktowe. Miało to zasadnicze znaczenie dla zastosowania geomatyki (geoinformatyki) w statystyce publicznej. Zmiana przyporządkowania umożliwiła bardziej elastyczne grupowanie danych statystycznych dla dowolnie małych obszarów. Pozwoliło to także na utworzenie bazy mikrodanych o charakterze przestrzennym, dającej możliwość przeprowadzania analiz geostatystycznych różnych zjawisk dotyczących m.in. demografii (np. średnia odległość od domu do pracy, szkoły lub szpitala).

Odniesienie danych statystycznych do punktu na mapie pozwoliło również na uniezależnienie się w prowadzonych badaniach od zmian w podziale terytorialnym kraju, skutkujących zwykle zmianami obwodów spisowych i wynikającymi stąd pracochłonnymi przeliczeniami. Ułatwiło także analizę porównawczą szeregów czasowych niezależnie od zmian zachodzących w tym podziale (Dygaszewicz, 2012).

Dzięki użyciu technologii GIS w statystyce publicznej GUS może od wielu lat brać udział w międzynarodowych projektach dotyczących wykorzystania informacji geograficznych w statystyce inicjowanych w ramach Europejskiego Systemu Statystycznego (ESS). Podczas pracy przy tych projektach stworzono m.in. metodę przedstawiania danych demograficznych w innym podziale niż dotychczas stosowany podział administracyjny. Wypracowano sposoby prezentacji danych o rozmieszczeniu ludności w podziale ewidencyjnym, statystycznym, a także za pomocą siatki kilometrowej i przeanalizowano te rozwiązania pod kątem jak największej przydatności dla potencjalnego użytkownika. Dzięki temu już dane z NSP 2011 zostały zaprezentowane w Portalu Geostatystycznym GUS[4] w przekroju siatki kilometrowej.

Zgodnie z definicją zawartą w Rozporządzeniu Rady Ministrów z dnia 15 października 2012 r. w sprawie państwowego systemu odniesień przestrzennych siatka kilometrowa to siatka odniesienia dla współrzędnych płaskich prostokątnych, przy czym:

- współrzędne geodezyjne narożników arkuszy map i linie siatki kartograficznej opisuje się w pełnych stopniach, minutach lub sekundach;
- linie siatki kilometrowej opisuje się w metrach lub kilometrach;
- dopuszcza się podawanie tylko punktów przecięcia siatek odniesienia, o których mowa powyżej.

Siatka kilometrowa otrzymuje nazwę od układu współrzędnych, dla którego została obliczona, przy czym:

- początek siatki pokrywa się z początkiem układu współrzędnych płaskich prostokątnych;

[4] https://geo.stat.gov.pl/.

- linie siatki biegną z południa na północ i z zachodu na wschód;
- punktem odniesienia komórki siatki jest lewy dolny narożnik komórki siatki.

Układ odniesienia siatki kilometrowej, w której GUS publikuje obecnie dane z NSP 2011 i 2021, to Europejski Ziemski System Odniesienia 1989 (ETRS89-LAEA). W skład kodu komórki siatki (np. PL_CRS3035RES1000mN3057000E5131000) wchodzą:

- CRS3035 – niepowtarzalny identyfikator przyjętego układu odniesienia (ETRS89-LAEA);
- RES1000 m – rozdzielczość komórki siatki (1000 m);
- N[y] – koduje wartość północną [N], gdzie y to współrzędna w metrach lewego dolnego rogu komórki siatki;
- E[x] – koduje wartość wschodnią [E], gdzie x to współrzędna w metrach lewego dolnego rogu komórki siatki.

Ponadto w Rozporządzeniu (WE) nr 1059/2003 Parlamentu Europejskiego i Rady z dnia 26 maja 2003 r. w sprawie ustalenia wspólnej klasyfikacji Jednostek Terytorialnych do Celów Statystycznych (NUTS) ustanowiono metodykę dla wszystkich członków Unii Europejskiej (UE) dotyczącą określania typologii terytorialnej na podstawie rozmieszczenia ludności w komórkach siatki kilometrowej. Z kolei Rozporządzenie wykonawcze Komisji (UE) 2018/1799 z dnia 21 listopada 2018 r. w sprawie ustanowienia tymczasowego bezpośredniego działania w dziedzinie statystyki dotyczącego rozpowszechniania wybranych tematów spisu ludności i mieszkań w 2021 r. i geokodowanych w siatce kilometrowej (dalej: rozporządzenie 2018/1799) jako tymczasowe bezpośrednie działanie w dziedzinie statystyki towarzyszące spisom powszechnym ludności i mieszkań w 2021 r. określa obowiązek opracowania głównych rezultatów ostatnich spisów powszechnych na podstawie siatki kilometrowej obejmującej całą Europę. Należy również wspomnieć, że Komisja Europejska planuje nałożenie na kraje członkowskie UE obowiązku corocznej publikacji danych demograficznych w przekroju siatek kilometrowych na podstawie danych z rejestrów administracyjnych. Według wstępnych założeń od 2025 r. corocznie mają być publikowane dane dotyczące ludności, a od 2031 r. – bardziej szczegółowe dane dotyczące mieszkań.

GUS opublikował dane demograficzne z NSP 2011 i 2021 w przekroju siatki kilometrowej (m.in. ogólną liczbę ludności, ludność według płci i biologicznych grup wieku oraz współczynnik feminizacji). Można je przeglądać w Portalu Geostatystycznym i pobierać z niego oraz za pośrednictwem usługi INSPIRE. Dzięki wykorzystaniu układu współrzędnych ETRS89-LAEA polskie dane mogą być łączone i porównywane z danymi europejskimi.

Należy również zaznaczyć, że w rozporządzeniu 2018/1799 określono zakres danych w przekroju siatki kilometrowej dostarczanych do Eurostatu przez kraje

członkowskie UE. Spośród nich w niniejszym artykule wzięto pod uwagę następujące zmienne i kategorie:

- ludność ogółem;
- płeć:
  – mężczyźni,
  – kobiety;
- grupy wieku:
  – poniżej 15 lat,
  – od 15 do 64 lat,
  – 65 lat i więcej.

W ramach dalszych prac autorzy planują analizę kolejnych kategorii i zmiennych[5].

Trzeba dodać, że informacji o liczbie osób pracujących dostarcza się, jeśli są dostępne na tym poziomie w kraju członkowskim. Zbiór przepisów dotyczących regulacji, które odnoszą się do rundy spisów z 2021 r. – ze szczególnym uwzględnieniem poziomu (rozmiaru komórek) siatki kilometrowej – zawiera publikacja Eurostatu (2019).

3. Podstawy formalne i rekomendacje międzynarodowe

Spośród rekomendacji Eurostatu w zakresie metod kontroli ujawniania danych zalecanych do stosowania w spisach powszechnych jedną z najważniejszych jest celowana wymiana rekordów (ang. *targeted record swapping* – TRS). To przykład podejścia pretablicowego, czyli metody ochrony stosowanej w odniesieniu do mikrodanych, zwłaszcza takich, na podstawie których powstają tablice wynikowe. Ten rodzaj wymiany rekordów zamienia nierówne wartości zmiennych – najczęściej geograficznych, choć nie tylko – pomiędzy odpowiednio sparowanymi jednostkami. Polega to na tym, że znajduje się rekordy z wysokim ryzykiem ujawnienia i wymienia się w nich określone dane na informacje pochodzące z innego (podobnego) rekordu. Najistotniejsze założenia TRS są następujące:

- wymiana dokonywana jest jedynie między gospodarstwami domowymi (lub innymi podobnie predefiniowanymi grupami jednostek);
- wyznaczane są jednostki o wysokim poziomie ryzyka ujawnienia w odniesieniu do wartości jednej lub większej liczby rozpatrywanych zmiennych;

[5] Będą to: (1) bieżąca aktywność ekonomiczna (osoby pracujące); (2) kraj/miejsce urodzenia (miejsce urodzenia w innym kraju członkowskim UE; miejsce urodzenia gdzie indziej); (3) miejsce zamieszkania na rok przed spisem (bez zmiany miejsca zamieszkania; przemieszczenie na terytorium kraju zgłaszającego; przemieszczenie poza terytorium kraju zgłaszającego).

- za pomocą schematu losowania probabilistycznego z wysokim prawdopodobieństwem wyłania się próbkę gospodarstw domowych (lub innych predefiniowanych grup jednostek) o wysokim stopniu ryzyka ujawnienia;
- każde gospodarstwo domowe (grupa) z tej próbki jest parowane z innym, podobnym w zakresie określonych zmiennych;
- dokonuje się wymiany danych z zakresu zmiennych docelowych (np. geograficznych) pomiędzy sparowanymi gospodarstwami domowymi (grupami).

Jak wspomniano, z uwagi na pierwotne zastosowanie to podejście odnosi się do najpowszechniej stosowanych w spisach grup jednostek, a więc do określonego poziomu podziału terytorialnego (np. podregionów lub powiatów), którego składowe podlegają wymianie w rekordach. Wymiana jest dokonywana w obrębie gospodarstw domowych. W innego rodzaju badaniach to odniesienie może być odmienne (np. w przypadku badań gospodarczych wymiana następuje między grupami Polskiej Klasyfikacji Działalności i klastrami przedsiębiorstw). Celowaną wymianę rekordów można przeprowadzić, używając pakietu `recordSwapping` środowiska R dostępnego na platformie GitHub[6], ostatnio włączonego także jako procedura `recordSwap` do pakietu `sdcMicro` (Gussenbauer, 2023; w kodzie źródłowym wykorzystano skrypty napisane w języku C++). Ocena ryzyka ujawnienia informacji wrażliwych oparta jest w tym przypadku głównie na regule k-anonimowości.

Innym podejściem rekomendowanym przez Eurostat jest metoda kluczy komórkowych (ang. *cell-key method* – CKM). Zalicza się ona do podejść posttablicowych, czyli stosowanych do ochrony już naliczonych tablic wynikowych, a polega na zakłócaniu wartości w komórkach tablicy z wykorzystaniem pewnych liczb losowych (kluczy rekordów – ang. *record keys*) przyporządkowywanych do rekordów, na podstawie których tworzona jest tablica, oraz specjalnej tabeli zakłóceń (p-*table* – p-tabeli). CKM składa się z następujących kroków:

- każdemu rekordowi przyporządkowywana jest losowa liczba naturalna (klucz rekordu) z zakresu od 1 do n, gdzie n jest predefiniowaną maksymalną możliwą wartością klucza;
- tworzona jest tablica docelowa, po czym dla każdej jej komórki sumowane są klucze tworzących ją rekordów modulo n;
- dzięki użyciu predefiniowanej (dobranej optymalnie dla danego przypadku) p-tablicy (zawierającej liczby 1 i –1) otrzymuje się wartość zakłócenia;
- do każdej komórki docelowej tablicy dodawane jest odpowiadające jej zakłócenie z p-tabeli.

[6] https://github.com/sdcTools/recordSwapping.

Metoda kluczy komórkowych może być stosowana z wykorzystaniem pakietu `cellKey` środowiska R. Warto jednak zauważyć, że ten pakiet wciąż jest testowany, dlatego znajduje się na platformie GitHub[7], choć 17 marca 2023 r. umieszczono go również w środowisku R[8]. W Polsce CKM testowano na danych z NSP 2011.

4. Ludność rezydująca według płci i grup wieku w przekroju siatki kilometrowej

Ochrona tajemnicy statystycznej zastosowana w odniesieniu do ludności rezydującej i skutkująca opublikowaniem danych na Portalu Geostatystycznym[9] opierała się na następujących założeniach:

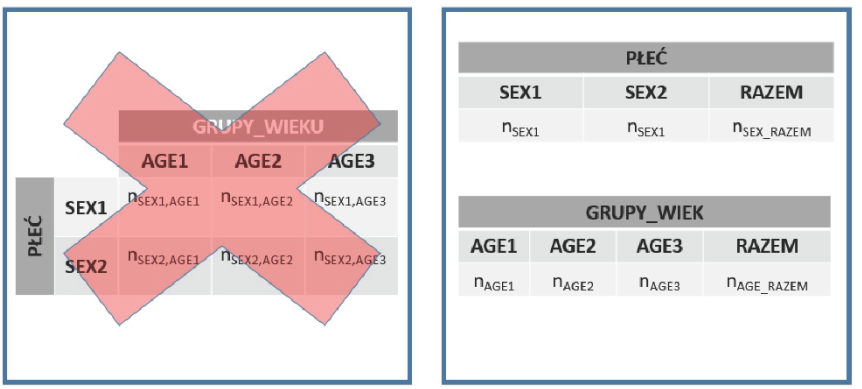
1. Nie należy uwzględniać niezamieszkanych komórek siatki kilometrowej.
2. Informacja na temat ogólnej liczby ludności rezydującej w komórce siatki kilometrowej nie jest objęta tajemnicą statystyczną, a więc jest znana.
3. Zostały udostępnione trzy zmienne na poziomie osób tworzących populację ludności rezydującej:
   - *identyfikator komórki siatki kilometrowej*;
   - *płeć*, przyjmująca warianty:
   – 1 dla mężczyzn (SEX1),
   – 2 dla kobiet (SEX2);
   - *grupa wieku*, przyjmująca warianty:
   – 1 dla osób w wieku poniżej 15 lat (AGE1),
   – 2 dla osób w wieku 15–64 lat (AGE2),
   – 3 dla osób w wieku 65 lat i więcej (AGE3).
4. Ochrona tajemnicy statystycznej za pomocą pierwotnego i wtórnego ukrywania komórek (ang. *primary/secondary supression*) realizowana jest z zastosowaniem zasad:
   - k-anonimowości, gdzie przyjęto $k = 3$;
   - l-różnorodności, gdzie przyjęto $l = 2$.
5. Publikowane są jedynie rozkłady brzegowe zmiennych płeć i grupa wieku w komórkach siatki kilometrowej (schemat).

[7] https://github.com/sdcTools/cellKey.

[8] https://cran.r-project.org/web/packages/cellKey/index.html.

[9] Zgodnie z harmonogramem podanym na stronie: https://stat.gov.pl/spisy-powszechne/nsp-2021/harmonogram-publikacji-wynikow-nsp-2021/.

Schemat. Rozkłady publikowane w Portalu Geostatystycznym



Źródło: opracowanie własne.

Należy dodać, że w przypadku gdy dwa warianty grup wieku były uznane za bezpieczne, czyli o niskim ryzyku ujawnienia ($n \geq 3$), a trzeci wariant był niebezpieczny ($n < 3$) i jednocześnie warianty bezpieczne były równoliczne, to obok oczywistego ukrycia wariantu niebezpiecznego należało zastosować odpowiedni mechanizm ukrywania jednego z wariantów bezpiecznych. Rozważano podejścia oparte na:

- konsekwentnym ukrywaniu określonego przekroju:

$$P(p_2 = \#|p_1 = u \wedge p_3 = s) = 1, \tag{1}$$

$$P(p_3 = \#|p_2 = u \wedge p_1 = s) = 1, \tag{2}$$

$$P(p_1 = \#|p_3 = u \wedge p_2 = s) = 1; \tag{3}$$

- losowym ukrywaniu określonego przekroju (opartym na rozkładzie równomiernym):

$$P(p_2 = \#|p_1 = u) = P(p_3 = \#|p_1 = u) = 0{,}5, \tag{4}$$

$$P(p_1 = \#|p_2 = u) = P(p_3 = \#|p_2 = u) = 0{,}5, \tag{5}$$

$$P(p_1 = \#|p_3 = u) = P(p_2 = \#|p_3 = u) = 0{,}5; \tag{6}$$

- losowym ukrywaniu określonego przekroju (opartym na rozkładzie grup wieku w populacji):

$$P(p_1 = \#|p_2 = u) = 1 - \frac{n_{p_1}}{n_{p_1} + n_{p_3}}$$
$$P(p_3 = \#|p_2 = u) = 1 - \frac{n_{p_3}}{n_{p_1} + n_{p_3}}, \tag{7}$$

$$P(p_1 = \#|p_3 = u) = 1 - \frac{n_{p_1}}{n_{p_1} + n_{p_2}}$$
$$P(p_2 = \#|p_3 = u) = 1 - \frac{n_{p_2}}{n_{p_1} + n_{p_2}}, \tag{8}$$

$$P(p_2 = \#|p_1 = u) = 1 - \frac{n_{p_2}}{n_{p_2} + n_{p_3}}$$
$$P(p_3 = \#|p_1 = u) = 1 - \frac{n_{p_3}}{n_{p_2} + n_{p_3}}, \tag{9}$$

gdzie:
p_1, p_2, p_3 – grupy wieku odpowiednio: do 15 lat, 15–64 lata oraz 65 lat i więcej,
$\#$ – przekrój z zastosowaną ochroną tajemnicy statystycznej,
s – przekrój bezpieczny, tj. $n \geq 3$,
u – przekrój niebezpieczny, tj. $n < 3$.

Ostatecznie zastosowano wariant oparty na rozkładzie częstości grup wieku w populacji. Podobne podejście przyjęto w przypadku, gdy liczebność dwóch wariantów grup wieku była równa 0, a trzeci był niebezpieczny ($n < 3$).

Analizę rozkładu komórek siatki w podziale na przekroje potencjalnie bezpieczne ($n \geq 3$) i potencjalnie niebezpieczne ($n < 3$) przedstawiono w tabl. 1.

Tabl. 1. Rozkład liczby rezydentów w komórkach siatki kilometrowej

Liczba rezydentów w komórkach siatki	Komórki siatki		Rezydenci		Kategoria komórek
	liczba	udział w ogólnej liczbie w %	liczba	udział w ogólnej liczbie w %	
1	3 424	1,74	3 424	0,01	potencjalnie niebezpieczne
2	4 213	2,14	8 426	0,02	potencjalnie niebezpieczne
3 i więcej	189 493	96,13	37 007 477	99,97	potencjalnie bezpieczne
Ogółem	197 130	100,00	37 019 327	100,00	.

Źródło: opracowanie własne na podstawie bazy danych z NSP 2021.

W wyniku prac przeprowadzonych z wykorzystaniem oprogramowania SAS dokonano pierwotnego albo pierwotnego i wtórnego ukrycia przekrojów niebezpiecznych, co skutkowało pewną utratą informacji. Należy jednak podkreślić, że w przypadku zmiennej:

- *płeć* bezwzględna strata informacji obliczona jako różnica między liczbą ludności rezydującej w oryginalnym zbiorze danych a liczbą ludności rezydującej opublikowaną na Portalu Geostatystycznym wyniosła 91 719 osób, tj. 0,25%;
- *grupa wieku* bezwzględna strata informacji obliczona jako różnica między liczbą ludności rezydującej w oryginalnym zbiorze danych a liczbą ludności rezydującej opublikowaną na Portalu Geostatystycznym wyniosła 261 560 osób, tj. 0,71%.

Analiza przestrzennego rozkładu zastosowanych technik ochrony tajemnicy statystycznej pokazuje, że ukrycia komórek siatki kilometrowej występują na terenie całego kraju, choć w nieco większym stopniu w północno-wschodniej Polsce (mapa, s. 54). Na submapie A przedstawiono jednoczesne osobne ukrycia poszczególnych kategorii płci i grup wieku, na submapie B – jednoczesne ukrycia agregacji krzyżowych.

5. Wyzwania związane z udostępnianiem danych o ludności w przekroju siatki kilometrowej

Opisane wyżej wyniki wskazują na złożoność problemu, z jakim przyszło się zmierzyć statystykom. Już w przypadku czynności przedstawionych we wcześniejszej części artykułu można było zauważyć potrzebę zastosowania metod probabilistycznych w ukrywaniu wtórnym. Dotyczyło to zwłaszcza sytuacji, w których typowe metody okazały się niewystarczające. Co ważne, w przypadku większej liczby zmiennych przewidzianych do udostępnienia (zob. cz. 2) kombinacje wartości podlegające ochronie mogą przyjmować coraz bardziej złożoną postać. W takiej sytuacji konieczne staje się zastosowanie bardziej zaawansowanych narzędzi SDC.

Z uwagi na rozmiar zbioru (ponad 197 tys. komórek siatki kilometrowej), charakter zawartych w nim informacji i potencjalne wykorzystywanie udostępnianych danych możliwe są dwa zasadnicze podejścia:

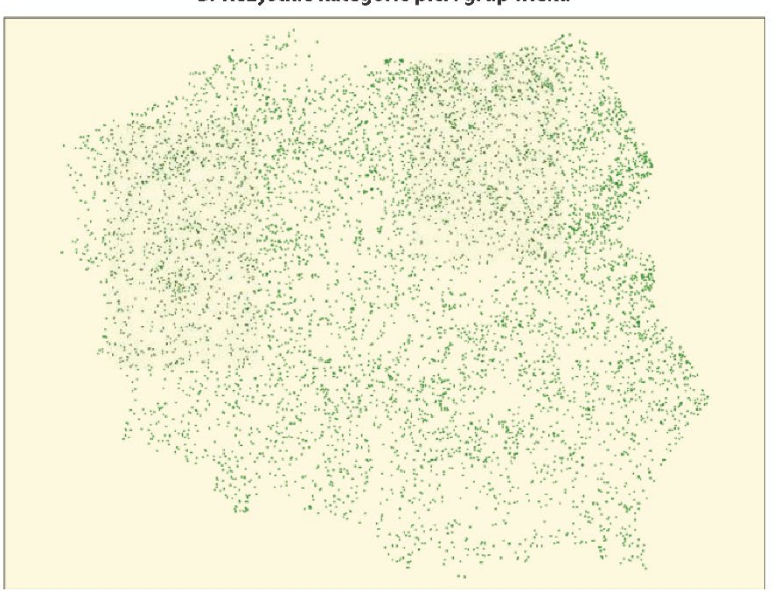
- potraktowanie zbioru jako bazy mikrodanych i zastosowanie w odniesieniu do niego odpowiednich metod SDC (np. celowana wymiana rekordów w połączeniu z mikroagregacją lub zaokrąglaniem[10]);

[10] Jak pokazują doświadczenia z innych badań – opisane m.in. w pracach Młodaka i in. (2022) oraz Pietrzaka i in. (2022) – z uwagi na ograniczone spektrum odniesienia zastosowanie TRS może się okazać niewystarczające.

Mapa. Rozkłady przestrzenne jednoczesnych ukryć

A. Wszystkie kategorie płci lub grup wieku

B. Wszystkie kategorie płci i grup wieku



Źródło: opracowanie własne na podstawie bazy danych z NSP 2021.

- przyjęcie, że zbiór jest tablicą wielowymiarową. Można wówczas wykorzystać narzędzia SDC, stosowane głównie w przypadku tablic częstości, przede wszystkim metodę kluczy komórkowych. Alternatywnie warto w tym kontekście rozważyć metodę kontrolowanego dopasowania tablicy (ang. *controlled tabular adjustment* – CTA) lub podejście hiperkostkowe (ang. *hypercube*), oparte na heurystycznej ochronie całościowej polegającej na ochronie podtablicy i bazującej na znajdowaniu i ukrywaniu struktur geometrycznych (kostek n-wymiarowych), co jest niezbędne do ochrony komórek wrażliwych wskazanych podczas ukrywania pierwotnego (Hundepool i in., 2012; Młodak i in., 2023).

Trzeba się liczyć z możliwością ujawnienia wrażliwego atrybutu osoby w wyniku połączenia dwóch różnych zbiorów. Z uwagi na to, że oprócz danych dotyczących ludności rezydującej udostępniane są także informacje dotyczące ludności według definicji krajowej (dla tych samych zmiennych), powiązanie obu zbiorów może pozwolić na uzyskanie dodatkowych wrażliwych informacji o osobie. Należy przypomnieć, że według definicji krajowej ludność naszego kraju tworzą stali mieszkańcy Polski, w tym osoby, które przebywają czasowo za granicą (bez względu na okres przebywania), ale zachowały stałe zameldowanie w Polsce. W tym ujęciu do ludności nie są natomiast zaliczani imigranci przebywający w Polsce czasowo. Z kolei ludność rezydująca to ogół osób mieszkających/przebywających na danym terenie przez co najmniej 12 miesięcy. Do rezydentów danej jednostki administracyjnej zalicza się wszystkie osoby mieszkające lub zamierzające mieszkać w tej jednostce nie krócej niż rok. Oznacza to, że w liczbie rezydentów danej jednostki administracyjnej ujęci są:

- stali mieszkańcy (osoby zameldowane lub stale mieszkające bez zameldowania), z wyjątkiem tych, którzy wyjechali na co najmniej 12 miesięcy do innego miejsca w kraju lub za granicę;
- osoby przybyłe z innego miejsca w kraju lub z zagranicy (imigranci bez karty stałego pobytu) na co najmniej 12 miesięcy.

Gdy dla danej zmiennej różnica między odpowiednimi agregatami w obu typach ludności jest niewielka, może się to przyczynić do identyfikacji osoby i ujawnienia jej danych wrażliwych. Oto przykład: jeśli w tym samym agregacie liczba ludności rezydującej jest o 1 większa od liczby ludności według definicji krajowej, to w konkretnych uwarunkowaniach lokalnych (np. brak osób, które wyjechały za granicę) może to oznaczać, że ową jedną osobą był imigrant przebywający w Polsce czasowo, ale dłużej niż 12 miesięcy. W konsekwencji można odczytać z danych jego płeć i wiek. Podczas przeprowadzania procesu SDC należy zatem uwzględnić i takie zagrożenie.

W tabl. 2 zaprezentowano możliwą skalę problemu przy założeniu, że zagrożenie dla poufności danych pojawia się przede wszystkim wtedy, gdy wartość bezwzględna wspomnianej różnicy jest większa od 0, ale mniejsza od 3.

Tabl. 2. Różnice między liczbą ludności rezydującej a liczbą ludności według definicji krajowej dla zmiennych prezentowanych w przekroju siatki kilometrowej

Zmienne	Różnice							
	–2		–1		1		2	
		w %		w %		w %		w %
SEX1	15200	7,71	28654	14,54	7599	3,86	1568	0,80
SEX2	15347	7,79	26964	13,68	7901	4,01	1609	0,82
AGE1	10523	5,34	17699	8,98	2809	1,43	1012	0,51
AGE2	16063	8,15	27250	13,83	8772	4,45	2028	1,03
AGE3	2614	1,33	8607	4,37	3909	1,98	397	0,20

Źródło: opracowanie własne na podstawie bazy danych z NSP 2021.

Jak widać, liczba i odsetek problemowych różnic są dość duże – dotyczy to zwłaszcza opisanej wcześniej sytuacji, gdy liczba ludności według definicji krajowej jest o 1 większa od liczby ludności rezydującej. Problem wydaje się wprawdzie w naj-większym stopniu dotyczyć mniejszych zbiorów, ale w określonych okolicznościach może okazać się istotny także dla tych większych.

Należy również pamiętać, że w przypadku ochrony zbioru o większej liczbie zmiennych – która wymaga wprowadzenia szerszego zakresu ukryć lub zniekształceń – wzrośnie też strata informacji spowodowana zastosowaniem SDC. Tymczasem osiągnięcie balansu między minimalizacją ryzyka ujawnienia a minimalizacją straty informacji (co oznacza maksymalizację użyteczności danych dla użytkownika) to kluczowy cel tego procesu. Istotne wyzwanie stanowi zatem wypracowanie kompromisu między tymi dwoma dążeniami. Zastosowanie rekomendowanych przez Eurostat metod TRS i CKM, względnie podejść dodatkowych lub alternatywnych, będzie wymagało dostępu do zbioru danych jednostkowych w ostatecznej postaci i w pełni funkcjonalnego środowiska operacyjnego obsługującego specjalistyczne pakiety programu R.

6. Podsumowanie

Skuteczna ochrona danych udostępnianych w przekroju siatki kilometrowej wymaga zaawansowanych technik wykorzystujących narzędzia kontroli ujawniania danych. Ochrona pojedynczego zasobu nie wiąże się z zastosowaniem szczególnie złożonych rozwiązań. Większe wyzwanie pojawia się w sytuacji, gdy należy ochronić kilka powiązanych ze sobą baz danych. Wówczas ukrywanie danych w każdym pliku niezależnie może nie wystarczyć, jako że występowanie logicznych i matematycznych powiązań między zbiorami – zwłaszcza gdy zmienne w obu zbiorach bazują na jednolitych fundamentach definicyjnych – stwarza ryzyko wtórnego odtworzenia danych wrażliwych, a tym samym identyfikacji jednostki. W takiej sytuacji konieczne

okazują się dodatkowe ukrycia dokonane na podstawie analizy możliwości reidentyfikacji jednostki oraz maksymalnego możliwego zachowania użyteczności danych. Być może niezbędne w tej sytuacji będzie np. użycie wyższych poziomów agregacji.

Inną kwestią pozostaje gęstość samej siatki. Obecnie najczęściej stosuje się komórki o boku 1 km×1 km. W wielu krajach w użyciu bywają jednak siatki gęstsze – o komórkach 500 m×500 m lub 200 m×200 m. Można też używać siatek o większych komórkach (np. 2 km×2 km lub 5 km×5 km). Ponadto metodologiczne podstawy konstrukcji takich siatek nie zawsze są spójne. Dodatkowo mogą one tworzyć strukturę hierarchiczną i wówczas mamy do czynienia z siatką wielopoziomową. W takim przypadku należy rozważyć zastosowanie podejścia zaproponowanego przez Branchu i in. (2018). Chodzi o metodę zakłócającą wartości zagregowane w sposób deterministyczny, która opiera się na dwóch krokach. W pierwszym kroku należy utworzyć podstawowe grupy komórek siatki w następujący sposób:

- połączyć komórki zawierające dane obarczone ryzykiem ujawnienia z innymi komórkami, tak aby utworzyć grupy komórek (mogą to też być komórki wyższego poziomu), w których dane pozostaną bezpieczne w myśl przyjętej reguły ryzyka ujawnienia (np. k-anonimowości);
- jeśli to możliwe, zastosować kilka siatek dla tego samego obszaru, co pomaga ograniczyć poszukiwanie najlepszych komórek do połączenia w określonym obszarze (na podstawie odpowiednio zdefiniowanej odległości, ale przy mniejszym nakładzie obliczeniowym);
- algorytm rozpoczyna się od poziomu najbardziej zgrubnego (czyli siatki, w której wszystkie komórki są bezpieczne, a zatem ryzyko ujawnienia jest małe), po czym stopniowo zniża się do poziomu najdokładniejszego.

W drugim kroku należy proporcjonalnie nałożyć zakłócenia. Oznacza to, że każda liczba na poziomie grupy otrzymanej w pierwszym kroku (np. liczba kobiet w wieku 65 lat i więcej) rozkłada się na komórki grupy proporcjonalnie do liczby ludności w każdej komórce siatki – wówczas ochrona jest silna. Ryzyko identyfikacji jednostki pojawia się jedynie wówczas, gdy rozkład danego atrybutu jest identyczny we wszystkich komórkach grupy (to bardzo rzadki przypadek). W takiej sytuacji komórki te należy złączyć z komórkami najbliższej grupy i poddać stosownemu zakłóceniu.

Zaletą prezentowanej metody (stosowanej np. we francuskim Institut national de la statistique et des études économiques – pol. Narodowy Instytut Statystyki i Badań Ekonomicznych) jest zachowanie addytywności w komórkach siatki. Ponadto zakłócenia są monitorowane geograficznie, jako że w wielu przypadkach odległość pomiędzy komórkami siatki kilometrowej nie jest zbyt duża. Dla zminimalizowania ryzyka ujawnienia informacji chronionych dokładny skład grup nie jest zazwyczaj ujawniany.

Analiza możliwości zastosowania w razie potrzeby tego rodzaju podejść oraz ich powiązania z omówionymi w artykule stanowi wyzwanie dla statystyki na najbliższe lata. Co ważne, w każdym przypadku konieczna jest ocena ryzyka ujawnienia i oczekiwanej straty informacji na skutek zastosowania opisanych mechanizmów.

Podziękowania

Autorzy składają podziękowania pani Dorocie Szałtys z Departamentu Badań Demograficznych GUS za cenne uwagi zgłoszone na etapie przygotowywania artykułu.

Bibliografia

Branchu, M., Costemalle, V., & Fontaine, M. (2018). Données carroyées et confidentialité. *Journées de Méthodologie Statistiques de l'Insee*, *13*, 1–23. http://www.jms-insee.fr/2018/S23_3_ACTE_BRANCHU_JMS2018.pdf.

Dygaszewicz, J. (2012). Spisy powszechne jako źrodło danych do analiz geoprzestrzennych. *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, *23*, 91–100. https://ptfit.sgp.geodezja.org.pl/wydawnictwa/archiwum_vol_23/Dygaszewicz.pdf.

Eurostat. (2019). *EU legislation on the 2021 population and housing censuses. Explanatory notes*. Publications Office of the European Union. https://ec.europa.eu/eurostat/docments/3859598/9670557/KS-GQ-18-010-EN-N.pdf/c3df7fcb-f134-4398-94c8-4be0b7ec0494?t=1552653277000.

Gussenbauer, J. (2023, 30 sierpnia). *Target Record Swapping*. https://cran.r-project.org/web/packages/sdcMicro/vignettes/recordSwapping.html.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., & de Wolf, P.-P. (2012). *Statistical Disclosure Control*. John Wiley & Sons. https://doi.org/10.1002/9781118348239.

Młodak, A., Pietrzak, M., & Józefowski, T. (2022). The trade-off between the risk of disclosure and data utility in SDC: A case of data from a survey of accidents at work. *Statistical Journal of the IAOS*, *38*(4), 1503–1511. https://doi.org/10.3233/SJI-220936.

Młodak, A., Pietrzak, M., Klimanek, T., Józefowski, T., & Lańduch, P. (2023). *Poufność a użyteczność informacji statystycznych. Dylematy ochrony udostępnianych danych*. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu. https://doi.org/10.18559/978-83-8211-168-2.

Pietrzak, M., Józefowski, T., Klimanek, T., & Młodak, A. (2022). An Optimized Selection of Statistical Disclosure Control Methods – A Case Study Involving Microdata from the Polish Survey of Accidents at Work. W: K. Jajuga, G. Dehnel, M. Walesiak (Eds.), *Modern Classification and Data Analysis: Methodology and Applications to Micro- and Macroeconomic Problems* (pp. 63–78). Springer. https://doi.org/10.1007/978-3-031-10190-8_6.

Rozporządzenie Rady Ministrów z dnia 15 października 2012 r. w sprawie państwowego systemu odniesień przestrzennych (Dz.U. 2012 poz. 1247 ze zm.).

Rozporządzenie (WE) nr 1059/2003 Parlamentu Europejskiego i Rady z dnia 26 maja 2003 r. w sprawie ustalenia wspólnej klasyfikacji Jednostek Terytorialnych do Celów Statystycznych (NUTS) (Dz.U. UE L 154/1).

Rozporządzenie wykonawcze Komisji (UE) 2018/1799 z dnia 21 listopada 2018 r. w sprawie ustanowienia tymczasowego bezpośredniego działania w dziedzinie statystyki dotyczącego rozpowszechniania wybranych tematów spisu ludności i mieszkań w 2021 r. i geokodowanych w siatce kilometrowej (Dz.U. L 296 z 22.11.2018).

System informacji przestrzennej. (n.d.). W: *Internetowy leksykon geomatyczny*. https://www.ptip.info/LMB/system-informacji-przestrzennej.

United Nations. (2017). *Principles and Recommendations for Population and Housing Censuses: the 2020 Round* (Revision 3). https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(7), 59–60
In memoriam

Dr Kazimierz Kruszka (1940–2024)

Fot. Archiwum Uniwersytetu Ekonomicznego w Poznaniu

Dr Kazimierz Kruszka urodził się 18 grudnia 1940 r. W latach 1959–1964 studiował w Wyższej Szkole Ekonomicznej w Poznaniu (obecnie Uniwersytet Ekonomiczny w Poznaniu – UEP), a po uzyskaniu tytułu magistra ekonomii rozpoczął pracę na tej uczelni. W 1971 r. obronił rozprawę doktorską pt. *Statystyczna analiza zużycia energii elektrycznej w ujęciu regionalnym*. Był cenionym nauczycielem akademickim, który wykształcił kilka pokoleń statystyków. Współtworzył Centrum Statystyki Regionalnej, organizował wykłady, warsztaty i szkolenia (m.in. dla pracowników administracji, dziennikarzy, księgowych i rewidentów), a niektóre także prowadził. Inicjował wiele wspólnych przedsięwzięć Katedry Statystyki UEP i Urzędu Statystycznego (US) w Poznaniu. W 2007 r. przeszedł na emeryturę, ale był związany z poznańską uczelnią do końca życia.

Przez kilkanaście lat piastował odpowiedzialne funkcje w US w Poznaniu: w latach 1990–1992 był zastępcą dyrektora, następnie – dyrektorem, a w latach 1996–2007 – ponownie zastępcą dyrektora. Zainicjował utworzenie Ośrodka Statystyki Miast, a także włączenie go do międzynarodowej sieci jednostek prowadzących badania Urban Audit; od 2007 r. pracował tam na stanowisku konsultanta. Położył podwaliny pod powstanie Ośrodka Statystyki Małych Obszarów.

Przyczynił się do reaktywowania w kwietniu 1981 r. Polskiego Towarzystwa Statystycznego (PTS), a w maju tego roku zorganizował poznański oddział PTS, którego pracami kierował przez wiele lat. Był również członkiem Rady Głównej i Prezydium Rady Głównej PTS. W latach 1994–2000 pełnił funkcję wiceprezesa PTS, następnie – sekretarza Rady Głównej, a w latach 2005–2010 – prezesa. W 2012 r. otrzymał tytuł Honorowego Członka PTS. Należał do inicjatorów i organizatorów Kongresu Statystyki Polskiej z okazji 100-lecia PTS, który odbył się w 2012 r. w Poznaniu. Również z jego inicjatywy zorganizowano wystawę *Statystyka w Wielkopolsce* otwartą podczas kongresu, a potem prezentowaną w różnych instytucjach w Poznaniu i regionie.

W swoim dorobku miał – jako autor lub redaktor – ponad sto publikacji naukowych i opracowań statystycznych oraz dotyczących historii statystyki

© Główny Urząd Statystyczny
Artykuł udostępniony na licencji CC BY-SA 4.0 / Article available under the CC BY-SA 4.0 licence

w Wielkopolsce, a także artykuły na temat rodzinnego Wągrowca i ziemi pałuckiej. Niestrudzenie popularyzował statystykę. Wygłaszał odczyty poświęcone statystyce i organizował konkursy dla szkół promujące tę dziedzinę wiedzy. Z jego inicjatywy PTS, US w Poznaniu i UEP zorganizowały Wielkopolski Konkurs *Statystyka mnie dotyka,* popularyzujący wiedzę o statystyce wśród uczniów szkół ponadgimnazjalnych.

Kazimierz Kruszka zasiadał w zarządzie Polskiego Towarzystwa Ekonomicznego i Poznańskiego Towarzystwa Przyjaciół Nauk. Nieoceniona była jego działalność ekspercka jako wieloletniego członka Komisji Programowo-Metodologicznej Głównego Urzędu Statystycznego, członka Rady Strategii Rozwoju Miasta Poznania i doradcy Marszałka Województwa Wielkopolskiego.

Za swoje zasługi był wielokrotnie nagradzany wysokimi odznaczeniami państwowymi i resortowymi: Krzyżem Kawalerskim Orderu Odrodzenia Polski, Srebrnym i Złotym Krzyżem Zasługi, Odznaką Honorową „Za Zasługi dla Statystyki Rzeczypospolitej Polskiej", Odznaką „Za Zasługi dla Rozwoju Województwa Wielkopolskiego", jak również Złotą Odznaką Polskiego Towarzystwa Ekonomicznego.

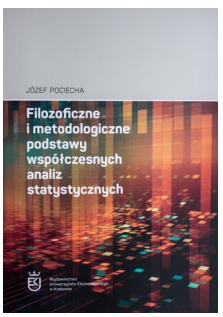
Zmarł 7 lipca 2024 r.

Redakcja „Wiadomości Statystycznych. The Polish Statistician"

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(7), 61–64 https://doi.org/10.59139/ws.2024.07.4
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

Recenzja książki Józefa Pociechy *Filozoficzne i metodologiczne podstawy współczesnych analiz statystycznych*

Review of Józef Pociecha's book *Philosophical and methodological foundations of modern statistical analyses*



Język/Language: polski/Polish
Wydawnictwo/Publisher: Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie
Miejsce i rok wydania / Place and year of publication: Kraków 2023
Liczba stron / Number of pages: 172

Józef Pociecha w monografii *Filozoficzne i metodologiczne podstawy współczesnych analiz statystycznych* przedstawia swoje spojrzenie na podwaliny współczesnej statystyki. Czytelnikowi mniej zorientowanemu w problematyce teorii statystyki i jej zastosowań mogłoby się wydawać, że tytułowe kwestie zostały ugruntowane i opisane w minionych dziesięcioleciach. Tak jednak nie jest. Po pierwsze rozwój technik przetwarzania danych liczbowych i coraz większe możliwości dotarcia do różnych zbiorów informacji liczbowych prowadzą do przewartościowania znanych dotychczas paradygmatów – zarówno jeśli chodzi o założenia (w tym ich realność), jak i walory poznawcze. Po drugie za sprawą postępu technologicznego w zakresie gromadzenia, przetwarzania, prezentowania i przesyłania danych statystycznych kształtuje się nowy interesujący paradygmat w obrębie statystyki czy też *nauki o danych* (ang. *data science*), który Józef Pociecha nazywa *paradygmatem uczenia statystycznego*. Zwrócenie uwagi na ten paradygmat i jego dogłębne scharakteryzowanie w recenzowanej monografii należy do jej zasadniczych zalet, ponieważ czyni tę pracę aktualną i ważną.

Innym jej walorem jest fachowe usystematyzowanie założeń filozoficznych i podstaw metodycznych badań naukowych wykorzystujących metody i techniki statystyczne. Już we *Wprowadzeniu* autor stwierdza jednoznacznie, że „nie ma żadnej analizy statystycznej bez przyjęcia określonych podstaw filozoficznych" (s. 11). A określając cele monografii, wyjaśnia, że jego intencją było połączenie problematyki filozoficznej, statystycznej i informatycznej w rozważaniach o obecnych i nowo powstających paradygmatach w obrębie nauk ilościowych. Należy przyznać, że ten cel

© Mirosław Szreder
Artykuł udostępniony na licencji CC BY-SA 4.0 / Article available under the CC BY-SA 4.0 licence

udało się autorowi w znacznym stopniu osiągnąć, a przy tym zawrzeć w tekście wiele oryginalnych i przemyślanych wywodów – niekiedy elementów ciekawej dyskusji naukowej – na tematy dotyczące współczesnych wyzwań statystyki.

Monografia składa się z sześciu rozdziałów. Jak na pracę z zakresu statystyki bardzo niewiele jest w niej liczb, tablic i wykresów, a ściślej rzecz biorąc, w czterech pierwszych rozdziałach nie ma ich wcale. Ogólna struktura monografii i kolejność prezentowanych w niej treści tworzą logiczną i spójną całość. Dużą dysproporcję dostrzega się jedynie między obszernym (15-stronicowym) *Wprowadzeniem* a krótkim (2,5-stronicowym) *Podsumowaniem*, które włączono do ostatniego rozdziału i które w niewielkim stopniu odnosi się do całości.

W rozdziale pierwszym *Poszukiwanie prawdy w badaniach naukowych* autor wyjaśnia, że kategoria prawdy jest dla niego ważna nie tylko jako jedno z podstawowych pojęć filozoficznych (s. 23), lecz także z powodu poszukiwania odpowiedzi na pytanie, „jakiej prawdy o otaczającej nas rzeczywistości poszukujemy", przeprowadzając badania statystyczne (s. 29). Omówienie celów i funkcji nauki – w tym statystyki – w szerokim kontekście poznawczym należy uznać za zadanie ambitne i wykonane z powodzeniem. Zagadnienia opisane w tym rozdziale dotyczą filozoficznych podstaw rozwoju poznania naukowego, w tym najważniejszych problemów filozofii nauki. Autor odwołuje się do dorobku wybitnych filozofów nauki, m.in. Karla Poppera, Thomasa Kuhna, Rudolfa Carnapa i Władysława Tatarkiewicza, i poprawnie omawia ich poglądy.

Rozdział drugi *Filozoficzne i metodologiczne założenia badań statystycznych* należy do najistotniejszych w tej monografii, biorąc pod uwagę jej cel i tytuł. Autor, opierając się na koncepcjach logików i filozofów nauki, pochyla się przede wszystkim nad takimi zagadnieniami, jak determinizm a indeterminizm, losowość a niepewność, wnioskowanie dedukcyjne a wnioskowanie indukcyjne. Jeden z kluczowych podrozdziałów traktuje o paradygmacie wnioskowania statystycznego. Ogólnie rozdział należy uznać za oryginalną prezentację najważniejszych zagadnień z zakresu filozoficznych i metodologicznych podstaw poznania statystycznego – opisu i wnioskowania.

W rozdziale trzecim *Wyzwania współczesnych badań statystycznych* autor koncentruje się na wyzwaniach wiążących się bezpośrednio z dużymi zbiorami danych, do których mają dostęp badacze, oraz wynikających z rozwoju technologii informatycznych i oprogramowania stosowanego do obliczeń. Generalnie wspomniane zagadnienia zostały omówione rzetelnie i syntetycznie. Jedynie podczas lektury podrozdziału *Duże zbiory danych* czytelnik może odczuwać brak głębszej refleksji filozoficznej na temat wpływu rozwoju big data na postawy naukowców oraz życie jednostek i społeczeństw. Wskazane byłoby dodanie krótkiego komentarza do stwierdze-

nia Mayera-Schönbergera i Cukiera (2013), że w big data ważna jest odpowiedź na pytanie, co się dzieje, a nie dlaczego się dzieje, a także do sugestii autorów, że nie zawsze musimy znać przyczyny jakiegoś zjawiska, ponieważ powinniśmy pozwolić danym mówić za siebie (zgodnie z hasłem *let data speak for themselves*).

Rozdział czwarty *Podstawy i zasady działania sztucznej inteligencji* jest poświęcony podwalinom filozoficznym i matematycznym sztucznej inteligencji, a także jej zastosowaniom. W pierwszym kroku autor wyjaśnia i systematyzuje definicje sztucznej inteligencji. Czyni to poprawnie, ale wydaje się, że dużym ułatwieniem w zrozumieniu i przyswojeniu wiedzy dotyczącej kryteriów klasyfikacji definicji sztucznej inteligencji byłoby posłużenie się schematem lub inną formą wizualizacji. Można mieć również wątpliwości, czy podrozdział *Filozoficzne podstawy sztucznej inteligencji* nie powinien poprzedzać podrozdziału *Matematyczne podstawy sztucznej inteligencji*, zwłaszcza że pozwoliłoby to uzyskać większą spójność treści ostatnich kilku podrozdziałów. Niemniej cały rozdział stanowi ważną część monografii, ponieważ trudno byłoby omawiać podstawy metodyczne współczesnych analiz statystycznych bez przyjrzenia się roli sztucznej inteligencji.

W rozdziale piątym *Metodologiczne podstawy uczenia statystycznego* autor przedstawia – nowy w statystyce – paradygmat uczenia statystycznego. Jego prezentacja jest poprawna, ale zilustrowana jedynie schematem i kilkoma równaniami. Według autora, o ile punktem wyjścia paradygmatu statystyki matematycznej jest teoria rachunku prawdopodobieństwa, a jej celem sprawdzenie, „na ile dane empiryczne można wpasować w teoretyczne ramy statystyki matematycznej" (s. 113), o tyle „w przypadku paradygmatu uczenia statystycznego jest odwrotnie: *niech dane mówią same za siebie*, *uczymy się z danych*. Punktem wyjścia jest dostępny zbiór danych" (s. 113). Szkoda, że zabrakło nawet krótkiego (może krytycznego?) komentarza do przytoczonego tu hasła.

W rozdziale szóstym *Przegląd metod nadzorowanego uczenia statystycznego* autor charakteryzuje metody należące do wielowymiarowej analizy statystycznej, które można zaliczyć do grupy metod uczenia statystycznego. Większość z nich została opatrzona odpowiednimi wzorami i klarownie zaprezentowana, tak aby czytelnik zrozumiał kolejne etapy ich implementacji. W rozdziale znajdują się także liczne odwołania do literatury przedmiotu oraz podsumowanie monografii.

Brak odrębnego podsumowania rozważań zawartych w recenzowanej książce pozostawia niedosyt. Gdyby autor się na takie zdecydował, to powołując się na przemyślenia zawarte w publikacji, mógłby skonstatować za Moore'em (1998,

s. 1254), że „statystyka jest ogólną intelektualną metodą, która znajduje zastosowanie gdziekolwiek pojawiają się dane, zróżnicowanie i losowość”[1].

Omawiana monografia może zainteresować naukowców stosujących analizy statystyczne w różnych obszarach współczesnych badań, nie tylko w ekonomii. Z pewnością będzie także wartościowym uzupełnieniem podstawowego kanonu literatury wykorzystywanej w nauczaniu statystyki w szkołach wyższych.

Bibliografia

Mayer-Schönberger, V., Cukier, K. (2013). *BIG DATA. Rewolucja, która zmieni nasze myślenie, pracę i życie*. MT Biznes.

Moore, D. S. (1998). Statistics Among the Liberal Arts. *Journal of the American Statistical Association*, *93*(444), 1253–1259. https://doi.org/10.1080/01621459.1998.10473786.

Mirosław Szreder
Uniwersytet Gdański, Wydział Zarządzania, Polska/ University of Gdańsk, Faculty of Management, Poland. ORCID: https://orcid.org/0000-0002-7597-0816

[1] W oryginale: „Statistics is a general intellectual method that applies wherever data, variation, and chance appear”.

Wiadomości Statystyczne. The Polish Statistician, 2024, *69*(7), 65–66
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

WYDAWNICTWA GUS. CZERWIEC 2024
PUBLICATIONS OF STATISTICS POLAND. JUNE 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:

Uczenie się osób dorosłych 2022
Adult Education 2022

W opracowaniu zamieszczono wyniki czwartej edycji międzynarodowego badania „Uczenie się osób dorosłych" (Adult Education Survey) przeprowadzonego na reprezentatywnej próbie gospodarstw domowych zgodnie z założeniami metodologicznymi Eurostatu. Publikację przygotował Urząd Statystyczny w Gdańsku we współpracy z Głównym Urzędem Statystycznym.

Język: polski (przedmowa, spis treści, synteza, tablice i wykresy również w języku angielskim)
Language: Polish (preface, contents, executive summary, tables and charts available also in English)
Seria: Analizy statystyczne
Series: Statistical analyses
Dostępne wersje: elektroniczna z tablicami w formacie Excel
Available in: electronic form with Excel tables

Badanie omówione w publikacji dostarczyło danych na temat uczestnictwa osób dorosłych (w wieku 18–69 lat) w edukacji formalnej i pozaformalnej oraz uczeniu się nieformalnym. Edukacja formalna często okazuje się niewystarczająca, a szkolenia, kursy i inne formy dokształcania pozwalają zdobyć lub poszerzyć wiedzę i umiejętności przydatne zarówno w życiu zawodowym, jak i poza nim. Najbardziej dostępną formą zdobywania lub podnoszenia kwalifikacji jest edukacja nieformalna, czyli uczenie się od współpracowników, znajomych i rodziny lub samodzielnie. W opracowaniu zamieszczono również dane dotyczące m.in. dostępu do informacji o możliwościach kształcenia, trudności ograniczających uczestniczenie w edukacji, czasu poświęcanego na naukę i znajomości języków obcych.

W czerwcu br. ukazały się ponadto:

- *Badanie Aktywności Ekonomicznej Ludności (folder dla rodzin biorących udział w Badaniu Aktywności Ekonomicznej Ludności) – 4 kwartał 2023*;

© Główny Urząd Statystyczny
Artykuł udostępniony na licencji CC BY-SA 4.0 / Article available under the CC BY-SA 4.0 licence

- *Bezrobocie rejestrowane 1 kwartał 2024 r.*;
- *Bilans przepływów międzygałęziowych w bieżących cenach bazowych w 2020 r.*;
- „Biuletyn statystyczny" nr 5/2024;
- *Ceny robót budowlano-montażowych i obiektów budowlanych (kwiecień 2024 r.)*;
- *Efekty działalności budowlanej w 2023 r.*;
- *Efektywność wykorzystania energii w latach 2012–2022*;
- *Energia 2024*;
- *Koniunktura w przetwórstwie przemysłowym, budownictwie, handlu i usługach 2000–2024 (czerwiec 2024). Z pogłębioną analizą aktualnych zagadnień gospodarczych*;
- *Nakłady i wyniki przemysłu – 1 kwartał 2024 r.*;
- *Popyt na pracę w 2023 r.*;
- *Produkcja ważniejszych wyrobów przemysłowych w maju 2024 r.*;
- *Przemysł – wyniki działalności w 2023 r.*;
- *Rachunki kwartalne produktu krajowego brutto w latach 2019–2023*;
- *Raport GUS 2023*;
- *Równoległy oraz wyprzedzający zagregowany wskaźnik koniunktury, zegar koniunktury – szereg do kwietnia 2024 r.*;
- „Statistics in Transition new series" nr 2/2024;
- *Sytuacja społeczno-gospodarcza kraju w maju 2024 r.*;
- *Sytuacja społeczno-gospodarcza województw Nr 1/2024*;
- *Turystyka w 2023 r.*;
- *Warunki pracy w 2023 r.*;
- *Wybrane wskaźniki przedsiębiorczości w latach 2018–2022*;
- *Zatrudnienie i wynagrodzenia w gospodarce narodowej w pierwszym kwartale 2024 r.*;
- *Zeszyt metodologiczny. Popyt na pracę.*

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 przesyłać 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 niezłożeniu 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 przeredagowania 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 korekcie).

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

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 w Departamencie Opracowań Statystycznych 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 opisania 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

sympozjum naukowego, 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 w Departamencie Opracowań Statystycznych 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) w Departamencie Opracowań Statystycznych 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 jakiekolwiek 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 opóźniać 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 jakiejkolwiek 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 formacie) 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 i 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 zwięzłe 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 łacińskim. 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 wyliczeniach należy posłużyć się listą punktowaną z punktorami 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. $\boldsymbol{a}$, $\boldsymbol{A}$, $\boldsymbol{w}$, $\boldsymbol{y}(x)$, $\boldsymbol{w}_i$); macierzy – pismem prostym i pogrubionym (np. $\mathbf{A}$, $\mathbf{a}$, $\mathbf{M}$, $\mathbf{Y}(x)$, $\mathbf{M}_i$).
19. Objaśnienia 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 jakiekolwiek 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. Przywoływanie źródeł w artykułach napisanych w języku polskim

4.3.1. Ogólne zasady APA

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)
Przypadek niepełnych danych bibliograficznych		
Brak ustalonego autorstwa	(*Skrócony tytuł* …, 2015)	*Pełny tytuł* (2015)
Brak roku wydania	(Iksiński, b.r.)	Iksiński (b.r.)
Inne przypadki		
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: Zieniecka, 2007)	Nowakowski (1990, za: Zieniecka, 2007)

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

4.3.2. Szczegółowe wewnętrzne zasady „WS"

4.3.2.1. Adresy portali internetowych, w tym baz danych Głównego Urzędu Statystycznego

Adresy portali internetowych, które są przywoływane w artykule jedynie w celach informacyjnych, należy umieszczać w przypisach dolnych.

W przypadku korzystania z danych pobranych z baz Głównego Urzędu Statystycznego prosimy o podanie w miejscu, w którym baza jest przywoływana po raz pierwszy, pełnej nazwy bazy i jej skrótu (jeśli istnieje), nazwy jej właściciela oraz adresu internetowego w przypisie dolnym. W kolejnych przywołaniach, np. w źródle pod wykresem, należy posługiwać się już tylko pełną lub skróconą nazwą bazy.

Przykłady baz danych GUS	
pierwsze przywołanie	kolejne przywołania
Bank Danych Lokalnych (BDL) Głównego Urzędu Statystycznego + link podany w przypisie dolnym: https://bdl.stat.gov.pl	BDL
Baza Demografia Głównego Urzędu Statystycznego + link podany w przypisie dolnym: https://demografia.stat.gov.pl	Baza Demografia
Dziedzinowe Bazy Wiedzy (DBW) Głównego Urzędu Statystycznego + link podany w przypisie dolnym: https://dbw.stat.gov.pl	DBW

4.3.2.2. Akty prawne

Jeśli autor powołuje się w pracy na akty prawne, powinien za pierwszym razem podać ich pełny oficjalny tytuł; przy kolejnych przywołaniach najczęściej wystarczy nazwa skrócona. W przypadku aktów prawnych zapisanych w innym alfabecie niż łaciński tytuł trzeba poddać transkrypcji. (Informacje dotyczące miejsca publikacji aktu prawnego, takie jak numer dziennika urzędowego, należy podać tylko w opisie zamieszczonym w bibliografii załącznikowej).

Przykłady aktów prawnych	
pierwsze przywołanie	kolejne przywołania
Ustawa z dnia 29 czerwca 1995 r. o statystyce publicznej (dalej: ustawa o statystyce publicznej)	ustawa o statystyce publicznej
Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 1260/2013 z dnia 20 listopada 2013 r. w sprawie statystyk europejskich w dziedzinie demografii (dalej: rozporządzenie w sprawie statystyk europejskich)	Rozporządzenie w sprawie statystyk europejskich
Statistics Act	Statistics Act

4.4. Bibliografia załącznikowa w artykułach napisanych w języku polskim

4.4.1. Zasady ogólne

Bibliografia powinna być zamieszczona na końcu opracowania.

Źródła należy uszeregować alfabetycznie według nazwiska pierwszego autora, a w przypadku dwóch lub więcej prac tego samego autora – chronologicznie według roku publikacji. Prace bez znanego roku publikacji (oznaczone „b.r.") występują przed pracami ze znanym rokiem publikacji. Jeśli kilka prac tego samego autora zostało opublikowanych w tym samym roku, należy podać je w kolejności alfabetycznej według tytułu i odpowiednio oznaczyć literami a, b, c itd.

Opis bibliograficzny materiałów dostępnych w internecie powinien zawierać link prowadzący do źródłowej strony internetowej lub link DOI. Nie należy podawać linków prowadzących do baz czasopism czy repozytoriów.

4.4.2. Przykłady opisów bibliograficznych

Typ źródła	Przykład opisu bibliograficznego
Artykuł w czasopiśmie	
W wersji: drukowanej	Nazwisko, X. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik* (zeszyt), strona początku–strona końca.
elektronicznej, z DOI	Nazwisko, X., Nazwisko 2, Y. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*(zeszyt), strona początku–strona końca. https://doi.org/xxx.
elektronicznej, bez DOI	Nazwisko, X., Nazwisko 2, Y., Nazwisko 3, Z. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*(zeszyt), strona początku–strona końca. https://xxx.
Opublikowany w trybie online first	Nazwisko, X. (rok). Tytuł artykułu. *Tytuł czasopisma.* Opublikowany w trybie online first. https://xxx.
Artykuł w gazecie codziennej	
W wersji: drukowanej	Nazwisko, X. (rok, miesiąc i dzień). Tytuł artykułu. *Tytuł czasopisma*, strona lub strona początku–strona końca.
elektronicznej	Nazwisko, X. (rok, miesiąc i dzień). Tytuł artykułu. *Tytuł czasopisma.* https://xxx. Nazwisko, X. (b.r.). Tytuł artykułu. *Tytuł czasopisma.* https://xxx. Tytuł artykułu. (rok, miesiąc i dzień). *Tytuł czasopisma.* https://xxx.
Książka	
W wersji: drukowanej	Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo.
elektronicznej, z DOI	Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo. https://doi.org/xxx.
elektronicznej, bez DOI	Nazwisko, X. (rok). *Tytuł książki*. Wydawnictwo. https://xxx.
W przekładzie	Nazwisko, X. (rok). *Tytuł książki* (tłum. Y. Nazwisko). Wydawnictwo.
Wydanie wielotomowe: tom zatytułowany	Nazwisko, X. (rok). *Tytuł książki: numer tomu. Tytuł tomu*. Wydawnictwo.
tom niezatytułowany	Nazwisko, X. (rok). *Tytuł książki* (numer tomu). Wydawnictwo.
Kolejne wydanie	Nazwisko, X. (rok). *Tytuł książki* (numer wydania). Wydawnictwo.
Pod redakcją (niezależnie od języka, w którym książka została wydana)	Nazwisko, X. (red.). (rok). *Tytuł książki*. Wydawnictwo.
Rozdział i hasło słownikowe/encyklopedyczne	
Rozdział w pracy zbiorowej	Nazwisko, X. (rok). Tytuł rozdziału. W: Y. Nazwisko, Z. Nazwisko 2 (red.), *Tytuł książki* (s. strona początku–strona końca). Wydawnictwo. https://doi.org/xxx lub https://xxx.
Hasło ze słownika lub z encyklopedii w wersji: drukowanej	Nazwisko autora hasła, X. (rok). Hasło. W: Y. Nazwisko (red.), *Tytuł*. Wydawnictwo. Hasło. (rok). W: Y. Nazwisko (red.), *Tytuł*. Wydawnictwo.
elektronicznej	Hasło. (rok, miesiąc i dzień lub „b.r."). W: *Tytuł* (np. *Wikipedia* lub *Słownik języka polskiego PWN*). https://xxx.
Raporty i szara literatura	
Autor: indywidualny	Nazwisko, X. (rok). *Tytuł raportu*. Wydawnictwo. https://doi.org/xxx lub https://xxx.
instytucjonalny	Nazwa instytucji. (rok). *Tytuł raportu.* Wydawnictwo (tylko jeśli wydawcą jest inna instytucja niż instytucja autorska). https://doi.org/xxx lub https://xxx.
Working papers	Nazwisko, X. (rok). *Tytuł pracy* (nazwa serii i numer). https://doi.org/xxx lub https://xxx.

Typ źródła	Przykład opisu bibliograficznego
Materiały z konferencji	
Opublikowane jako: druk zwarty	zob. przykład opisu książki lub rozdziału
druk ciągły	zob. przykład opisu artykułu w czasopiśmie
Niepublikowane (jedynie wygłoszone)	Nazwisko, X. (rok, dzień i miesiąc). *Tytuł pracy* [typ wystąpienia, np. referat lub prezentacja]. Nazwa i miejsce (miasto, kraj) konferencji.
Rozprawa doktorska	
Niepublikowana	Nazwisko, X. (rok). *Tytuł pracy* [niepublikowana rozprawa doktorska]. Nazwa instytucji nadającej tytuł doktorski.
Opublikowana, dostępna w internecie	Nazwisko, X. (rok). *Tytuł pracy* [rozprawa doktorska, nazwa instytucji nadającej tytuł doktorski]. https://xxx.
Maszynopis	
Niepublikowany / przygotowywany przez autora / zgłoszony do publikacji, ale jeszcze niezaakceptowany	Nazwisko, X. (rok). *Tytuł* [maszynopis niepublikowany / w przygotowaniu / zgłoszony do publikacji].
Artykuł zaakceptowany do publikacji w czasopiśmie	Nazwisko, X. (w druku). Tytuł artykułu. *Tytuł czasopisma.*
Opublikowany nieformalnie (np. na stronie internetowej autora)	Nazwisko, X., Nazwisko 2, Y. (rok). *Tytuł*. https://xxx.
Akt prawny[a]	
Polski i UE	Pełny tytuł aktu prawnego wraz z numerem/pozycją w dzienniku urzędowym.
Inny	Pełny tytuł aktu prawnego w języku oryginalnym (w przypadku zapisu w innym alfabecie niż łaciński należy podać tylko transkrypcję) wraz z numerem/pozycją w dzienniku urzędowym. https://xxx.
Tekst na stronie internetowej (dostępny tylko online)	
Znana data publikacji, zawartość strony się nie zmienia (jest archiwizowana)	Nazwisko, X. (rok, dzień i miesiąc). *Tytuł*. https://xxx.
Nieznana data publikacji, zawartość strony się zmienia (nie jest archiwizowana)	Nazwa instytucji. (b.r.). *Tytuł*. Pobrane dzień, miesiąc i rok pobrania z https://xxx.
Zbiór danych	
Dane opublikowane: znana data publikacji, zawartość zbioru się nie zmienia (jest archiwizowana)	Nazwisko, X. (rok). *Nazwa zbioru danych* [zbiór danych]. Wydawca. https://xxx.
nieznana data publikacji, zawartość zbioru się zmienia (nie jest archiwizowana)	Nazwa instytucji. (b.r.). *Nazwa zbioru danych* [zbiór danych]. Wydawca (tylko jeśli wydawcą jest inna instytucja niż instytucja autorska / właściciel danych). Pobrane dzień, miesiąc i rok pobrania z https://xxx.
Materiały audiowizualne	
Nagranie wideo	Nazwisko, X. (rok, dzień i miesiąc). *Tytuł* [wideo]. Nazwa kanału, na którym nagranie zostało udostępnione (np. YouTube). https://xxx.
Webinar	Nazwisko, X. (rok, dzień i miesiąc). *Tytuł* [webinar]. Nazwa instytucji. https://xxx.

[a] Wewnętrzne zasady „WS".

Typ źródła	Przykład opisu bibliograficznego
Posty w serwisach społecznościowych	
Post na portalu X lub Instagramie	Nazwisko, X. lub nazwa instytucji [@nazwa użytkownika] (rok, dzień i miesiąc). *Treść – do 20 wyrazów* [post]. Nazwa serwisu społecznościowego (X lub Instagram). https://xxx.
Post na Facebooku	Nazwisko, X. lub nazwa instytucji (rok, dzień i miesiąc). *Treść – do 20 wyrazów* [post]. Facebook. https://xxx. Nazwa instytucji [nazwa użytkownika] (rok, dzień i miesiąc). *Treść – do 20 wyrazów* [post]. Facebook. https://xxx.

Źródło: opracowanie na podstawie: American Psychological Association. (2020). *Publication manual of the American Psychological Association* (wyd. 7). 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.

STAŁE DZIAŁY „WS" – ZAKRES TEMATYCZNY
PERMANENT SECTIONS OF *WS* – THEMATIC SCOPE

Tematy artykułów	Topics of the articles
Studia metodologiczne / Methodological studies	
• Oryginalne lub udoskonalone rozwiązania metodologiczne, które mogą znaleźć zastosowanie w analizach statystycznych i służyć podnoszeniu ich jakości • Projektowanie badań statystycznych	• Original or developed methodological solutions which can be applied to statistical analyses and serve to improve their quality • Planning statistical surveys
Statystyka w praktyce / Statistics in practice	
• Nowatorskie zastosowania narzędzi i modeli statystycznych oraz analiza i ocena statystyczna zjawisk społeczno-gospodarczych i innych, prowadzona w szczególności na danych pochodzących z zasobów statystyki publicznej • Wykorzystanie narzędzi informatycznych do uzyskiwania i przetwarzania informacji statystycznych, naliczania i kontroli ujawniania danych oraz prezentacji i rozpowszechniania danych wynikowych	• Innovative applications of statistical tools and models as well as statistical analysis and assessment of social, economic and other phenomena, performed mainly on data produced by official statistics • Application of IT tools to obtain and process statistical information, to calculate data and control the statistical disclosure, and to present and disseminate output data
Studia interdyscyplinarne. Wyzwania badawcze / Interdisciplinary studies. Research challenges	
• Wyzwania badawcze wynikające z rosnących potrzeb użytkowników danych statystycznych i wymagające stosowania rozwiązań z różnych dziedzin nauki • Problematyka wykraczająca poza konwencjonalne tematy związane ze statystyką • Wyniki badań prowadzonych w obrębie różnych dyscyplin z wykorzystaniem metod statystycznych	• Research challenges resulting from growing needs of statistical data users and requiring the application of solutions from various fields of science • Problems beyond the conventional thematic scope related to statistics • Results of research carried out in the framework of several fields of science using statistical methods
Edukacja statystyczna / Statistical education	
• Metody i efekty nauczania statystyki na wszystkich poziomach edukacji • Popularyzacja myślenia statystycznego i rzetelnego posługiwania się informacjami statystycznymi	• Methods and effects of statistical education at all levels of education • Popularisation of statistical thinking and of diligent use of statistical information
Spisy powszechne – problemy i wyzwania / Issues and challenges in census taking	
• Rozwiązania metodologiczne i organizacyjne możliwe do zastosowania podczas przygotowywania i prowadzenia spisów • Praktyczne aspekty związane z gromadzeniem i udostępnianiem danych ze spisów, w tym dotyczące obciążenia odpowiedzi i ochrony tajemnicy statystycznej	• Methodological and organisational solutions which may be implemented in the process of preparing and conducting censuses • Practical aspects of collecting and disseminating census data, including those related to response burden and the protection of statistical confidentiality
Z dziejów statystyki / From the history of statistics	
• Historia prowadzenia obserwacji statystycznych, w tym rozwój metodologii i narzędzi oraz instytucji statystycznych w Polsce i za granicą • Życie i osiągnięcia wybitnych statystyków	• History of statistical observations, including the development of statistical methodologies, tools and institutions in Poland and abroad • Life and achievements of prominent statisticians
In memoriam	
• Nekrologi i artykuły wspomnieniowe o osobach zasłużonych dla statystyki	• Obituaries and articles remembering important people in the world of statistics
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information	
• Dyskusje i polemiki • Sprawozdania z konferencji naukowych • Recenzje książek oraz zestawienia nowości wydawniczych GUS	• Discussions and polemics • Reports from scientific conferences • Book reviews and compilations of Statistics Poland's new publications