Cena 13,00 zł
(VAT 8%)

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

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

SIERPIEŃ / AUGUST
ROK / VOLUME 65

2020 | 8

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

POLSKIE TOWARZYSTWO STATYSTYCZNE
POLISH STATISTICAL ASSOCIATION

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

SIERPIEŃ / AUGUST
ROK / VOLUME 65

2020 | 8 (711)

ADRES REDAKCJI / EDITORIAL OFFICE ADDRESS

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

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

Redakcja techniczna, skład i łamanie, wykresy, korekta: Zakład Wydawnictw Statystycznych – zespół pod kierunkiem Wojciecha Szuchty
Technical editing, typesetting, figures, proof-reading: Statistical Publishing Establishment – team supervised by Wojciech Szuchta

Zakład Wydawnictw Statystycznych

Druk i oprawa / Printed and bound:
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 original version of the journal is the electronic issue, available at ws.stat.gov.pl

Indeks 381306

Informacje w sprawie nabywania czasopism / Information on purchasing of the journal:
Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment
tel./phone +48 22 608 32 10, +48 22 608 38 10
Prenumerata jest prowadzona przez / Subscription is available at RUCH S.A.
Zamówienia na prenumeratę można składać na stronie / Subscriptions can be ordered at
www.prenumerata.ruch.com.pl

Zam. 236/2020 – nakład 320 egz.

SPIS TREŚCI
CONTENTS

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 8, 4–5
Od redakcji / From the editorial team

OD REDAKCJI

W sierpniowym wydaniu „Wiadomości Statystycznych. The Polish Statistician" inaugurujemy dział Spisy powszechne – problemy i wyzwania, poświęcony jednemu z największych przedsięwzięć podejmowanych w ramach statystyki publicznej. Prowadzenie spisów, podobnie jak innych badań, wymaga zmierzenia się z wyzwaniami współczesności, a także – co pokazały ostatnie miesiące – z nieprzewidywalnymi zdarzeniami losowymi. W związku ze zbliżającą się rundą spisów pragniemy wyeksponować prace zawierające propozycje efektywnych rozwiązań, zarówno organizacyjnych, jak i metodologicznych, dotyczących spisów powszechnych. Przyświeca nam myśl, aby wymiana doświadczeń w gronie statystyków i badaczy przyczyniła się do zwiększenia skuteczności i rozwoju systemów statystycznych.

W tę problematykę wprowadza pierwsze opracowanie zamieszczone w nowym dziale – *Organizacja Powszechnego Spisu Rolnego w 2020 r.* Autorzy, dr inż. Janusz Dygaszewicz i Magdalena Janczur-Knapek, omawiają planową organizację tegorocznego spisu oraz kluczowe zmiany organizacyjne, techniczne i metodologiczne wprowadzone po wybuchu pandemii Covid-19. W pracy poruszono zagadnienia związane z organizacją zbierania danych, zadaniami rachmistrzów spisowych, zastosowaną architekturą IT i popularyzacją badania (m.in. za pośrednictwem internetu i mediów społecznościowych).

Pozostałe artykuły w bieżącym numerze przedstawiają propozycje rozwiązań metodologicznych i wyniki badań z obszaru praktycznych zastosowań statystyki.

Artykuł *Estymacja pośrednia wskaźników ubóstwa na poziomie powiatów* dr. Łukasza Wawrowskiego dotyczy estymacji stopy ubóstwa i głębokości ubóstwa z zastosowaniem empirycznej metody bayesowskiej (w tym przypadku liniowego modelu mieszanego) i symulacji Monte Carlo. Autor wykorzystał dane z Europejskiego Badania Dochodów i Warunków Życia przeprowadzonego w 2011 r. oraz Narodowego Spisu Powszechnego Ludności i Mieszkań 2011. Oszacowane w ten sposób wskaźniki pozwalają na ocenę zróżnicowania ubóstwa na poziomie lokalnym i co ważne, cechują się większą precyzją i zbieżnością z rejestrami administracyjnymi niż w przypadku zastosowania estymacji bezpośredniej.

Analiza rozbieżności danych lustrzanych dotyczących masy towarów w statystykach handlu wewnątrzunijnego jest przedmiotem artykułu dr hab. Iwony Markowicz i dr. Pawła Barana. W badaniu omawianym w tej pracy do oceny jakości danych wykorzystano dane dotyczące nie tylko wartości, lecz także masy towarów. Na podstawie informacji za 2017 r., pochodzących z bazy Eurostatu Comext, oraz uwzględniając dynamikę w latach: 2005, 2008, 2011 i 2014, wykazano, że udział wywozu towarów z Polski do danego kraju na obszarze Unii Europejskiej ogółem różni się dla danych wyrażonych wartościowo (wartość towarów) i ilościowo (masa towarów). Autorzy postulują więc, aby w badaniu jakości danych lustrzanych stosować oba podejścia.

Dr inż. Jolanta Wojnar w artykule *Zróżnicowanie wykorzystania technologii informacyjno-komunikacyjnych w krajach Unii Europejskiej* analizuje 15 wybranych wskaźników opisujących wykorzystanie technologii informacyjno-komunikacyjnych (ICT) przez osoby fizyczne i gospodarstwa domowe. Autorka posłużyła się danymi ze sprawozdań Głównego Urzędu Statystycznego i danymi Eurostatu za 2017 r.; zastosowała metodę składowych głównych i analizę skupień z użyciem metody *k*-średnich. Liderami w zakresie wykorzystania ICT okazały się kraje skandynawskie i kraje Beneluksu, najniżej oceniono zaś kraje południowej i południowo-wschodniej Europy oraz Polskę.

Na zakończenie sierpniowego wydania prezentujemy wyniki XLIX Ogólnopolskiego Konkursu Statystycznego, którego przebieg w wyjątkowych okolicznościach, wynikających z ogłoszenia stanu epidemii, oraz poziom nadesłanych prac omawia dr Bożena Łazowska. Wybrane nowości wydawnicze GUS tradycyjnie poleca Justyna Gustyn.

Zapraszamy do lektury.

FROM THE EDITORIAL TEAM

The August issue of *Wiadomości Statystyczne. The Polish Statistician* introduces a new section of the journal entitled Issues and challenges in census taking, dedicated to one of the major projects undertaken by official statistics. Conducting censuses, like any other type of research, requires us to respond to the challenges of today's world and also – as recent months have shown – to unforeseeable random events. As a round of censuses is nearing, we would like to highlight articles which include propositions of effective organisational and methodological solutions relating to censuses. Our hope is that the exchange of experiences among statisticians and researchers will enhance the effectiveness and development of statistical systems.

The first article of the new section introducing the aforementioned topic is entitled *The organisation of the National Agricultural Census 2020*. The authors, Janusz Dygaszewicz, BEng, PhD, and Magdalena Janczur-Knapek, discuss the planned organisation of this year's census and the key organisational, technical, and methodological changes introduced following the outbreak of the COVID-19 pandemic. The article describes issues relating to the organisation of data collection, tasks performed by census enumerators, the applied IT architecture, and the promotion of the census through, among others, the Internet and social media.

The remaining articles in this issue present some propositions of methodological solutions and the results of research from within the scope of practical applications of statistics.

The article *Indirect estimation of poverty indicators at poviat level* by Łukasz Wawrowski, PhD, describes the estimation of the poverty rate and depth of poverty with the application of the Empirical Bayes method (in this case of a linear mixed model) and Monte Carlo simulations. The author used data from the European Union Statistics on Income and Living Conditions of 2011 and the National Census of Population and Housing 2011. The indicators estimated this way allow for an assessment of the diversity of poverty at a local level and, what is important, they are more precise and consistent with administrative registers than if direct estimation were applied.

Iwona Markowicz, PhD, DSc, and Paweł Baran, PhD, present an *Analysis of discrepancies in mirror data relating to the weight of goods in intra-EU trade statistics* in their article. The described research used data relating not only to the quality, but also to the weight of goods in the process of assessing data quality. The information obtained from Eurostat's Comext database relating to the year 2017 and the dynamics in the years 2005, 2008, 2011, and 2014 indicated that the share of Polish total export to a given country from within the European Union is different for data expressed in value (value of goods) and in quantity (weight of goods). Consequently, the authors recommend using both approaches when studying the quality of mirror data.

Jolanta Wojnar, BEng, PhD, in her article entitled *Diversity in the use of information and communication technologies among European Union countries* analyses 15 selected indicators describing the application of the information and communication technologies (ICT) by natural persons and households. The author uses data from Statistics Poland reports and Eurostat covering the year 2017. The method of principal components analysis was applied and a cluster analysis based on the *k*-means method was performed. The Scandinavian and Benelux countries were the leaders in ICT usage, while countries of southern and south-eastern Europe, jointly with Poland, were the lowest rated.

The August issue closes with a presentation of the results of the 49th Polish Nationwide Statistical Competition. Bożena Łazowska, PhD, discusses the exceptional course of the competition, resulting from the announcement of a state of epidemic, and the level of the submitted papers. Justyna Gustyn, as always, recommends a selection of new publications by Statistics Poland.

We wish you pleasant reading.

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 8, 7–26 DOI: 10.5604/01.3001.0014.3524
Studia metodologiczne / Methodological studies

Estymacja pośrednia wskaźników ubóstwa na poziomie powiatów

Łukasz Wawrowski[a]

Streszczenie. Dysponowanie szczegółowymi i precyzyjnymi danymi na temat ubóstwa na niskim poziomie agregacji przestrzennej jest ważne dla prowadzenia skutecznej polityki spójności. W Polsce tego typu informacje są gromadzone w ramach badań gospodarstw domowych, prowadzonych przez Główny Urząd Statystyczny, i udostępniane na poziomie kraju, regionów i wybranych grup społeczno-ekonomicznych. Oszacowania bezpośrednie w domenach, których badanie nie obejmuje, są obarczone dużym błędem szacunku. W sytuacji ograniczonej, w skrajnym przypadku zerowej, liczebności próby estymację umożliwia zastosowanie metod statystyki małych obszarów – estymacji pośredniej. Techniki te wykorzystują cechy silnie skorelowane z badanym zjawiskiem, pochodzące ze spisu powszechnego lub z rejestru administracyjnego.

Celem badania omawianego w artykule jest estymacja dwóch wskaźników: stopy ubóstwa i głębokości ubóstwa na poziomie powiatów, z zastosowaniem empirycznej metody bayesowskiej (EB). Pierwszy wskaźnik informuje o skali zjawiska, a drugi – o jego intensywności, więc są one komplementarnymi miarami ubóstwa. W badaniu wykorzystano dane z Europejskiego Badania Dochodów i Warunków Życia przeprowadzonego w 2011 r. oraz Narodowego Spisu Powszechnego Ludności i Mieszkań 2011. Za pomocą metody EB, bazującej na liniowym modelu mieszanym i symulacjach Monte Carlo, uzyskano informacje o wielkości i intensywności ubóstwa na poziomie powiatów. Oszacowane w ten sposób wskaźniki pozwalają na ocenę zróżnicowania ubóstwa na poziomie lokalnym. Ponadto cechują się większą precyzją i zbieżnością z rejestrami administracyjnymi w porównaniu do rezultatów estymacji bezpośredniej.

Słowa kluczowe: ubóstwo, estymacja pośrednia, empiryczna metoda bayesowska
JEL: C13, C51, I32

Indirect estimation of poverty indicators at poviat level

Abstract. The availability of detailed and precise data on poverty at a low level of spatial aggregation is important when pursuing an effective cohesion policy. In Poland, this type of information is gathered during household surveys conducted by Statistics Poland and is made available at country, region, and selected socio-economic group level. Direct estimates relating to domains not included in a survey are burdened with a serious estimation error. In a situation of a limited (or in extreme cases zero) sample size, an estimation becomes possible through the application of small area estimation methods – indirect estimation. These techniques use variables which are strongly correlated with the researched phenomenon and which come from a census or from an administrative register.

The aim of the study discussed in the article is to estimate two indicators: the rate of poverty and the depth of poverty at a poviat level, with the application of the Empirical Bayes (EB) method. The first indicator provides information on the scale of the phenomenon and the other one on its intensity, and so they constitute complementary measures of poverty. The study used data from the European Union Statistics on Income and Living Conditions of 2011 and the National Census of Population and Housing 2011. Information about the scale and intensity of poverty at the poviat level was obtained through the adaptation of the EB method based on the linear mixed model and Monte Carlo simulations. The indicators estimated this way allow for an assessment of the diversity of poverty at a local level. In addition, they are more precise and consistent with administrative registers in comparison to direct estimation results.

Keywords: poverty, small area estimation, Empirical Bayes method

[a] Uniwersytet Ekonomiczny w Poznaniu, Instytut Informatyki i Ekonomii Ilościowej.
ORCID: https://orcid.org/0000-0002-1201-5344.

1. Wprowadzenie

Badania empiryczne prowadzone w Polsce, przede wszystkim badania reprezentacyjne Głównego Urzędu Statystycznego (GUS): badanie budżetów gospodarstw domowych (BBGD) oraz Europejskie Badanie Dochodów i Warunków Życia (EU-SILC), dostarczają informacji na temat ubóstwa na bardzo ogólnym poziomie. Estymacja na niższym poziomie agregacji przestrzennej z wykorzystaniem klasycznych metod nie jest możliwa ze względu na duże błędy takich oszacowań.

Dostęp do szczegółowych danych jest podstawą funkcjonowania społeczeństwa informacyjnego. Obserwuje się zapotrzebowanie na dane, które mogłyby zostać wykorzystane m.in. przez władze samorządowe do oceny i planowania polityki społecznej. Na podstawie wyników BBGD i EU-SILC można stwierdzić, że sytuacja dochodowa gospodarstw domowych w Polsce poprawia się z roku na rok. Przeciętny miesięczny dochód rozporządzalny *per capita* w 2009 r. wynosił 1114 zł, a w 2019 r. – 1819 zł, co oznacza wzrost o 63% (GUS, 2020). Stopa ubóstwa szacowana na podstawie EU-SILC w 2018 r. wynosiła 14,8%, podczas gdy 10 lat wcześniej – 16,9%. Obserwuje się także coraz mniejsze nierówności dochodowe. W 2008 r. wartość współczynnika Giniego kształtowała się na poziomie 32,0, a w 2018 r. było to 27,8 (GUS, 2019). Wymienione wskaźniki dotyczą całego kraju; sytuacja dochodowa gospodarstw domowych na poziomie lokalnym może być bardziej zróżnicowana.

Rozszerzenie pokrycia informacyjnego bez konieczności podnoszenia kosztów badania, którą spowodowałoby m.in. zwiększenie wielkości próby, jest możliwe dzięki zastosowaniu metod statystyki małych obszarów – estymacji pośredniej. Umożliwiają one estymację parametrów w przypadku ograniczonej, a nawet zerowej liczebności próby, przy wykorzystaniu wszystkich dostępnych źródeł danych (Żądło, 2015), z tym że wymagają identyfikacji zbiorów danych stanowiących odpowiednie źródło zmiennych pomocniczych w estymacji pośredniej. Cechy te powinny być silnie skorelowane z analizowanym zjawiskiem, a zatem w przypadku analizy ubóstwa poszukiwane są jego ilościowe determinanty (Panek, 2011). Oprócz badań nad ubóstwem (Chambers i Tzavidis, 2006; Molina i Rao, 2010; Wawrowski, 2016) metody statystyki małych obszarów znajdują zastosowanie m.in. w szacowaniu charakterystyk rynku pracy (Gołata, 2004; Wilak, 2014) oraz w statystyce gospodarczej (Chandra, Chambers i Salvati, 2012; Dehnel, 2010).

Badania nad wykorzystaniem metod estymacji pośredniej poziomu ubóstwa są prowadzone także w ramach krajowych i międzynarodowych projektów badawczych. W latach 2013 i 2014 GUS wraz z Bankiem Światowym realizowały projekt *Mapy ubóstwa na poziomie podregionów w Polsce z wykorzystaniem estymacji pośredniej* (Szymkowiak, Młodak i Wawrowski, 2017). Jego bezpośrednim rezultatem było opracowanie raportu *Pomiar ubóstwa na poziomie powiatów (LAU 1)* (Szym-

kowiak, 2015). Estymatorami statystyki małych obszarów posłużono się do oszacowania wartości stopy ubóstwa dla lat 2005, 2008 i 2011. W badaniu omawianym w niniejszym artykule przyjęto szersze podejście – włączono dodatkowy wskaźnik ubóstwa, diagnostykę modelu oraz ocenę wyników. Zastosowano także inną metodę transformacji zmiennej zależnej, z uwzględnieniem aktualnego stanu badań (Rojas-Perilla, Pannier, Schmid i Tzavidis, 2020).

Celem badania omawianego w artykule jest estymacja dwóch wskaźników: stopy ubóstwa i głębokości ubóstwa na poziomie powiatów, z zastosowaniem empirycznej metody bayesowskiej (Molina i Rao, 2010), bazującej na liniowym modelu mieszanym i symulacjach Monte Carlo. Dane wykorzystane w analizie pochodzą z badania EU-SILC przeprowadzonego w 2011 r. oraz Narodowego Powszechnego Spisu Ludności i Mieszkań (NSP) 2011. Otrzymane wyniki poddano wielopłaszczyznowej ocenie z użyciem danych z rejestrów administracyjnych oraz miar precyzji i autokorelacji przestrzennej.

2. Pomiar ubóstwa

Ubóstwo jest uznawane za jeden z najważniejszych i najbardziej złożonych problemów społecznych współczesnego świata. Występuje w każdym społeczeństwie. Kluczowy w ocenie tego zjawiska jest jego pomiar. Nadanie mu uniwersalnego charakteru wymagałoby określenia jednolitych zasad i definicji. Rada Europejskiej Wspólnoty Gospodarczej (WE) w dokumencie z 19 grudnia 1984 r. zdefiniowała je następująco: „ubóstwo odnosi się do osób, rodzin lub grup osób, których zasoby (materialne, kulturowe i społeczne) są ograniczone w takim stopniu, że poziom ich życia obniża się poza akceptowalne minimum w kraju zamieszkania" (EEC, 1985).

W powyższej definicji kluczowym aspektem jest ustalenie kryterium przynależności do sfery ubóstwa. Można zastosować kryterium jedno- lub wielowymiarowe. Za pomocą tego pierwszego ocenia się stopień zaspokojenia potrzeb jednostki w odniesieniu do dochodów (wydatków) wyrażonych w formie monetarnej. W literaturze przedmiotu podkreśla się jednak niedoskonałość tego podejścia i wyraża się przekonanie, że zjawisko ubóstwa nie jest jednowymiarowe (GUS i US w Łodzi, 2013a; Panek, 2010). Wielu badaczy postulowało uwzględnienie w jego analizie także czynników pozamonetarnych, m.in. dochodów i zasobów materialnych nagromadzonych we wcześniejszych okresach, warunków mieszkaniowych, edukacji oraz zasobów zawodowych i finansowych (Panek, 2011; Ulman i Ćwiek, 2020).

Podejście jednowymiarowe wymaga określenia poziomu dochodów (wydatków), poniżej którego osoba jest uznawana za ubogą. W Polsce zdefiniowano kilka rodzajów granicy (linii) ubóstwa. Instytut Pracy i Spraw Socjalnych (IPiSS) wyznacza wartości minimum egzystencji (granicę wydatków pozwalających na skromne wyżywienie oraz utrzymanie małego mieszkania) i minimum socjalnego (granicę wydatków

umożliwiającą minimalnie godziwy standard życia), a także bierze udział w konsultacji ustawowej granicy ubóstwa (kwoty dochodu, która uprawnia do ubiegania się o przyznanie świadczenia z systemu pomocy społecznej) (Kurowski, 2002). W badaniach reprezentacyjnych granicę ubóstwa ustala się jako stałą część mediany lub średniej arytmetycznej rozkładu dochodów w całej populacji. Główny Urząd Statystyczny w BBGD jako linię ubóstwa przyjmuje 50% średnich wydatków. Natomiast w EU-SILC określono, za rekomendacją Eurostatu, że granicę ubóstwa stanowi 60% mediany rozkładu dochodów ekwiwalentnych (uwzględniających skład demograficzny gospodarstwa domowego). Tę definicję przyjęto także w niniejszej pracy.

2.1. Wskaźniki ubóstwa

Pomiar ubóstwa bazuje na dwóch podstawowych wskaźnikach: stopie ubóstwa i głębokości ubóstwa. Pierwsza miara, informująca o odsetku ubogich w populacji, jest prosta w budowie i interpretacji, ale niepozbawiona wad. Przede wszystkim nie uwzględnia intensywności ubóstwa. Osoby znajdujące się poniżej granicy ubóstwa mogą mieć dochody bardzo zbliżone do tej granicy lub bardzo od niej oddalone, ale w obu przypadkach stopa ubóstwa będzie taka sama, mimo że zjawisko jest bardziej nasilone w drugiej sytuacji. Intensywność tę odzwierciedla głębokość ubóstwa, która informuje o poziomie ubóstwa wśród osób znajdujących się poniżej linii ubóstwa. Oparcie analizy na obu wymienionych wskaźnikach daje więc pełniejszy obraz zjawiska (Haughton i Khandker, 2009).

Na podstawie wartości dochodów ekwiwalentnych można ustalić odpowiednie miary ubóstwa. Ogólny wzór na wskaźniki ubóstwa wyznaczyli Foster, Greer i Thorbecke (1984); miary te, od inicjałów nazwisk autorów, nazwano FGT. Zakłada się, że dana jest skończona populacja $U = 1, \ldots, j, \ldots, N$ podzielona na D domen lub obszarów o liczebnościach $N_1, \ldots, N_D$. Wartość granicy ubóstwa określona jest przez z, a E_{dj} oznacza dochód j-tej jednostki w d-tej domenie. Jeśli $E_{dj} < z$, wówczas jednostka j z obszaru d określana jest jako znajdująca się w sferze ubóstwa. Ogólny wzór na wskaźniki ubóstwa z rodziny FGT wyrażony jest następująco:

$$F_{\alpha d} = \frac{1}{N}\sum_{j=1}^{N_d}\left(\frac{z - E_{dj}}{z}\right)^{\alpha} I\left(E_{dj} < z\right), \quad \alpha \geq 0, \quad d = 1, \ldots, D \tag{1}$$

gdzie $I\left(E_{dj} < z\right) = 1$, jeśli $E_{dj} < z$, oraz $I\left(E_{dj} < z\right) = 0$ w przeciwnym przypadku.

Dla $\alpha = 0$ otrzymuje się wskaźnik zasięgu ubóstwa, czyli stopę ubóstwa. Z kolei po podstawieniu $\alpha = 1$ wynikiem jest głębokość ubóstwa, która informuje o odległości dochodów osób ubogich od granicy ubóstwa. Innymi słowy, wskaźnik ten określa stopień ubóstwa osób ubogich.

2.2. Determinanty ubóstwa

Działania prowadzące do redukcji poziomu ubóstwa wymagają identyfikacji jego symptomów. Haughton i Khandker (2009) wskazują trzy poziomy, na których można obserwować kluczowe przyczyny ubóstwa i mierzyć charakterystyki tego zjawiska:

- regionalny;
- środowiska lokalnego;
- jednostki (gospodarstwa domowego lub osoby), obejmujący charakterystyki demograficzne, ekonomiczne i społeczne.

Główny Urząd Statystyczny w swoich publikacjach rozpatruje zjawisko ubóstwa w bardzo zbliżonych przekrojach społeczno-demograficzno-ekonomicznych, przy czym wśród typów gospodarstwa domowego wyróżnia także małżeństwa z co najmniej czworgiem dzieci na utrzymaniu, w przypadku których zasięg ubóstwa jest bardzo wysoki i w 2011 r. wynosił 47,5% (GUS i US w Łodzi, 2013b).

Uwzględnienie grup wieku pozwala zauważyć, że wartości stopy ubóstwa w grupie 0–17 lat są wyższe (23,1%) niż wśród osób w wieku 18–64 lat (15,9%) czy 65 lat i więcej (11,2%).

W badaniu *Jakość życia, kapitał społeczny, ubóstwo i wykluczenie społeczne w Polsce* (GUS i US w Łodzi, 2013a) opracowano model regresji logistycznej dla ubóstwa dochodowego. W kategorii głównego źródła utrzymania gospodarstwa domowego największe prawdopodobieństwo znalezienia się w sferze ubóstwa miały gospodarstwa utrzymujące się ze świadczeń społecznych i z rent – w odniesieniu do poziomu referencyjnego, jakim była praca najemna. W omawianym badaniu uwzględniono także zawód głowy gospodarstwa domowego, a za poziom referencyjny przyjęto wartości wskaźników dla robotników przemysłowych i rzemieślników. Stwierdzono, że większe ryzyko ubóstwa dotyczy osób, które pracują jako rolnicy, ogrodnicy, leśnicy i rybacy, a także pracownicy zatrudnieni przy pracach prostych. Najmniejsze prawdopodobieństwo ubóstwa odnotowano wśród kadry menedżerskiej, wyższych urzędników i kierowników.

3. Metody estymacji poziomu ubóstwa

Podstawowym narzędziem w szacowaniu parametrów populacji na podstawie danych z badań reprezentacyjnych jest estymacja bezpośrednia. Przy odpowiednio licznej próbie estymator bezpośredni jest nieobciążony i efektywny (Rao i Molina, 2015), ale na niższym poziomie agregacji, co oznacza mniej liczną próbę, traci swoje własności. Pomimo to oceny parametrów uzyskane za pomocą estymatora bezpośredniego są wykorzystywane jako wielkości wejściowe przy estymacji na poziomie obszaru i stanowią punkt odniesienia w ocenie jakości oszacowań.

3.1. Estymacja bezpośrednia

W badaniach reprezentacyjnych stosuje się estymator bezpośredni zaproponowany przez Horvitza i Thompsona (1952). Niech s oznacza próbę wylosowaną z populacji U, a s_d będzie podpróbą z obszaru d o liczebności $n_d < N_d$. Przez $w_{dj} = \pi_{dj}^{-1}$ oznaczono wagę z próby będącą odwrotnością prawdopodobieństwa inkluzji pierwszego rzędu. Oszacowanie wartości globalnej w domenie d uzyskuje się z wykorzystaniem wzoru:

$$\hat{Y}_d^{HT} = \sum_{j=1}^{n_d} y_{dj} w_{dj} \tag{2}$$

gdzie:
$\hat{Y}_d^{HT}$ – oszacowanie globalnej wartości cechy Y w d-tej domenie,
y_{dj} – wartość cechy Y dla j-tej jednostki w d-tym obszarze,
w_{dj} – wartość wagi wynikającej ze schematu losowania dla j-tej jednostki w d-tym obszarze.

W celu estymacji bezpośredniej wskaźników z rodziny FGT należy zmodyfikować formułę (2) w następujący sposób:

$$\hat{F}_{\alpha d}^{HT} = N_d^{-1} \sum_{j \in s_d} w_{dj} \left(\frac{z - E_{dj}}{z} \right)^{\alpha} I(E_{dj} < z), \quad j = 1, \ldots, N_d \tag{3}$$

gdzie:
w_{dj} – wartość wagi wynikającej ze schematu losowania dla j-tej jednostki w d-tym obszarze,
N_d – liczba osób w d-tym obszarze,
E_{dj} – dochód j-tej jednostki,
z – granica ubóstwa,
$I(.)$ – funkcja indykatorowa przyjmująca wartość 1, jeśli wyrażenie wewnątrz funkcji jest prawdziwe, a wartość 0 w przeciwnym przypadku.

Estymator bezpośredni wykorzystuje wyłącznie dane pochodzące z próby i dla dużych wartości n_d jest nieobciążony i efektywny, natomiast mała liczebność próby implikuje duże wartości wariancji. Jeśli dana domena w próbie nie jest w ogóle reprezentowana ($n_d = 0$), to ten sposób estymacji nie ma zastosowania (Guadarrama, Molina i Rao, 2016).

3.2. Estymacja pośrednia

W metodach estymacji pośredniej przyjmuje się podejście obszarowe lub jednostkowe. W podejściu obszarowym zmienną zależną jest wskaźnik ubóstwa mierzony na określonym poziomie terytorialnym, a w podejściu jednostkowym – dochód gospodarstwa domowego. W omawianym badaniu zastosowano ten drugi sposób.

3.2.1. Empiryczna metoda bayesowska

Empiryczna metoda bayesowska (EB) polega na dopasowaniu liniowego modelu mieszanego opisującego dochód na poziomie gospodarstwa domowego na podstawie danych z badania reprezentacyjnego, a następnie na przeprowadzeniu dużej liczby symulacji Monte Carlo z wykorzystaniem danych z badania pełnego w celu wyznaczenia teoretycznych wartości dochodu dla osób spoza próby na podstawie zmiennych pomocniczych (Molina i Rao, 2010). Metoda ta jest wykorzystywana do estymacji ubóstwa m.in. przez Bank Światowy (Van der Weide, 2014).

Dany jest wektor $y = (y_1, \ldots, y_N)'$, zawierający wartości zmiennej losowej powiązanej z N jednostkami skończonej populacji. Niech y_s będzie fragmentem wektora y zawierającym informacje z wylosowanej próby s, a y_r – wektorem elementów spoza próby. Po uporządkowaniu elementów można zapisać, że $y = (y'_s, y'_r)$. Celem jest oszacowanie wartości funkcji $\delta = h(y)$ zmiennej losowej y z wykorzystaniem danych z próby y_s. Dla estymatora δ średni kwadrat błędu (MSE) jest dany wzorem:

$$\mathrm{MSE}(\hat{\delta}) = E_y\{(\hat{\delta} - \delta)^2\} \tag{4}$$

gdzie E_y – wartość oczekiwana z uwzględnieniem łącznego rozkładu wektora y.

Najlepszym estymatorem (ang. *best predictor*, BP) parametru δ jest funkcja y_s, która minimalizuje (4) i jest dana warunkową wartością oczekiwaną:

$$\widehat{\delta^B} = E_{y_r}(\delta | y_s) \tag{5}$$

gdzie wartość oczekiwana uwzględnia warunkowy rozkład y_r.

Warto zauważyć, że ten estymator jest nieobciążony, ponieważ:

$$E_{y_s}(\hat{\delta}^B) = E_{y_s}\{E_{y_r}(\delta | y_s)\} = E_y(\delta) \tag{6}$$

Zwykle $\hat{\delta}^B$ zależy od wektora θ nieznanych parametrów modelu. Wówczas empiryczny BP δ otrzymuje się poprzez zastąpienie θ odpowiednim estymatorem $\hat{\theta}$ i następnie oszacowanie (5), przyjmując $\theta = \hat{\theta}$.

Poniżej zaprezentowano sposób oszacowania BP dla rodziny wskaźników ubóstwa FGT w przypadku małych domen. Zakłada się transformację zmiennej dochodowej E_{dj} według formuły $Y_{dj} = T(E_{dj})$, biorąc pod uwagę, że wektor y zawiera wartości transformowanej zmiennej Y_{dj} dla wszystkich jednostek populacji, tak że $y \sim N(\mu, V)$. Wówczas zmienną losową

$$\hat{F}_{\alpha dj} = \left(\frac{z - E_{dj}}{z}\right)^{\alpha} I(E_{dj} < z), \quad j = 1, \ldots, N_d \tag{7}$$

można wyrazić w kontekście Y_{dj} jako:

$$F_{\alpha dj} = \left(\frac{z - T^{-1}(Y_{dj})}{z}\right)^{\alpha} I\{T^{-1}(Y_{dj}) < z\} =: h_{\alpha}(Y_{dj}), \quad j = 1, \ldots, N_d \tag{8}$$

Wobec tego miara ubóstwa (1) jest nieliniową funkcją wektora y. Po przyjęciu za $\delta = F_{\alpha d}$ i podstawieniu do (5) najlepszy estymator (BP) parametru $F_{\alpha d}$ można wyrazić wzorem:

$$\hat{F}^{B}_{\alpha d} = E_{y_r}(F_{\alpha d}|y_s) \tag{9}$$

Dekomponując $F_{\alpha d}$ (1) na jednostki wylosowane i spoza próby oraz uwzględniając warunkową wartość oczekiwaną (9), otrzymuje się estymator wskaźnika ubóstwa:

$$\hat{F}^{B}_{\alpha d} = \frac{1}{N_d}\left\{\sum_{j \in s_d} F_{\alpha dj} + \sum_{j \in r_d} \hat{F}^{B}_{\alpha dj}\right\} \tag{10}$$

gdzie $\hat{F}^{B}_{\alpha dj}$ jest najlepszym estymatorem $F_{\alpha dj} = h_{\alpha}(Y_{dj})$, danym przez

$$\hat{F}^{B}_{\alpha dj} = E_{y_r}[h_{\alpha}(Y_{dj})|y_s] = \int_{IR} h_{\alpha}(y) f_{Y_{dj}}(y|y_s)\,\mathrm{d}y, \quad j \in r_d \tag{11}$$

gdzie $f_{Y_{dj}}(y \vee y_s)$ jest warunkową gęstością Y_{dj} wektora y_s.

Wartość oczekiwana (11) nie może zostać wyliczona bezpośrednio z powodu złożoności $h_{\alpha}(y)$. Jednakże w przypadku gdy $y = ({y'}_s, {y'}_r)'$ ma rozkład normalny ze średnią daną wektorem $\mu = ({\mu'}_s, {\mu'}_r)'$, a rozkład y_r pod warunkiem y_s dany jest przez

$$y_r|y_s \sim N(\mu_{r|s}, V_{r|s}) \tag{12}$$

gdzie:

$$\mu_{r|s} = \mu_r + V_{rs}V_s^{-1}(y_s - \mu_s) \qquad \text{i} \qquad V_{r|s} = V_r - V_{rs}V_s^{-1}V_{sr} \tag{13}$$

to empiryczne przybliżenie rozwiązania (11) jest możliwe z wykorzystaniem symulacji Monte Carlo o dużej liczbie L wektorów y_r wygenerowanych z (12).

Niech $Y_{dj}^{(l)}$ oznacza wartości zmiennej Y_{dj}, $j \in r_d$ spoza próby, otrzymane w l symulacjach $l = 1, \ldots, L$. Przybliżenie Monte Carlo najlepszego estymatora (BP) Y_{dj} dla $j \in r_d$ można wyrazić jako:

$$\hat{F}_{\alpha dj}^{B} = E_{y_r}\left[h_\alpha\left(Y_{dj}\right)|y_s\right] \approx \frac{1}{L}\sum_{\ell=1}^{L} h_\alpha\left(Y_{dj}^{(\ell)}\right), \qquad j \in r_d \tag{14}$$

Końcowy estymator to $\hat{F}_{\alpha dj}^{EB}$, będący najlepszym empirycznym estymatorem (EBP) parametru $F_{\alpha dj}$. Ostatecznie EBP miary ubóstwa $F_{\alpha d}$ jest określony wzorem:

$$\hat{F}_{\alpha d}^{EB} = \frac{1}{N_d}\left\{\sum_{j \in s_d} F_{\alpha dj} + \sum_{j \in r_d} \hat{F}_{\alpha dj}^{EB}\right\} \tag{15}$$

Wartość oczekiwana (9) może być bezpośrednio przybliżona przez symulację Monte Carlo, co pozwala na estymację właściwie dowolnego parametru dla domeny $\delta_d = h(y_d)$. Wówczas procedura estymacji parametru dla domeny $\delta_d = h(y_d)$ za pomocą metody EB jest następująca:

1. estymacja nieznanych parametrów rozkładu θ transformowanego wektora y z wykorzystaniem danych z próby y_s;
2. wygenerowanie L wektorów spoza próby $y_r^{(l)}$, $l = 1, \ldots, L$ na podstawie (12) i (13); θ jest zastępowane estymatorem $\hat{\theta}$ uzyskanym w pkt 1;
3. rozszerzenie każdego z L wygenerowanych wektorów $y_r^{(l)}$ danymi z próby y_s do postaci wektora populacji $y^l = \left(y'_s, \left(y_r^{(l)}\right)'\right)', l = 1, \ldots, L$;
4. wyznaczenie wartości parametru dla domeny $\delta_d^{(l)} = h\left(y_d^{(l)}\right)$ z wykorzystaniem elementów wektora $y^{(l)}$ dla d-tej domeny – $y_d^l = \left(y'_{ds}, \left(y_{dr}^{(l)}\right)'\right)'$;
5. przybliżenie Monte Carlo estymatora EBP parametru δ_d poprzez uśrednienie wartości parametru dla domeny na podstawie L pseudopopulacji:

$$\delta_d^{EB} = \frac{1}{L}\sum_{\ell=1}^{L} \delta_d^{(\ell)} \tag{16}$$

3.2.2. Liniowy model regresji z zagnieżdżonym składnikiem losowym

Model nadpopulacji ξ oparty na liniowym modelu mieszanym (Battese, Harteri i Fuller, 1988) może zostać wykorzystany do oszacowania (15). Model opisuje liniową dla wszystkich domen relację pomiędzy transformowaną zmienną dochodową Y_{dj} a wektorem x_{dj}, zawierającym wartości p zmiennych niezależnych, a także uwzględnia efekt losowy dla domeny u_d wraz z resztami e_{dj}:

$$\xi\colon\ Y_{dj} = x'_{dj}\beta + u_d + e_{dj},\quad j = 1, \ldots, N_d,\quad d = 1, \ldots, D \tag{17}$$

gdzie $u_d \overset{iid}{\sim} N(0, \sigma_u^2)$ i $e_{dj} \overset{iid}{\sim} N(0, \sigma_e^2)$, a u_d oraz e_{dj} są niezależne.

Wektory i macierze otrzymuje się poprzez wydzielenie elementów dla domeny d:

$$y_d = col_{1\le j\le N_d}(Y_{dj}),\quad e_d = col_{1\le j\le N_d}(e_{dj}),\quad X_d = col_{1\le j\le N_d}(x'_{dj}) \tag{18}$$

Wówczas wektor $y_d, d = 1, \ldots, D$ jest niezależny z $y_d \sim N(\mu_d, V_d)$, gdzie:

$$\mu_d = X_d\beta \ \text{ i } \ V_d = \sigma_u^2 1_{N_d} 1'_{N_d} + \sigma_e^2 I_{N_d} \tag{19}$$

przy czym 1_k oznacza kolumnowy wektor jedynek o rozmiarze k, a I_k jest macierzą jednostkową $k \times k$.

Rozważana jest dekompozycja y_d na jednostki wylosowane i te spoza próby $y_d = (y'_{ds}, y'_{dr})'$ w przypadku $n_d > 0$ i podobnie dla X_d, μ_d i V_d. Przy założeniu normalności modelu (17) można zauważyć, że $y_{dr} \vee y_{ds} \sim N(\mu_{dr\vee s}, V_{dr\vee s})$, gdzie:

$$\mu_{dr|s} = X_{dr}\beta + \sigma_u^2 1_{N_d - n_d} 1'_{n_d} V_{ds}^{-1}(y_{ds} - X_{ds}\beta) \tag{20}$$

i

$$V_{dr|s} = \sigma_u^2(1 - \gamma_d) 1_{N_d - n_d} 1'_{N_d - n_d} + \sigma_e^2 I_{N_d - n_d} \tag{21}$$

dla $V_{ds} = \sigma_u^2 1_{n_d} 1'_{n_d} + \sigma_e^2 I_{n_d}$ oraz $\gamma_d = \sigma_u^2$.

Zastosowanie przybliżenia Monte Carlo (14) pociąga za sobą symulację D wektorów y_{dr} o rozmiarze $N_d - n_d, d = 1, \ldots, D$ i wielowymiarowym rozkładzie normalnym. Powtórzenie tego procesu L razy sprawia, że jest on wymagający obliczeniowo, a dla dużych N_d – niewykonalny. Można tego uniknąć, gdy się zauważy, że macierz $V_{dr\vee s}$ (21) odpowiada macierzy kowariancji wektora y_{dr} wygenerowanego przez model

$$y_{dr} = \mu_{dr|s} + v_d 1_{N_d - n_d} + \varepsilon_{dr} \tag{22}$$

z nowym efektem losowym v_d oraz błędem ε_{dr}, które są niezależne, o rozkładach:

$$v_d \sim N\{0, \sigma_u^2(1-\gamma_d)\}, \quad d = 1, \ldots, D, \quad \varepsilon \sim N\left(0_{N_d - n_d}, \sigma_e^2 I_{N_d - n_d}\right) \tag{23}$$

Jeśli wykorzystuje się (22), to zamiast generowania wektora y_{dr} o wielowymiarowym rozkładzie normalnym wystarczy wygenerować jednowymiarowe zmienne $v_d \sim N\{0, \sigma_u^2(1-\gamma_d)\}$ i $\varepsilon_{dj} \sim N(0, \sigma_e^2)$ niezależnie dla $j \in r_d$. Wówczas uzyskuje się odpowiadające elementy wektora $y_{dj}, j \in r_d$ z (22) przy wykorzystaniu $\mu_{dr \vee s}$ danego przez (23). Jak już wspomniano, w praktyce parametry modelu $\theta = (\beta', \sigma_u^2, \sigma_e^2)'$ są zastępowane przez odpowiednie estymatory $\hat{\theta} = \left(\hat{\beta}', \hat{\sigma}_u^2, \hat{\sigma}_e^2\right)\hat{\theta}'$ i wówczas wartości zmiennej Y_{dj} są generowane z rozkładu normalnego o odpowiednim odchyleniu standardowym.

Jeśli domena d nie znalazła się w próbie, to wartości $Y^{(\ell)_{dj}}$ dla $j = 1, \ldots, N_d$ są generowane metodą bootstrap z wykorzystaniem $Y_{dj} = x'_d j\hat{\beta} + u_d^* + e_{dj}^*$, gdzie $u_d^* \overset{iid}{\sim} N(0, \hat{\sigma}_u^2)$ i $e_{dj}^* \overset{iid}{\sim} N(0, \widehat{\sigma_e^2})$ oraz u_d^* jest niezależne od e_{dj}^*. Wzór (14) wykorzystuje się wówczas do otrzymania estymatora $\hat{F}_{\alpha dj}^{EB}$ dla $F_{\alpha dj}$ i estymatora EB $F_{\alpha d}$ jako:

$$\hat{F}_{\alpha d}^{EB} = N_d^{-1} \sum_{j=1}^{N_d} \hat{F}_{\alpha dj}^{EB} \tag{24}$$

Estymator (24) ma charakter syntetyczny, jeśli żadna jednostka nie jest obserwowana w domenie d.

Opisane podejście bazuje na transformacji zmiennej zależnej w celu zbliżenia rozkładu cechy do rozkładu normalnego. W przypadku występowania silnej skośności operacja ta może nie przynieść pożądanych efektów. Można wtedy zastosować model wykorzystujący uogólniony rozkład beta (Graf, Marín i Molina, 2019), ale wiąże się to ze skomplikowaniem obliczeń.

3.3. Ocena jakości oszacowań

Jako miarę precyzji oszacowań uzyskanych za pomocą opisanych metod przyjęto pierwiastek względnego średniego kwadratu błędu oszacowania – RRMSE (ang. *relative root mean square error*), dany wzorem:

$$\mathrm{RRMSE}\left(\hat{F}_{\alpha d}\right) = \frac{\sqrt{MSE\left(\hat{F}_{\alpha d}\right)}}{\hat{F}_{\alpha d}} \tag{25}$$

gdzie:

$\sqrt{MSE\left(\hat{F}_{\alpha d}\right)}$ – średni błąd oszacowania wskaźnika ubóstwa,

$\hat{F}_{\alpha d}$ – oszacowanie wskaźnika ubóstwa.

Wartości średniego błędu oszacowania można aproksymować z wykorzystaniem metody bootstrap. W przypadku estymacji bezpośredniej przegląd podejść można znaleźć w pracy Woltera (2007), natomiast w przypadku metody EB odpowiedni algorytm zaproponowali González-Manteiga, Lombardía, Molina, Morales i Santamaría (2008).

4. Estymacja pośrednia ubóstwa na poziomie powiatów

W EU-SILC zrealizowanym w 2011 r. wzięło udział 12871 gospodarstw domowych z 375 spośród wszystkich 379 powiatów (cztery powiaty nie były reprezentowane w próbie). Analiza prowadzona na poziomie powiatów wykazała, że w kolejnych 12 powiatach nie znalazła się ani jedna osoba uboga, co uniemożliwiło estymację bezpośrednią dla tych jednostek terytorialnych.

Kolejnym etapem badania było oszacowanie wskaźników ubóstwa na poziomie powiatów. Z uwagi na duże błędy standardowe bezpośrednich oszacowań wskaźników ubóstwa, a tym samym ich małą precyzję, zrezygnowano z podejścia klasycznego i zastosowano estymację pośrednią. Estymacja charakterystyk ubóstwa z wykorzystaniem metod statystyki małych obszarów wymaga – w celu poprawy jakości – użycia zmiennych pomocniczych. Cechy te powinny pochodzić ze źródeł pozbawionych błędów losowych, np. z badań pełnych bądź rejestrów administracyjnych. W podejściu modelowym kluczowy jest dobór zmiennych, które dobrze opisują zróżnicowanie estymowanych cech. Nieodpowiednia specyfikacja zmiennych niezależnych może powodować obciążenie wyników.

4.1. Liniowy model mieszany dochodu ekwiwalentnego

Na podstawie danych z EU-SILC z 2011 r. oszacowano parametry liniowego modelu regresji z zagnieżdżonym składnikiem losowym. Wartości zmiennej zależnej – dochodu ekwiwalentnego gospodarstwa – zostały przekształcone za pomocą transformacji Boxa-Coxa. Jako zmienne pomocnicze wykorzystano cechy demograficzne, ekonomiczne i społeczne mierzone na poziomie gospodarstwa domowego, które wybrano na podstawie studiów literatury dotyczących symptomów ubóstwa. Za efekt losowy przyjęto powiat. Obliczenia wykonano w programie R z wykorzystaniem pakietu lme4 (Bates, Mächler, Bolker i Walker, 2015). Parametry modelu przedstawia tabl. 1[1].

[1] Dane zaprezentowane w tablicy są tożsame z wartościami zamieszczonymi w raporcie *Pomiar ubóstwa na poziomie powiatów (LAU 1)* (Szymkowiak, 2015) ze względu na zastosowanie innej transformacji zmiennej zależnej.

Tabl. 1. Parametry liniowego modelu mieszanego objaśniającego dochód ekwiwalentny na poziomie gospodarstwa domowego

Zmienne niezależne	Parametr β	Błąd standardowy	Statystyka t
(wyraz wolny)	27,65	0,09	319,37
Odsetek mężczyzn w gospodarstwie	0,79	0,10	7,76
Odestek osób w wieku: 30–44 lat w gospodarstwie	0,56	0,12	4,58
65 lat i więcej w gospodarstwie	–0,32	0,08	–3,91
Odsetek osób: bezrobotnych w gospodarstwie	–5,05	0,19	–26,54
niepełnosprawnych w gospodarstwie	–2,44	0,16	–15,15
z wykształceniem podstawowym w gospodarstwie	–1,00	0,10	–10,30
z wykształceniem wyższym w gospodarstwie	4,05	0,11	38,50
Gospodarstwo posiadające: 1 pokój (zmienna binarna)	–0,89	0,10	–9,26
3 pokoje i więcej (zmienna binarna)	0,77	0,06	13,10
Miejsce zamieszkania: wieś lub miasto do 20 tys. mieszkańców (zmienna binarna)	–1,11	0,06	–17,11
Wskaźnik obciążenia demograficznego dzieci w gospodarstwie	–1,07	0,07	–14,69

Źródło: opracowanie własne na podstawie danych z EU-SILC 2011.

Wszystkie zmienne niezależne są istotne, a odpowiadające im znaki przy parametrze β są zgodne z logicznym kierunkiem wpływu na zmienną zależną. Większe wartości czterech zmiennych: odsetka mężczyzn w gospodarstwie, odsetka osób w wieku 30–44 lat w gospodarstwie, odsetka osób z wykształceniem wyższym w gospodarstwie oraz gospodarstwo posiadające 3 pokoje i więcej (zmienna binarna) wpływają na wzrost dochodu w gospodarstwie. Wartości parametru β przy pozostałych zmiennych mają znak ujemny, co oznacza, że większe wartości tych cech powodują spadek dochodu w gospodarstwie.

Wariancja efektu losowego wyniosła $\widehat{\sigma_u} = 0{,}2702$, natomiast wariancja błędu losowego $\widehat{\sigma_e} = 8{,}7083$. Na podstawie testu permutacyjnego wykazano, że efekt losowy jest w modelu istotny. Przeprowadzono diagnostykę modelu z wykorzystaniem wykresów kwantylowych. Stwierdzono, że zarówno rozkład efektów, jak i reszt jest zbliżony do normalnego. Pozytywnie zweryfikowano także założenie o niezależności efektów losowych i reszt (Biecek, 2011).

W kolejnym kroku zastosowano algorytm estymacji metodą EB dla danych pochodzących z NSP 2011; przyjęto $L = 200$ i $B = 200$. Obliczenia przeprowadzono z wykorzystaniem pakietu sae (Molina i Marhuenda, 2015) w programie R. Na tej podstawie uzyskano oszacowania stopy i głębokości ubóstwa na poziomie powiatów w Polsce. Tablica 2 zawiera rozkład względnych błędów oszacowań stopy i głębokości ubóstwa.

Tabl. 2. Rozkład względnych błędów oszacowań stopy i głębokości ubóstwa oszacowanych z wykorzystaniem estymacji bezpośredniej (HT) i pośredniej (EB)

Wskaźniki i metody estymacji	Minimum	Q1	Mediana	Średnia	Q3	Maksimum
	w %					
Stopa ubóstwa: HT	11,50	22,28	28,52	32,81	37,57	99,97
EB	6,45	10,36	11,58	11,81	13,08	19,31
Głębokość ubóstwa: HT	14,41	25,05	31,90	37,00	42,46	99,97
EB	9,77	13,19	14,46	14,78	16,17	23,62

Źródło: opracowanie własne na podstawie danych EU-SILC 2011 i NSP 2011.

Estymacja pośrednia poprawia precyzję oszacowań w porównaniu do estymacji bezpośredniej. Średni wskaźnik precyzji dla stopy ubóstwa oszacowanej z wykorzystaniem metody EB wyniósł 11,81%, podczas gdy dla estymacji bezpośredniej było to 32,81%. Podobną poprawę średniej wartości RRMSE można zaobserwować także w przypadku głębokości ubóstwa – z 37,00% (estymacja HT) do 14,78% (estymacja EB). Względny błąd oszacowania stopy ubóstwa nie przekroczył 20%.

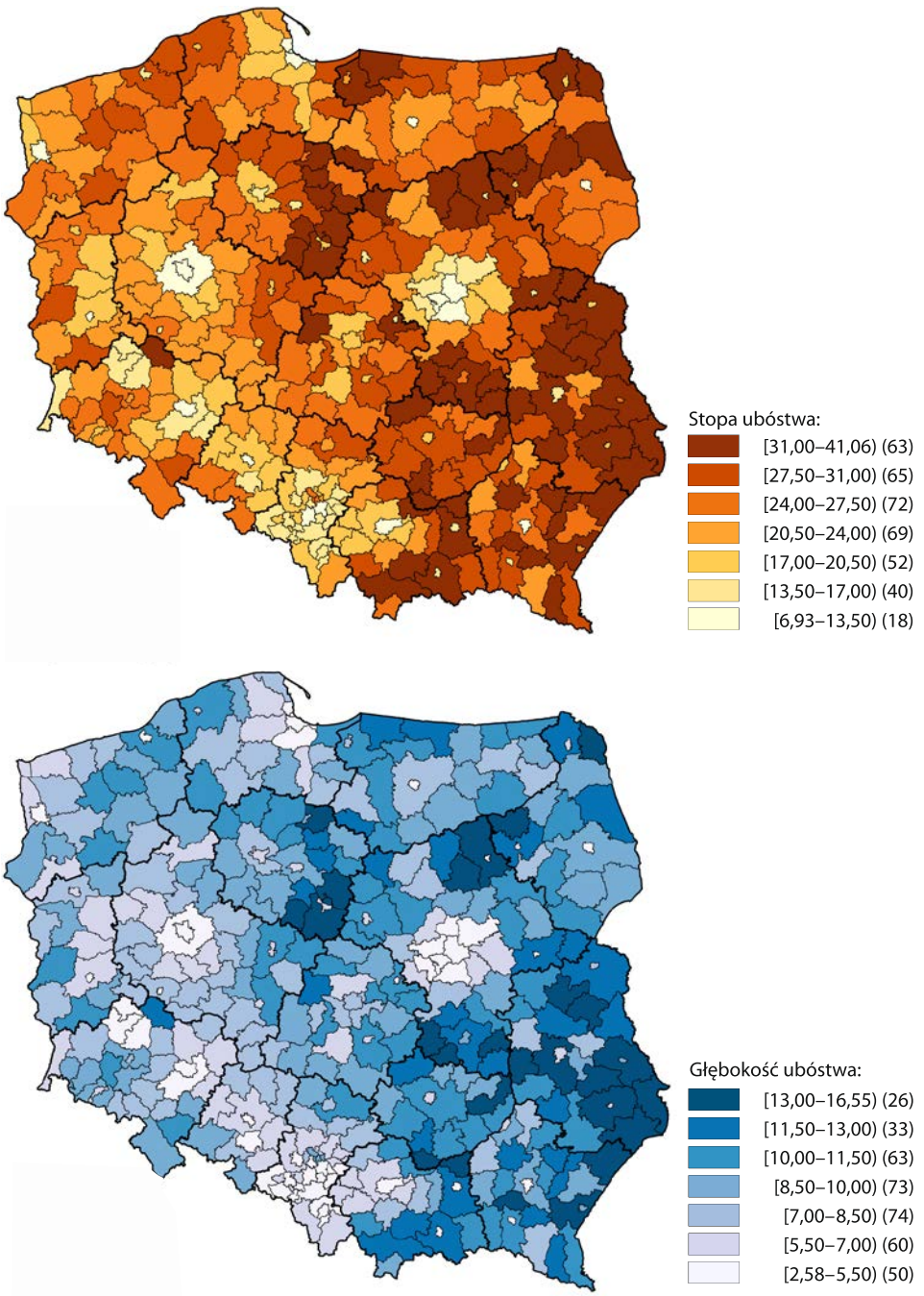
4.2. Terytorialne zróżnicowanie ubóstwa w Polsce

Przeprowadzono także przestrzenną analizę ubóstwa w Polsce. Jej rezultaty przedstawiono na kartogramach (mapa), które pozwalają na zobrazowanie przestrzennych zależności (Bedi, Coudouel i Simler, 2007). Zidentyfikowano obszary najmniej i najbardziej narażone na występowanie zjawiska ubóstwa. Ponadto przeprowadzono ocenę spójności przestrzennej otrzymanych oszacowań wskaźników ubóstwa z wykorzystaniem statystyki Morana I.

Największe wartości stopy ubóstwa obserwuje się w powiatach we wschodniej części kraju oraz jednostkach położonych w znacznej odległości od stolic województw. Występowaniem największego ubóstwa – powyżej 40% – charakteryzują się powiaty chełmski (41,1%) i hrubieszowski (40,9%). Wśród 20 powiatów o największych wartościach stopy ubóstwa znajdują się trzy powiaty położone we wschodniej części woj. kujawsko-pomorskiego, pięć powiatów woj. lubelskiego i jeden powiat – dąbrowski – z woj. małopolskiego. Pięć kolejnych powiatów należy do woj. mazowieckiego, przy czym są one znacznie oddalone od Warszawy. Są to powiaty: ostrołęcki i makowski (północna część województwa) oraz przysuski, szydłowiecki i zwoleński (południowa część województwa). Grupę dopełniają po dwie jednostki z województw podkarpackiego, podlaskiego i świętokrzyskiego.

Wśród 20 powiatów o najniższej stopie ubóstwa znajdują się wyłącznie miasta na prawach powiatu oraz jednostki bezpośrednio z nimi graniczące. Na pierwszym miejscu plasuje się Warszawa (6,9%), na drugim – Poznań (10,9%), a na trzecim – graniczący z Warszawą powiat pruszkowski (11,2%). Dalsze pozycje zajmują miasta na prawach powiatu: Gdynia i Sopot (po 11,7%) oraz Opole, Rzeszów i Olsztyn (po 12,8%).

Mapa terytorialnego zróżnicowania stopy i głębokości ubóstwa w przekroju powiatów w 2011 r.



Uwaga. W nawiasach okrągłych podano liczbę powiatów.
Źródło: opracowanie własne na podstawie danych EU-SILC 2011 i NSP 2011.

Głębokość ubóstwa jest w dużym stopniu skorelowana ze stopą ubóstwa ($r = 0{,}99$). Najmniejsze ubóstwo wśród osób ubogich obserwuje się w Warszawie (2,6%), a następnie w powiecie pruszkowskim i Sopocie (po 3,8%) oraz powiecie poznańskim i Poznaniu (po 3,9%). Największe wartości wskaźnika głębokości ubóstwa odnotowano w tych samych powiatach, w których występowała wysoka stopa ubóstwa.

Na podstawie wartości wskaźników ubóstwa oszacowanych na szczeblu powiatowym oraz macierzy sąsiedztwa obliczono także statystykę Morana I. Stwierdzono istnienie dodatniej korelacji przestrzennej – statystyka wyniosła odpowiednio 0,45 dla stopy ubóstwa i 0,47 dla głębokości ubóstwa. Istnienie umiarkowanie dodatniej autokorelacji przestrzennej świadczy o podobieństwie wartości sąsiadujących.

4.3. Zbieżność wyników estymacji pośredniej z danymi z rejestrów

W sytuacji gdy wartości wskaźników ubóstwa w populacji nie są znane, w ocenie ich poprawności można posłużyć się zmiennymi proxy (Zhang i Giusti, 2016). Wybór tych cech powinien być uzasadniony, a pomiędzy analizowaną cechą a zmienną proxy musi występować związek. Osoby, które korzystają ze świadczeń pomocy społecznej (zmienna proxy), są ubogie (zmienna analizowana). To samo można zauważyć w przypadku osób otrzymujących zasiłek dla bezrobotnych, których odsetek stanowi zmienną proxy dla liczby osób bezrobotnych. Trzeba jednak zauważyć, że nie zachodzi relacja odwrotna – nie wszystkie osoby bezrobotne otrzymują zasiłek (Fenton, 2013).

Oszacowania stopy i głębokości ubóstwa porównano z danymi na temat bezrobocia rejestrowanego oraz korzystania z pomocy społecznej. Zmienne te pochodziły z rejestrów, a zatem nie były obciążone błędem losowym, dzięki czemu można je uznać za precyzyjne miary porównawcze. Ponadto są to cechy ściśle związane ze zjawiskiem ubóstwa (GUS i US w Łodzi, 2017).

Oszacowania bezpośrednie są w bardzo małym stopniu skorelowane ze wskaźnikami z zakresu bezrobocia ($r = 0{,}198$) i pomocy społecznej ($r = 0{,}324$). Zastosowanie estymacji pośredniej bazującej na podejściu jednostkowym przynosi istotną poprawę. Porównanie tych oszacowań ze stopą bezrobocia długotrwałego pokazuje istotny wzrost wartości w porównaniu do estymacji bezpośredniej – w przypadku metody EB współczynnik korelacji liniowej Pearsona wynosi $r = 0{,}641$. Największe wartości współczynników obserwuje się dla cechy zasięg korzystania ze środowiskowej pomocy społecznej ogółem – $r = 0{,}747$.

Wyniki estymacji świadczą o tym, że oszacowania pośrednie uzyskane przy zastosowaniu podejścia jednostkowego na poziomie powiatów są w dużo większym stopniu zbieżne z rzeczywistymi wartościami niż oszacowania bezpośrednie. Większe podobieństwo oszacowań i wskaźników opracowanych na podstawie rejestrów administracyjnych wskazuje na lepszą jakość wyników otrzymanych metodami statystyki małych obszarów.

5. Podsumowanie

Wartości wskaźników ubóstwa publikowane są dla całego kraju oraz w przekroju regionów i wybranych grup społeczno-demograficznych. Wielkość próby, jak również wykorzystywana obecnie metoda szacunku – estymacja bezpośrednia – z powodu dużych błędów oszacowań nie pozwalają na publikację wyników na niższych poziomach agregacji. Ponadto z roku na rok coraz więcej osób odmawia udziału w badaniach reprezentacyjnych. Wskaźnik kompletności w przypadku Europejskiego Badania Dochodów i Warunków Życia (EU-SILC) w 2006 r. wyniósł 69,9%, a w 2016 r. – już tylko 48,4% (GUS, 2008, 2017). Niemniej równocześnie z obserwowanym spadkiem wartości wskaźników realizacji badań reprezentacyjnych rosną potrzeby odbiorców informacji. Oczekuje się danych dla szczegółowych przekrojów oraz coraz mniejszych jednostek administracyjnych czy terytorialnych.

W badaniu omawianym w artykule do oszacowania wskaźników ubóstwa na poziomie powiatów zastosowano estymację pośrednią – metodę empiryczną bayesowską (EB), która polega na generowaniu pseudopopulacji metodą Monte Carlo. W tym celu wykorzystano dane z EU-SILC oraz Narodowego Spisu Powszechnego Ludności i Mieszkań (NSP). Metoda EB bazuje na danych jednostkowych, które lepiej opisują ubóstwo gospodarstw domowych niż dane zagregowane. Tym samym parametry liniowego modelu mieszanego szacowane są na podstawie większej liczby obserwacji. Przyjęcie za zmienną zależną dochodu ekwiwalentnego gospodarstwa umożliwia szacowanie, oprócz wskaźników ubóstwa, dowolnego parametru opartego na dochodzie, np. współczynnika Giniego.

Zastosowanie metod statystyki małych obszarów umożliwiło uzyskanie precyzyjnych oszacowań stopy i głębokości ubóstwa na poziomie powiatów oraz oszacowanie wartości wskaźników ubóstwa także dla tych 16 jednostek terytorialnych, w których nie było żadnego reprezentanta populacji ubogich lub które nie znalazły się w próbie EU-SILC. Otrzymane wyniki istotnie zwiększają szczegółowość dostępnych informacji, zwłaszcza w przypadku wskaźnika głębokości ubóstwa, który publikowany jest wyłącznie w wybranych przekrojach społeczno-demograficznych. W celu merytorycznej oceny oszacowań zaproponowano wykorzystanie danych na temat bezrobocia i pomocy społecznej pochodzących z rejestrów administracyjnych. Wykazano, że korelacja pomiędzy tymi zmiennymi a oszacowaniami pośrednimi jest wyższa niż w przypadku oszacowań bezpośrednich.

Spis ludności stanowi podstawowe źródło danych o stanie i strukturze populacji kraju. Ze względu na koszty i zakres badanie to jest realizowane co dekadę. Wyniki omawianego badania, w którym wykorzystano dane z NSP 2011, nie odnoszą się zatem do sytuacji gospodarstw domowych w Polsce aktualnej w czasie publikacji artykułu. Od 2011 r. wdrożono wiele programów pomocy społecznej skierowanych

do gospodarstw domowych, takich jak „Rodzina 500+", „Aktywne formy przeciwdziałania wykluczeniu społecznemu", „Program Operacyjny Pomoc Żywnościowa", „Karta Dużej Rodziny" oraz wiele innych. Mają one na celu m.in. ograniczenie ubóstwa. Jędrzejczak i Pekasiewicz (2020), na podstawie badania budżetów gospodarstw domowych, wskazują na istotną poprawę sytuacji dochodowej rodzin wielodzietnych w latach 2015 i 2016. Ocenę wpływu transferów społecznych na problem ubóstwa można znaleźć m.in. w pracach Brzezińskiego i Najsztuba (2017), Szarfenberga (2018) i Szulca (2019).

Zastosowanie zaproponowanej metodyki oraz wykorzystanie danych ze spisu powszechnego, który będzie prowadzony w 2021 r., powinno umożliwić porównanie poziomu ubóstwa w gospodarstwach domowych w latach 2011 i 2021. Oszacowanie wskaźników ubóstwa z wykorzystaniem metody EB dla pozostałych lat wymaga dostępu do danych jednostkowych, np. z rejestrów administracyjnych. Dane z rejestrów rządowych mogą być wykorzystane także w celu korekty rozkładu dochodów, ponieważ osoby osiągające bardzo wysokie dochody zwykle odmawiają udziału w badaniach ankietowych (Bukowski i Novokmet, 2019).

Bibliografia

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: 10.18637/jss.v067.i01.

Battese, G. E., Harter, R. M., Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, *83*(401), 28–36. DOI: 10.1080/01621459.1988.10478561.

Bedi, T., Coudouel, A., Simler, K. (2007). *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*. Washington: The World Bank.

Biecek, P. (2011). *Analiza danych z programem R: Modele liniowe z efektami stałymi, losowymi i mieszanymi*. Warszawa: Wydawnictwo Naukowe PWN.

Brzeziński, M., Najsztub, M. (2017). The impact of „Family 500+" programme on household incomes, poverty and inequality. *Polityka Społeczna*, (1), 16–25.

Bukowski, P., Novokmet, F. (2019). Between Communism and Capitalism: Long-Term Inequality in Poland, 1892–2015. *CEP Discussion Paper,* (1628).

Chambers, R., Tzavidis, N. (2006). M-quantile models for small area estimation. *Biometrika*, *93*(2), 255–268. DOI: 10.1093/biomet/93.2.255.

Chandra, H., Chambers, R., Salvati, N. (2012). Small area estimation of proportions in business surveys. *Journal of Statistical Computation and Simulation*, *82*(6), 783–795. DOI: 10.1080 /00949655.2011.554834.

Dehnel, G. (2010). *Rozwój mikroprzedsiębiorczości w Polsce w świetle estymacji dla małych domen*. Poznań: Wydawnictwo Uniwersytetu Ekonomicznego.

EEC. (1985). Council Decision of 19 December 1984 on specific community action to combat poverty. *Official Journal of the European Communities*, *28*(L 2), 24–25.

Fenton, A. (2013). *Small-area measures of income poverty*. London: Centre for Analysis of Social Exclusion, London School of Economics.

Foster, J., Greer, J., Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, *52*(3), 761–776. DOI: 10.2307/1913475.

Gołata, E. (2004). *Estymacja pośrednia bezrobocia na lokalnym rynku pracy*. Poznań: Wydawnictwo Akademii Ekonomicznej.

González-Manteiga, W., Lombardía, M., Molina, I., Morales, D., Santamaría, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Computational Statistics and Data Analysis*, *52*(12), 5242–5252. DOI: 10.1016/j.csda.2008.04.031.

Graf, M., Marín, J. M., Molina, I. (2019). A generalized mixed model for skewed distributions applied to small area estimation. *Test*, *28*(2), 565–597. DOI: 10.1007/s11749-018-0594-2.

Guadarrama, M., Molina, I., Rao, J. N. K. (2016). A comparison of small area estimation methods for poverty mapping. *Statistics in Transition new series* and *Survey Methodology*, *17*(1), 41–66. DOI: 10.21307/stattrans-2016-005.

GUS. (2008). *Dochody i warunki życia ludności Polski (raport z badania EU-SILC 2006)*. Warszawa: Główny Urząd Statystyczny.

GUS. (2017). *Dochody i warunki życia ludności Polski (raport z badania EU-SILC 2015)*. Warszawa: Główny Urząd Statystyczny.

GUS. (2019). *Dochody i warunki życia ludności Polski – raport z badania EU-SILC 2018*. Warszawa: Główny Urząd Statystyczny.

GUS. (2020). *Sytuacja gospodarstw domowych w 2019 r. w świetle wyników badania budżetów gospodarstw domowych.* Informacja sygnalna z dnia 29.05.2020. Pobrane z: https://stat.gov.pl/files/gfx/portalinformacyjny/pl/defaultaktualnosci/5486/3/19/1/sytuacja_gospodarstw_domowych_w_2019_r._w_swietle_wynikow_badania_budzetow_gospodarstw_domowych.pdf.

GUS, US w Łodzi. (2013a). *Jakość życia, kapitał społeczny, ubóstwo i wykluczenie społeczne w Polsce*. Warszawa: Główny Urząd Statystyczny.

GUS, US w Łodzi. (2013b). *Ubóstwo w Polsce w świetle badań GUS*. Warszawa: Główny Urząd Statystyczny.

GUS, US w Łodzi. (2017). *Jakość życia w Polsce w 2015 roku: Wyniki badania spójności społecznej*. Warszawa: Główny Urząd Statystyczny.

Haughton, J., Khandker, S. R. (2009). *Handbook on Poverty and Inequality*. Washington: The World Bank.

Horvitz, D. G., Thompson, D. J. (1952). A Generalization of Sampling Without Replacement From a Finite Universe. *Journal of the American Statistical Association*, *47*(260), 663–685. DOI: 10.2307/2280784.

Jędrzejczak, A., Pekasiewicz, D. (2020). Changes in Income Distribution for Different Family Types in Poland. *International Advances in Economic Research*, *26*(2), 135–146. DOI: 10.1007/s11294-020-09785-1.

Kurowski, P. (2002). *Koszyki minimum socjalnego i minimum egzystencji – dotychczasowe podejście*. Pobrane z: https://www.ipiss.com.pl/wp-content/uploads/downloads/2012/08/rola_funk_min_soc_egz.pdf.

Molina, I., Marhuenda, Y. (2015). Sae: An R Package for Small Area Estimation. *The R Journal*, *7*(1), 81–98. DOI: 10.32614/rj-2015-007.

Molina, I., Rao, J. N. K. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, *38*(3), 369–385. DOI: 10.1002/cjs.10051.

Panek, T. (2010). Multidimensional approach to poverty measurement: Fuzzy measures of the incidence and the depth of poverty. *Statistics in Transition new series*, *11*(2), 361–380.

Panek, T. (2011). *Ubóstwo, wykluczenie społeczne i nierówności*: *Teoria i praktyka pomiaru*. Warszawa: Oficyna Wydawnicza Szkoła Główna Handlowa.

Rao, J. N. K., Molina, I. (2015). *Small Area Estimation*. Hoboken, New Jersey: John Wiley & Sons.

Rojas-Perilla, N., Pannier, S., Schmid, T., Tzavidis, N. (2020). Data-driven transformations in small area estimation. *Journal of the Royal Statistical Society: Series A: Statistics in Society*, *183*(1), 121–148. DOI: 10.1111/rssa.12488.

Szarfenberg, R. (2018). Ubóstwo i pogłębiona deprywacja materialna rodzin w kontekście wdrożenia programu „Rodzina 500 plus". *Polityka Społeczna*, *45*(1), 11–17.

Szulc, A. (2019). Transfery socjalne w Polsce w kontekście ubóstwa monetarnego i wielowymiarowego. *Wiadomości Statystyczne. The Polish Statistician*, (3), 7–26. DOI: 10.5604/01.3001.0013.8519.

Szymkowiak, M., Młodak, A., Wawrowski, Ł. (2017). Mapping poverty at the level of subregions in Poland using indirect estimation. *Statistics in Transition new series*, *18*(4), 609–635. DOI: 10.21307/stattrans-2017-003.

Szymkowiak, M. (red). (2015). *Pomiar ubóstwa na poziomie powiatów (LAU 1) – etap II*. Jachranka: Centrum Badań i Edukacji Statystycznej GUS.

Ulman, P., Ćwiek, M. (2020). Measuring housing poverty in Poland: a multidimensional analysis. *Housing Studies*, 1–19. DOI: 10.1080/02673037.2020.1759515.

Van der Weide, R. (2014). *A Review of Empirical Bayes Prediction Using Sampling Weights with Poverty Mapping in Mind*. Washington: The World Bank.

Wawrowski, Ł. (2016). The Spatial Fay-Herriot Model in Poverty Estimation. *Folia Oeconomica Stetinensia*, *16*(2), 191–202. DOI: 10.1515/foli-2016-0034.

Wilak, K. M. (2014). Strukturalne modele szeregów czasowych w estymacji stopy bezrobocia w dezagregacji na województwa, płeć i wiek. *Przegląd Statystyczny*, (4), 409–431.

Wolter, K. (2007). *Introduction to variance estimation*. New York: Springer.

Zhang, L.-C., Giusti, C. (2016). Small Area Methods and Administrative Data Integration. W: M. Pratesi (red.), *Analysis of Poverty Data by Small Area Estimation* (s. 61–82). Hoboken, New Jersey: John Wiley & Sons.

Żądło, T. (2015). *Statystyka małych obszarów w badaniach ekonomicznych: Podejście modelowe i mieszane*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego.

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 8, 27–38 DOI: 10.5604/01.3001.0014.3525
Statystyka w praktyce / Statistics in practice

Analiza rozbieżności danych lustrzanych dotyczących masy towarów w statystykach handlu wewnątrzunijnego

Iwona Markowicz[a], Paweł Baran[b]

Streszczenie. W badaniach prowadzonych dotychczas przez autorów artykułu ocena jakości danych lustrzanych w wymianie towarowej między krajami Unii Europejskiej (UE) opierała się na wartości towarów. Analogiczne podejście stosuje wielu badaczy. Celem badania omawianego w artykule jest ocena jakości danych dotyczących obrotu wewnątrzunijnego na podstawie nie tylko wartości, lecz także ilości towarów. W analizie rozbieżności danych w handlu między krajami UE, ze szczególnym uwzględnieniem Polski, wzorowano się na wybranych metodach badania znanych z literatury przedmiotu. Podejście zarówno wartościowe, jak i ilościowe stanowi wkład własny autorów w metodykę badawczą.

Zaproponowano wskaźniki jakości danych oraz wykorzystano dane dotyczące masy towarów. Analizie poddano informacje na temat obrotu towarowego między krajami unijnymi w 2017 r. pochodzące z bazy Eurostatu Comext. Badanie dynamiki obejmowało również lata: 2005, 2008, 2011 i 2014. Wyniki analizy pokazały, że ogółem udział wywozu z Polski towarów do danego kraju na obszarze UE jest różny dla danych wyrażonych wartościowo (wartość towarów) i ilościowo (masa towarów). W badaniu jakości danych lustrzanych należy zatem stosować oba podejścia.

Słowa kluczowe: handel, dane lustrzane, wewnątrzunijna wymiana towarowa, jakość danych, podejście wartościowe, podejście ilościowe

JEL: F14, C10, C82

Analysis of discrepancies in mirror data relating to the weight of goods in intra-EU trade statistics

Abstract. In the research carried out to date by the authors of the article, the assessment of the quality of mirror data in the exchange of goods between European Union (EU) countries was based on the value of goods. A similar approach is applied by many researchers. The aim of the research discussed in the article is to assess the quality of data relating to intra-EU trade based on not only the value, but also on the quantity of goods. The analysis of discrepancies in data relating to trade between EU countries, with a particular emphasis on Poland, was based on selected research methods known from literature. Both the value-based and the quantitative approach constitute the authors' contribution to the development of research methodology.

Data quality indicators were proposed and data pertaining to the weight of goods were used. Information on trade in goods between EU countries in 2017 obtained from Eurostat's Comext database was analysed. The research relating to the dynamics also covered the years 2005, 2008, 2011, and 2014. The results of the analysis demonstrated that the total share of export of goods from Poland to a given country within the EU is different for data expressed in value (value of goods) and in quantity (weight of goods). Therefore, both approaches should be applied in the study of the quality of mirror data.

Keywords: trade, mirror data, intra-EU trade of goods, data quality, value-based approach, quantitative approach

[a] Uniwersytet Szczeciński, Wydział Ekonomii, Finansów i Zarządzania.
ORCID: https://orcid.org/0000-0003-1119-0789.

[b] Uniwersytet Szczeciński, Wydział Ekonomii, Finansów i Zarządzania.
ORCID: https://orcid.org/0000-0002-7687-4041.

1. Wprowadzenie

Dane na temat handlu między krajami Unii Europejskiej (UE), a szerzej – handlu międzynarodowego, mają charakter lustrzany. Oznacza to, że są one rejestrowane w dwóch źródłach. Jak się okazuje, między danymi dotyczącymi tej samej transakcji istnieją rozbieżności. Stanowi to poważny problem w analizach makroekonomicznych, niemniej dane lustrzane umożliwiają stwierdzenie tych rozbieżności i podjęcie prób ich weryfikacji oraz rozwiązanie problemu.

Dane lustrzane w handlu zagranicznym od dawna stanowiły przedmiot zainteresowania badaczy. Wielu autorów zwracało uwagę na rozbieżność tzw. danych bilateralnych (ang. *bilateral data*, ten Cate, 2014), czyli rejestrowanych w dwóch źródłach (Ferrantino i Wang, 2008; Guo, 2010; Hamanaka, 2012; Parniczky, 1980). Jedni tłumaczyli to brakiem rzetelności danych statystyki publicznej, inni – wręcz przeciwnie – uważali, że problem wynika jedynie z błędnego zadeklarowania kierunków wywozu czy grup towarowych, i próbowali potwierdzić poprawność tych danych. Ustalono, że na rozbieżność danych wpływa sposób kwalifikowania kosztów transportu i ubezpieczenia, w tym franco granica lub port kraju dostawcy dla eksportu (FOB) oraz port polski lub franco granica polska dla importu (CIF), przez kontrahentów z różnych krajów (Carrère i Grigoriou, 2014; Tsigas, Hertel i Binkley, 1992). Należy dodać, że w statystykach unijnych wartość importu jest wykazywana na warunkach CIF, czyli łącznie z kosztami transportu i ubezpieczenia do granicy kraju odbiorcy (GUS, 2018). Część autorów jako przyczynę asymetrii danych wskazywała oszustwa podatkowo-celne (Federico i Tena, 1991; Fisman i Wei, 2004; Javorcik i Narciso, 2008).

Do oceny jakości danych lustrzanych stosuje się różne wskaźniki. Autorzy niniejszego artykułu od kilku lat badają niespójności tych danych za pomocą wskaźników proponowanych w literaturze przedmiotu oraz stworzonych samodzielnie (Markowicz i Baran, 2019b, 2019c). Wyznaczenie wskaźników jest ważne, ponieważ umożliwia uszeregowanie krajów czy grup towarowych według poziomu jakości. Ponieważ jednak takie podejście nie pozwala stwierdzić, czy poziom konkretnego wskaźnika świadczy o niskiej czy wysokiej jakości danych, autorzy szukali w literaturze propozycji oceniania jakości danych.

Badacze proponują różnorodne metody oceny jakości danych dotyczących zagranicznej wymiany towarowej o różnym zakresie przestrzennym i czasowym. Autorzy niniejszego artykułu również podejmowali badania w tym zakresie (Markowicz i Baran, 2019a, 2019d).

Tematem artykułu jest analiza rozbieżności danych w handlu między krajami UE ze szczególnym uwzględnieniem Polski. Cel badania omawianego w artykule to ocena jakości danych dotyczących obrotu wewnątrzunijnego na podstawie nie

tylko wartości, lecz także ilości towarów. Inspiracją do podjęcia badania były publikacje Morgensterna (1963), Federica i Teny (1991) oraz Ferrantina i Wanga (2008) na temat oceny jakości danych lustrzanych. Badacze ci zajmowali się pomiarem i oceną jakości danych dotyczących handlu zagranicznego. Podejście zarówno wartościowe, jak i ilościowe (wartość i masa towarów) jest wkładem własnym autorów niniejszego artykułu w metodykę badania rozbieżności danych lustrzanych.

2. Metoda badania

Za pierwszego badacza asymetrii danych w handlu zagranicznym uważa się Morgensterna (1963). Obserwował on różnice w danych dotyczących światowego eksportu i importu. Do badania par krajów zaproponował wskaźniki, spośród których dwa można zapisać następująco[1]:

$$W_{1i} = \frac{I_{ij} - E_{ji}}{I_{ij}} 100 \tag{1}$$

$$W_{2i} = \frac{E_{ij} - I_{ji}}{E_{ij}} 100 \tag{2}$$

gdzie:
I_{ij} – import kraju i deklarowany przez kraj i,
I_{ji} – import kraju j deklarowany przez kraj j,
E_{ij} – eksport kraju i deklarowany przez kraj i,
E_{ji} – eksport kraju j deklarowany przez kraj j.

Morgenstern stwierdził, że duże różnice w danych lustrzanych wskazują na brak wiarygodności danych statystycznych na temat handlu zagranicznego. Przyjął, że wartości wskaźników mniejsze niż –25% lub większe niż 25% świadczą o braku poprawności danych dotyczących wymiany handlowej analizowanej pary krajów. Uznał również, że wartość bezwzględna wskaźnika większa niż 50% wskazuje na występowanie dodatkowych przyczyn błędów poza jednostronnym uwzględnianiem kosztów transportu i cła. Znak (dodatni lub ujemny) przy wartości wskaźników może sugerować miejsce powstania niedoszacowania lub przeszacowania wartości towarów w handlu między krajami (zestawienie).

[1] Symbole w tych i kolejnych wzorach zostały zmienione w stosunku do oryginału w celu nadania im jednolitości w całym artykule.

Zestawienie. Wnioski na podstawie znaku przy wartości wskaźników (1) i (2)

Znak przy W_{1i} (1)	Znak przy W_{2i} (2)	Wniosek
+	+	przeszacowanie wartości obrotu kraju i lub niedoszacowanie kraju j w obu kierunkach (eksport i import)
+	–	przeszacowanie wartości importu lub niedoszacowanie wartości eksportu w obu krajach (i i j)
–	+	przeszacowanie wartości eksportu lub niedoszacowanie wartości importu w obu krajach (i i j)
–	–	niedoszacowanie wartości obrotu kraju i lub przeszacowanie kraju j w obu kierunkach (eksport i import)

Źródło: opracowanie własne na podstawie: Morgenstern (1963).

Odmienne stanowisko niż Morgenstern zaprezentowali Federico i Tena (1991). Uważali oni, że błędna klasyfikacja przepływów handlowych w podziale na towary lub według krajów powoduje równoległą błędną klasyfikację przeciwnego znaku w innej kategorii. Zdaniem tych badaczy problem da się wyeliminować dzięki agregacji danych, w związku z czym wiarygodność danych może zostać znacznie lepiej zweryfikowana poprzez porównanie dotyczące całkowitej wartości wymiany handlowej zarejestrowanej w krajach partnerskich w sferze handlu. Wyróżnili oni następujące grupy przyczyn rozbieżności danych: nieuniknione (koszty transportu CIF i FOB), strukturalne (niejednakowa klasyfikacja towarów) i błędy rzeczywiste (brak rejestracji z powodu przemytu, błędy w deklaracjach z powodu zaniedbania lub oszustwa, różne kursy walutowe).

Federico i Tena nie porównywali krajów parami w celu wyeliminowania błędów wynikających z położenia geograficznego, ale zastosowali wskaźnik – stosunek całkowitej wartości eksportu/importu i-tego kraju według jego statystyk do sumy tych samych przepływów zgodnie ze statystykami krajów partnerskich (kraje j):

$$W_{3i} = \frac{\sum_{j=1}^{N} E_{ij}}{\sum_{j=1}^{N} I_{ji}} \tag{3}$$

$$W_{4i} = \frac{\sum_{j=1}^{N} I_{ij}}{\sum_{j=1}^{N} E_{ji}} \tag{4}$$

Współczynniki te obrazują udział kosztów transportu, czyli współczynnik frachtu. Federico i Tena uznali, że wskaźnik eksportu kraju i (3) powinien wynosić 80–100%, a wskaźnik importu kraju i (4) – 100–120%, a następnie stwierdzili, że choć wskaźniki dla niektórych krajów nie mieszczą się w ustalonych granicach, to jednak wyniki zagregowane dla wszystkich krajów okazują się nadspodziewanie dobre i stabilne w czasie.

Ferrantino i Wang (2008) zajmowali się analizą rozbieżności w handlu dwustronnym w relacji kraj – kraj. Uznali, że przekazywane przez importerów i eksporterów dane dotyczące handlu międzynarodowego z wielu różnych powodów nie są zbieżne i mogą

się znacznie od siebie różnić. Autorzy ci badali, na podstawie wartości obrotów deklarowanych przez parę krajów, jak rozbieżności danych lustrzanych zmieniają się w czasie. Stwierdzili, że w związku z działaniami służb statystyki publicznej mającymi na celu poprawę jakości danych różnice w informacjach pochodzących z dwóch źródeł powinny się zmniejszać. Jeśli tak się nie dzieje, to znaczy, że mogą istnieć przyczyny nieoczywiste, takie jak celowe ukrywanie prawdziwych transakcji lub ich sztuczna rejestracja.

Autorzy niniejszego artykułu we wcześniejszych badaniach wykorzystywali dane na temat handlu wewnątrzunijnego w ujęciu wartościowym, zgodnie z którym analizuje się wartości wywozu i przywozu towarów deklarowane przez podmioty gospodarcze. W związku z mankamentami danych deklarowanych w ujęciu wartościowym (które wynikają z różnic kursu walut oraz formuł CIF i FOB) postanowiono wykorzystać dane o masie towarów w handlu między krajami UE. Na wzór analizy w ujęciu wartościowym przeprowadzono zatem analizę na podstawie ilości towarów i porównano wyniki tych badań. Założono, że wartości wskaźników jakości danych dotyczących handlu wewnątrzwspólnotowego poprawią się, to znaczy, że dane lustrzane wyrażone masą towarów będą bardziej zbieżne niż wyrażone ich wartością.

W badaniu wykorzystano dane statystyki publicznej pochodzące z bazy Comext, tworzonej i udostępnianej przez Eurostat. Były to informacje zarówno o wartości, jak i o masie wewnątrzwspólnotowych dostaw towarów (WDT) i wewnątrzwspólnotowego nabycia towarów (WNT) poszczególnych krajów UE w 2017 r. W celu analizy zmian w czasie poszerzono zakres danych o lata: 2005, 2008, 2011 i 2014.

Wzorując się na badaniach znanych z literatury przedmiotu, wyznaczono:

- wskaźniki (1) i (2) dla Polski w relacji z poszczególnymi krajami UE na podstawie danych z 2017 r. (Morgenstern, 1963);
- wskaźniki (3) i (4) dla poszczególnych krajów unijnych, odpowiednio dla wartości i masy towarów (Federico i Tena, 1991).

Sprawdzono także, jak rozbieżności danych lustrzanych – zarówno wartości, jak i masy WDT i lustrzanych WNT w relacji Polski z poszczególnymi krajami unijnymi – zmieniały się w czasie (Ferrantino i Wanga, 2008); do tej części badania wybrano lata: 2005, 2008, 2011, 2014 i 2017.

3. Wyniki badania

W tablicy przedstawiono wskaźniki (1) i (2) wyznaczone dla Polski w relacji z poszczególnymi krajami UE na podstawie danych za 2017 r.[2] Jak zostało wspomniane, wzorowano się na pracy Morgensterna (1963).

[2] Wskaźniki oparte na wartości transakcji pochodzą z wcześniejszego badania, którego wyniki zaprezentowano na konferencji Wielowymiarowa Analiza Statystyczna 2019, która odbyła się w Łodzi od 4 do 6 listopada 2019 r.

Tablica. Wartości wskaźników w handlu relacji Polska (kraj *i*) – kraj UE (kraj *j*) oraz UE ogółem w 2017 r.

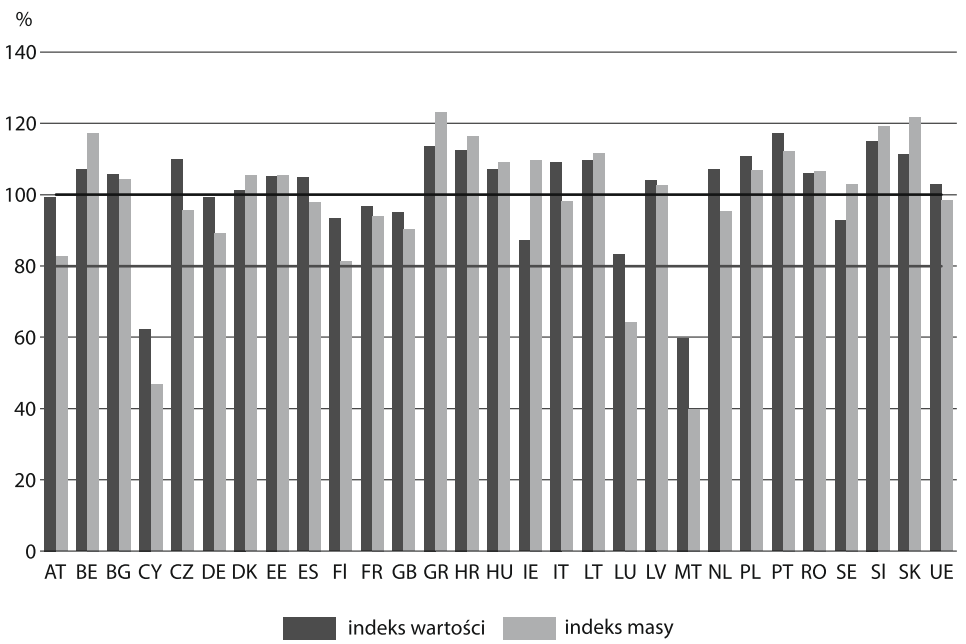
Kraje *j*	Wskaźniki dla wartości		Wskaźniki dla masy	
	W_{1i}	W_{2i}	W_{1i}	W_{2i}
	w %			
AT	–2,7	15,7	13,9	21,6
BE	–6,1	6,2	–10,1	–0,4
BG	–10,9	12,7	13,1	6,8
CY	**93,9**	**65,7**	**83,0**	**60,7**
CZ	–15,0	1,9	20,9	–3,0
DE	2,7	10,6	–6,9	8,0
DK	7,6	14,1	20,7	5,1
EE	5,4	17,7	–11,0	17,8
ES	–8,2	7,3	–8,4	9,0
FI	12,0	17,3	18,1	24,9
FR	–2,5	11,7	0,7	5,3
GB	4,6	16,4	–7,3	13,2
GR	–11,1	14,3	–17,6	3,2
HR	–13,7	–7,6	**–40,4**	–9,9
HU	–8,2	12,1	–5,9	8,1
IE	11,6	**42,8**	22,6	**29,7**
IT	–12,0	4,2	–8,2	18,7
LT	–21,1	5,4	–18,2	–7,4
LU	**39,3**	**26,5**	**36,2**	**48,2**
LV	–6,5	11,2	6,4	6,3
MT	**62,2**	**58,9**	**68,3**	–22,1
NL	–12,0	4,8	–11,0	–1,9
PT	–4,9	5,1	2,5	–2,5
RO	–11,0	–2,0	–19,4	–6,5
SE	19,5	16,9	–3,5	18,9
SI	–18,7	0,1	–16,8	8,7
SK	–15,6	8,0	–7,8	2,8
UE-27	–2,5	9,7	0,3	6,4

Uwaga. AT – Austria, BE – Belgia, BG – Bułgaria, CY – Cypr, CZ – Czechy, DE – Niemcy, DK – Dania, EE – Estonia, ES – Hiszpania, FI – Finlandia, FR – Francja, GB – Wielka Brytania, GR – Grecja, HR – Chorwacja, HU – Węgry, IE – Irlandia, IT – Włochy, LT – Litwa, LU – Luksemburg, LV – Łotwa, MT – Malta, NL – Holandia, PT – Portugalia, RO – Rumunia, SE – Szwecja, SI – Słowenia, SK – Słowacja. Pogrubieniem wyróżniono wartości bezwzględne przekraczające 25%.
Źródło: obliczenia własne.

Przeprowadzona analiza wykazała, że większość wskaźników W_1 i W_2 – zarówno dla wartości, jak i dla masy towarów – przyjmuje wartość bezwzględną poniżej 25%. Zgodnie z sugestią Morgensterna (1963) oznacza to dobrą jakość danych. Przyjęty dla wszystkich wyznaczonych miar poziom został wyraźnie przekroczony w przypadku relacji Polski z Cyprem (kolejno: 93,9%, 65,7%, 83,0% i 60,7%). Bardzo duże wartości przyjmują też wskaźniki (choć nie wszystkie) dotyczące relacji Polski z Maltą (kolejno: 62,2%, 58,9% i 68,3%). W tym przypadku jedynie wskaźnik W_2 dla masy

zawiera się w przyjętych za Morgensternem granicach normy (–22,1%). Świadczy to o niewielkich rozbieżnościach między danymi wyrażonymi masą towarów, tj. wywozem z Polski na Maltę deklarowanym przez przedsiębiorców w Polsce i przywozem na Maltę z Polski deklarowanym przez przedsiębiorców na Malcie. Przyjętą normę przekraczają także, choć w mniejszym stopniu, wskaźniki wyznaczone dla transakcji Polski z Luksemburgiem (39,3%, 26,5%, 36,2% i 48,2%) oraz po jednym wskaźniku w relacji Polska – Irlandia (W_2 dla wartości towarów wynoszące 42,8%) i Polska – Chorwacja (W_1 dla masy towarów wynoszące 40,4%).

Wykr. 1. Wskaźniki wartości i masy WDT W_{3i} (3) poszczególnych krajów UE i UE ogółem w 2017 r.



Uwaga. Jak przy tabl. 1.
Źródło: opracowanie własne.

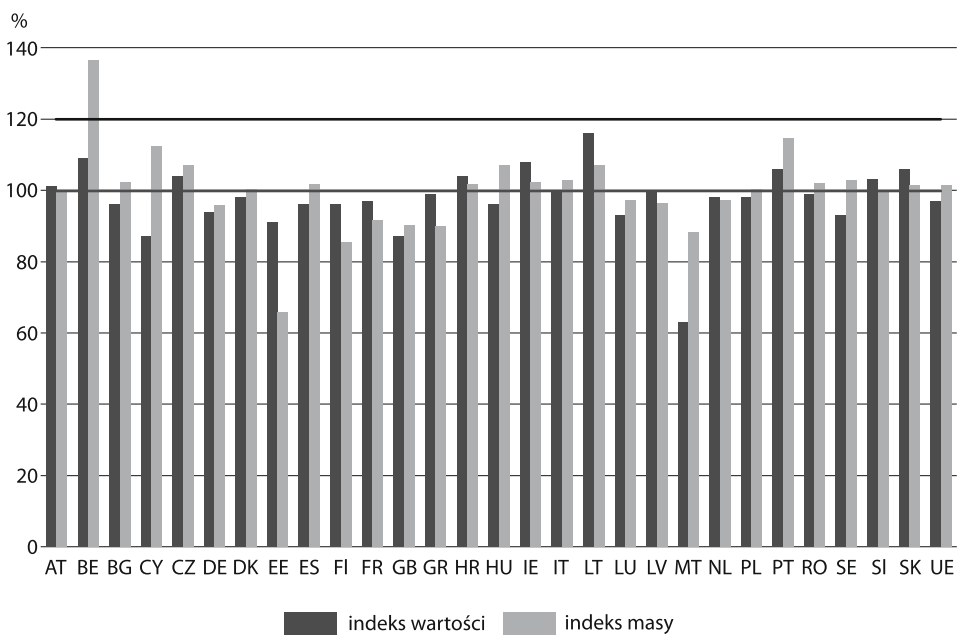
Wskaźniki (3) i (4) wyznaczono odpowiednio dla wartości i masy towarów w nawiązaniu do badań Federica i Teny (1991). Konstrukcja tych wskaźników (agregacja wartości wywozu lub przywozu w handlu krajów partnerskich) miała spowodować wynikanie dyspersji w danych z różnic w kosztach transportu (CIF i FOB), a nie z błędnego klasyfikowania towarów. Dla krajów UE wyznaczono, na podstawie danych za 2017 r., wskaźnik W_3 (3) dla WDT (wykr. 1) i wskaźnik W_4 (4) dla WNT (wykr. 2). Na wykresach zaznaczono granice wartości wskaźników określających udział kosztów transportu (współczynnik frachtu) zaproponowane przez Federica

i Tenę (1991). W przypadku wywozu jest to przedział 80–100% kosztów po stronie nabywcy, a w przypadku przywozu – 100–120%. Okazuje się, że w wielu krajach wartość wskaźnika WDT dla wartości towarów jest większa niż wartość wskaźnika WNT, co stoi w sprzeczności z wnioskami przywołanych autorów. Dla krajów UE ogółem wartości te wynoszą odpowiednio 103% i 97%. W proponowanych przedziałach znajdują się wskaźniki wartości towarów dla ośmiu (WDT) i jedenastu (WNT) krajów. Analiza masy towarów nie uwzględnia zasady odpowiedniej proporcji, ponieważ nie występuje tu problem kosztów transportu i ubezpieczenia, w związku z czym wartości wskaźników W_3 i W_4 dla masy nie powinny być różne od 100%.

Wzorując się na badaniach Ferrantina i Wanga (2008), sprawdzono, jak rozbieżności danych lustrzanych zmieniają się w czasie. Analizie poddano zmiany wartości oraz masy WDT i lustrzanego WNT w relacji Polski z poszczególnymi krajami unijnymi w latach: 2005, 2008, 2011, 2014 i 2017. Zwrócono uwagę na następujące relacje, charakterystyczne zarówno dla wartości, jak i dla masy (wykr. 3):

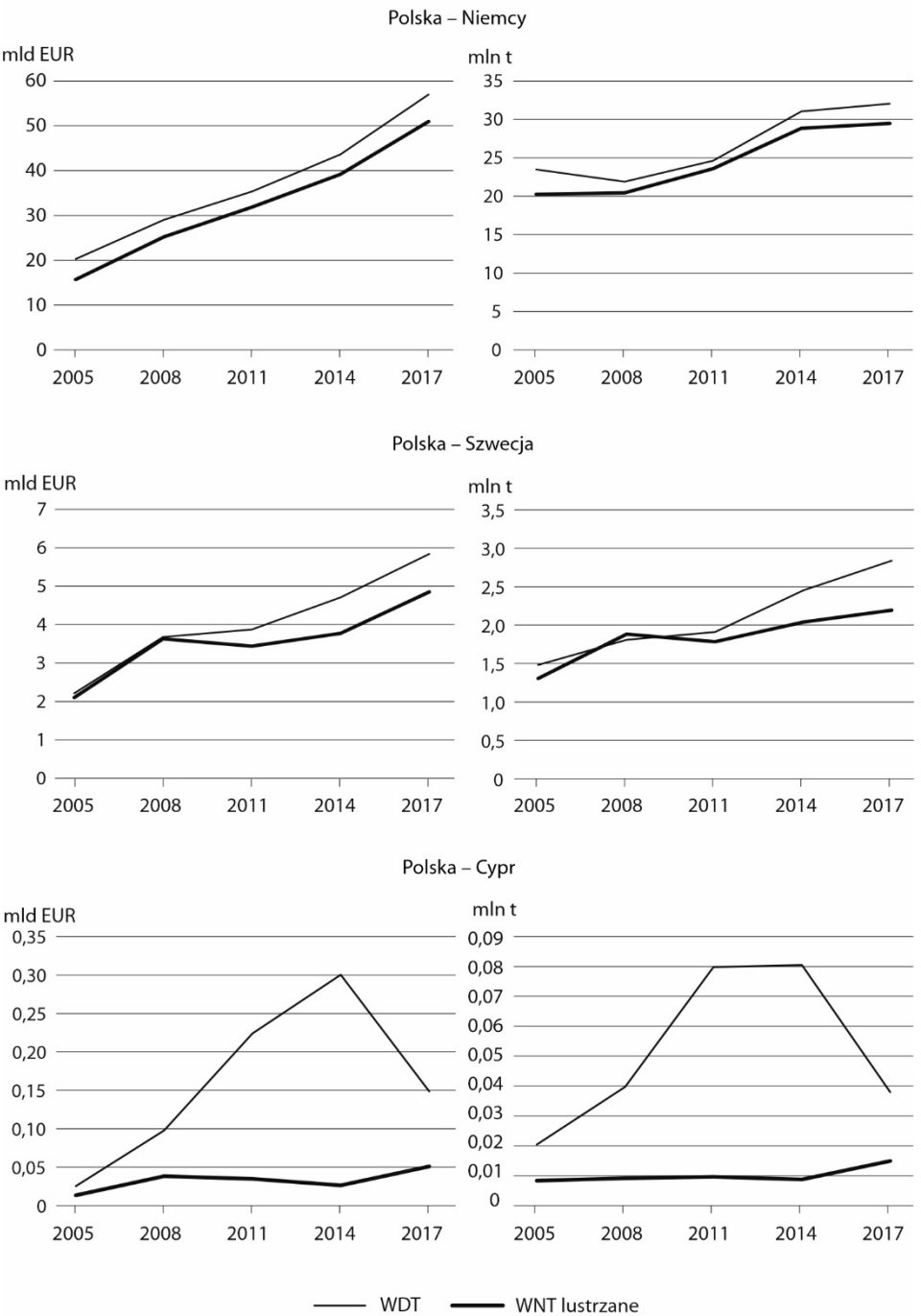
- Polska – Niemcy (PL – DE) – relacja stabilna;
- Polska – Szwecja (PL – SE) – rosnąca rozbieżność danych;
- Polska – Cypr (PL – CY) – duże i niestabilne rozbieżności.

Wykr. 2. Wskaźniki wartości i masy WNT W_{4i} (4) poszczególnych krajów UE i UE ogółem w 2017 r.



Uwaga. Jak przy tabl. 1.
Źródło: opracowanie własne.

Wykr. 3. Wartość i masa WDT i WNT lustrzanego w relacji Polska – wybrany kraj UE



Źródło: opracowanie własne.

Okazuje się, że relacje handlowe Polski z krajami UE, wyrażone zarówno wartościowo, jak i ilościowo, nie są jednolite. W przypadku gdy różnice w informacjach pochodzących z dwóch źródeł są znaczne i nie zmniejszają się w czasie, można domniemywać – podobnie jak w innych tego typu sytuacjach czynią to Ferrantino i Wang (2008) – że zachodzi tu wpływ przyczyn nieoczywistych, czyli celowego ukrywania prawdziwych transakcji lub ich sztucznej rejestracji. Należy dodać, że założenie badawcze o większej zbieżności danych lustrzanych wyrażonych masą towarów niż wyrażonych ich wartością nie zostało potwierdzone.

4. Wnioski i podsumowanie

Przegląd literatury w zakresie analiz rozbieżności danych lustrzanych w handlu zagranicznym oraz wyniki przeprowadzonych badań własnych umożliwiają sformułowanie następujących wniosków:

- literatura przedmiotu zawiera różne propozycje pomiaru jakości danych na temat zagranicznej, w tym wewnątrzunijnej, wymiany towarowej, z wykorzystaniem danych wyrażonych wartościowo;
- niewielu badaczy proponuje optymalne (czyli pozwalające na uznanie jakości danych za dobrą lub złą) zakresy wartości w odniesieniu do wybranych wskaźników;
- w analizach należy uwzględniać dane dotyczące nie tylko wartości, lecz także masy towarów będących w obrocie między krajami;
- zarówno proponowane w literaturze wskaźniki, jak i odnoszące się do nich normy można wykorzystać w badaniu innego zakresu czasowego, przestrzennego czy relacyjnego oraz można je modyfikować;
- w nowszej literaturze coraz częściej wskazuje się, że przyczyną rozbieżności danych poza względami statystycznymi, czyli systemem zbierania danych, są względy podatkowe.

Różnice między danymi lustrzanymi, czyli rejestrowanymi w dwóch źródłach, świadczą niewątpliwie o ich jakości. W przeprowadzonym badaniu wykorzystano wybrane propozycje zaczerpnięte z literatury dotyczące określania niskiej lub wysokiej jakości danych na temat wewnątrzunijnej wymiany towarowej. Wskazano także na potrzebę porównywania danych wyrażonych wartością towarów i danych wyrażonych ich masą (podejście autorskie). Przyjęto założenie o poprawie jakości danych w przypadku wyznaczania wskaźników rozbieżności dla masy towarów będących przedmiotem handlu wewnątrz UE. Zasadzało się ono na tym, że w przypadku wskaźników dla masy towarów nie występują różnice w danych związane z kosztami transportu, ubezpieczenia i kursu walut. Badanie wykazało, że wartości wskaźników w podejściach wartościowym i ilościowym różnią się, ale nie wpływa to znacząco na ogólną poprawę jakości danych.

W Polsce prowadzi się działania, które służą zarówno poprawieniu adekwatności danych statystyki publicznej, jak i ograniczeniu nieprawidłowości podatkowych, głównie związanych z VAT. Najnowsze przedsięwzięcia administracji skarbowej w tym zakresie dotyczą takich zagadnień, jak wprowadzenie jednolitego pliku kontrolnego (w 2020 r. – od kwietnia dla dużych podmiotów, a od lipca dla pozostałych), Biała Lista Podatników VAT (od stycznia 2020 r. podatnicy mają obowiązek zapłaty faktury o wartości powyżej 15 tys. zł na rachunek zamieszczony w wykazie Ministerstwa Finansów) czy Mechanizm Podzielonej Płatności (od listopada 2019 r. dla wybranych towarów w miejsce dotychczasowego odwrotnego obciążenia; VAT jest płacony na specjalny rachunek, co ma zapewnić bezpieczeństwo transakcji). Prowadzenie badań jakości danych dotyczących zagranicznego obrotu towarowego przyczynia się do wykrywania nieprawidłowości. Rozmiary tych nieprawidłowości mogą z kolei skłaniać służby statystyczne oraz podatkowe do podejmowania działań zmierzających do poprawy jakości danych.

Bibliografia

Carrère, C., Grigoriou, Ch. (2014). Can mirror data help to capture informal international trade? *Policy issues in international trade and commodities research study series*, *65*. New York: UNCTAD.

ten Cate, A. (2014). The Identification of Reporting Accuracies from Mirror Data. *Jahrbücher für Nationalökonomie und Statistik, 234*(1). DOI: 10.1515/jbnst-2014-0106.

Federico, G., Tena, A. (1991). On the Accuracy of Foreign Trade Statistics (1909–1935). Morgenstern Revisited. *Explorations in Economic History*, *28*(3), 259–273.

Ferrantino, M. J., Wang, Z. (2008). Accounting for discrepancies in bilateral trade: The case of China, Hong Kong, and the United States. *China Economic Review*, *19*(3), 502–520.

Fisman, R., Wei, S.-J. (2004). Tax Rates and Tax Evasion: Evidence from 'Missing Imports' in China. *Journal of Political Economy*, *112*(2), 471–496.

Guo, D. (2010). Mirror Statistics of International Trade in Manufacturing Goods:The Case of China. (UNIDO Working Paper No. 19/2009). Pobrane z: https://www.unido.org/api/opentext/documents/download/10079519/unido-file-10079519.

GUS. (2018). *Handel zagraniczny. Statystyka lustrzana i statystyka asymetrii*. Warszawa: Główny Urząd Statystyczny.

Hamanaka, S. (2012). Whose trade statistics are correct? Multiple mirror comparison techniques: a test of Cambodia. *Journal of Economic Policy Reform*, *15*(1), 33–56.

Javorcik, B. S., Narciso, G. (2008). Differentiated products and evasion of import tariffs. *Journal of International Economics, 76*(2), 208–222.

Markowicz, I., Baran, P. (2019a). ICA and ICS-based rankings of EU countries according to quality of mirror data on intra-Community trade in goods in the years 2014–2017. *Oeconomia Copernicana*, *10*(1), 55–68. DOI: 10.24136/oc.2019.003.

Markowicz, I., Baran, P. (2019b). Intra-Community Trade Asymmetries-Based Clustering and Linear Ordering of Combined Nomenclature Chapters Using Generalized Distance Measure (GDM). *Econometrics. Ekonometria, 23*(3), 50–58.

Markowicz, I., Baran, P. (2019c). Jakość danych dotyczących wewnątrzunijnej wymiany towarowej. *Wiadomości Statystyczne. The Polish Statistician,* (6) 5–15.

Markowicz, I., Baran, P. (2019d). Quality of Intrastat data. Comparison between 'old' and 'new' EU member states. *Acta Universitatis Lodziensis. Folia Oeconomica, 2*(341), 69–80. DOI: 10.18778/0208-6018.341.05.

Morgenstern, O. (1963). *On the Accuracy of Economic Observations. Second edition*. New Jersey: Princeton University Press.

Parniczky, G. (1980). On the Inconsistency of World Trade Statistics. *International Statistical Review*, *48*(1), 43–48.

Tsigas, M. E., Hertel, T. W., Binkley, J. K. (1992). Estimates of systematic reporting biases in trade statistics. *Economic Systems Research*, *4*(4), 297–310.

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 8, 39–56 DOI: 10.5604/01.3001.0014.3526
Statystyka w praktyce / Statistics in practice

Zróżnicowanie wykorzystania technologii informacyjno-komunikacyjnych w krajach Unii Europejskiej

Jolanta Wojnar[a]

Streszczenie. Celem badania omawianego w artykule jest ocena zróżnicowania krajów Unii Europejskiej pod względem stopnia wykorzystania technologii informacyjno-komunikacyjnych (ICT). Do analizy wybrano 15 wskaźników opisujących wykorzystanie ICT przez osoby fizyczne i gospodarstwa domowe. Dane pochodziły ze sprawozdań Głównego Urzędu Statystycznego oraz bazy Eurostatu i dotyczyły 2017 r. W analizie zróżnicowania zastosowano metodę analizy składowych głównych. Wykonano także analizę skupień za pomocą metody *k*-średnich.

Z badania wynika, że liderami w dziedzinie wykorzystania ICT są kraje skandynawskie i kraje Beneluksu. Wśród najniżej ocenionych znajdują się kraje południowej i południowo-wschodniej Europy oraz Polska.

Słowa kluczowe: analiza składowych głównych, metoda *k*-średnich, technologie informacyjno-komunikacyjne, ICT

JEL: C38, O52

Diversity in the use of information and communication technologies among European Union countries

Abstract. The aim of the research discussed in the article is to assess the diversity among European Union countries in terms of the use of information and communication technologies (ICT). Fifteen indicators describing the use of ICT by natural persons and households were selected for the analysis. The data were obtained from Statistics Poland reports and from the Eurostat database for the year 2017. The method of principal components analysis was applied in the process of analysing the diversity. Moreover, a cluster analysis based on the *k*-means method was performed.

The analysis demonstrates that Scandinavian and Benelux countries are the leaders in using ICT, while countries of southern and south-eastern Europe as well as Poland are the lowest rated.

Keywords: principal component analysis, *k*-means method, information and communication technologies

[a] Uniwersytet Rzeszowski, Kolegium Nauk Społecznych. ORCID: https://orcid.org/0000-0001-6962-4610.

1. Wprowadzenie

Technologie informacyjno-komunikacyjne (information and communication technologies, ICT) uznano w ostatnim dziesięcioleciu za kluczowy czynnik wzrostu i rozwoju gospodarczego (Campisi, Nicola, Farhadi i Mancuso, 2013; Madon, 2000; Papaioannou i Dimelis, 2007; Pejić Bach, 2014; Roztocki, Soja i Weistroffer, 2019; Roztocki i Weistroffer, 2016; Toader, Firtescu, Roman i Sorin, 2018), a także ważny czynnik warunkujący dobrobyt (Kasprzyk, 2013; Palvia, Baqir i Nemati, 2018). Rozwój ICT i coraz większa dostępność tych technologii powodują konieczność uwzględnienia ich wpływu na szeroko rozumianą konkurencyjność krajów (Kowal i Paliwoda-Pękosz, 2017; Zoroja, 2015; Zoroja i Pejić Bach, 2016).

Zakres wykorzystania ICT w poszczególnych krajach kształtuje się odmiennie, co wynika przede wszystkim z różnic kulturowych, ekonomicznych, technologicznych i społecznych (Pejić Bach, Zoroja i Bosilj Vuksic, 2013; Roztocki i in., 2019; Zoroja, 2015). Zróżnicowania te mogą występować na poziomie nie tylko kraju, lecz także regionów i mniejszych jednostek, np. miast, w związku z czym istotne staje się prowadzenie badań w zakresie oceny wykorzystania ICT także w mniejszej skali.

Literatura dotycząca wykorzystania ICT jest bardzo bogata – obejmuje publikacje naukowe, raporty (Baller, Dutta i Lanvin, 2016; ITU, 2017) i opracowania statystyczne (GUS, 2017). Dostępne są prace na temat wpływu ICT na gospodarkę zarówno na poziomie globalnym (Doong i Ho, 2012; Pick i Nishida, 2015), jak i regionalnym (Kos-Łabędowicz, 2017; Savulescu, 2015; Vicente i López, 2011) oraz krajowym (Szkudlarek i Milczarek, 2014; Talar i Kos-Łabędowicz, 2015; Tomaszewska, 2013; Wojnar, 2015).

Przedmiotem wielu opracowań jest zrównoważone społeczeństwo informacyjne (Karczmarczyk, Wątróbski, Jankowski i Ziemba, 2019; Szkudlarek i Milczarek, 2014; Ziemba, 2018, 2019). Autorzy tych prac przeanalizowali, w jaki sposób wdrażanie ICT (oceniane poprzez nakłady na te technologie, zarządzanie nimi i ich jakość oraz kulturę informacyjną) wpływa na ekonomiczny, społeczny i ekologiczny wymiar zrównoważonego rozwoju. W badaniach wykazywano dużą zależność pomiędzy wykorzystaniem ICT a zrównoważonym rozwojem zarówno w gospodarstwach domowych, jak i przedsiębiorstwach.

Pomiar dostępu do ICT i ich wykorzystania, będący jednocześnie pomiarem rozwoju społeczeństwa informacyjnego, jest zagadnieniem złożonym ze względu na mnogość wskaźników oraz problem wyboru tych, które najlepiej odwzorowują istotne zależności (Misuraca, Codagnone i Rossel, 2013; Schlichter i Danylchenko, 2014). Trudności pomiaru wynikają również ze stopnia skomplikowania tematyki i jej wieloaspektowości. Wskaźniki identyfikujące poziom dostępu do ICT wchodzą w skład różnych miar zagregowanych, konstruowanych w celu pomiaru nie tylko poziomu

rozwoju ICT, lecz generalnie – społeczeństwa informacyjnego. Można wśród nich wskazać takie miary, jak liczone dla krajów ICT Development Index (ITU, 2017), DESI (European Commission, 2018) i Networked Readiness Index (Baller i in., 2016) czy Indeks Społeczeństwa Informacyjnego ESPON, skonstruowany dla regionów NTS 2 w Europie (ESPON, 2007).

Przy dokonywaniu oceny wykorzystania ICT w krajach Unii Europejskiej (UE), regionach czy nawet przedsiębiorstwach autorzy sięgają zarówno po metody proste, oparte na wartościach pojedynczych wskaźników (Bajdor, 2015; Olszak i Ziemba, 2011; Tomaszewska, 2013; Zecevic, Radovic Stojanovic i Cudan, 2019), jak i metody wielowymiarowe, umożliwiające konstruowanie własnych agregatowych mierników oceny (Wojnar, 2015; Ziemba, 2018; Zoroja i Pejić Bach, 2015), pozwalających na porównanie krajów pod względem wdrażania ICT. Niektórzy badacze oprócz analizy statycznej, za pomocą której oceniają poziom wykorzystania ICT w wybranym momencie czasowym (Zoroja i Pejić Bach, 2016), stosują podejście dynamiczne i analizują zmiany w określonym przedziale czasowym, z wykorzystaniem danych panelowych (Toader, Firtescu, Roman i Sorin, 2018; Zoroja, 2015).

Pomiaru wykorzystania ICT w regionach dokonuje się za pomocą różnych wskaźników i różnych metod badawczych. Trwają dyskusje na temat ich wyboru i interpretacji wyników. Zastosowanie dotychczasowych metod pomiaru i zestawów wskaźników często nie daje pełnego obrazu rozwoju społeczeństwa informacyjnego, który – co warto podkreślić – przebiega dynamicznie, i nie pozwala na rzetelną ocenę zjawiska. Rodzi to konieczność ciągłego poszukiwania nowych wskaźników, metod pomiaru i analizy materiału empirycznego do oceny i porównania dostępności i wykorzystania ICT. Ograniczeniem jest jednak dostęp do danych gromadzonych przez statystykę publiczną w porównaniu z potrzebami informacyjnymi.

Celem badania omawianego w artykule jest ocena zróżnicowania krajów UE pod względem stopnia wykorzystania ICT. W badaniu zastosowano metodę analizy składowych głównych, a także wykonano analizę skupień za pomocą metody *k*-średnich.

2. Metoda badania

Analizę zróżnicowania procesu kształtowania się społeczeństwa informacyjnego w Polsce i pozostałych krajach UE przeprowadzono na podstawie 15 wskaźników opisujących wykorzystanie ICT przez osoby fizyczne i gospodarstwa domowe. Dane pochodzą z bazy Eurostatu oraz ze sprawozdań GUS (GUS, 2018) i dotyczą 2017 r. Pod uwagę wzięto wskaźniki zarówno dostępu do ICT, jak i wykorzystania tych technologii. Wybrano te wskaźniki, które obejmują różne obszary wykorzystania ICT, tj. zasoby komputerowe, internet, działalność i usługi biznesowe, handel elektroniczny, e-administrację, sieci społecznościowe online oraz umiejętności cyfrowe.

O doborze wskaźników decydowała również ich porównywalność i dostępność dla wszystkich krajów UE. W badaniu uwzględniono następujące wskaźniki (wyrażone w %):

X_1 – dostęp do internetu;
X_2 – szerokopasmowy dostęp do internetu;
X_3 – dostęp do internetu poza domem i miejscem pracy za pośrednictwem urządzeń mobilnych;
X_4 – dostęp do komputera w domu;
X_5 – regularne korzystanie z internetu;
X_6 – czytanie internetowych serwisów informacyjnych;
X_7 – wysyłanie/odbieranie poczty elektronicznej;
X_8 – korzystanie z serwisów społecznościowych;
X_9 – dokonywanie zakupów przez internet;
X_{10} – korzystanie z bankowości internetowej;
X_{11} – wyszukiwanie usług turystycznych i zakwaterowania przez internet;
X_{12} – poszukiwanie pracy lub wysłanie aplikacji przez internet;
X_{13} – wyszukiwanie informacji na stronach administracji publicznej;
X_{14} – wysyłanie wypełnionych formularzy przez internet;
X_{15} – umiejętności cyfrowe podstawowe lub powyżej podstawowych.

W badaniu zastosowano analizę składowych głównych (principal component analysis, PCA). To jedna z najwcześniej opracowanych metod wielowymiarowych, której konstrukcję zapoczątkował Pearson (1901), a rozwijali ją Hotteling (1933, 1936), Rao (1964) oraz Eriksson, Johansson i Kettaneh-Wold (1999). Ideę analizy składowych głównych stanowi redukcja znacznej liczby zmiennych do kilku nieskorelowanych wskaźników zachowujących możliwie dużą część informacji o badanym zjawisku zawartych w zmiennych pierwotnych. To metoda nieparametryczna, a więc nie wymaga założeń o rozkładzie analizowanych zmiennych, a dzięki zminimalizowaniu liczby zmiennych potrzebnych do wyjaśnienia danego zagadnienia znacząco upraszcza interpretację wyników. Znajduje zastosowanie w licznych dziedzinach nauki (Czernyszewicz, 2008; Kamińska i Andrejko, 2006; Kolasa-Więcek, 2012). Jednak tylko niewiele opracowań dotyczy zastosowania tej techniki do oceny wykorzystania ICT w krajach UE (Zamfir i Iordache, 2019). W tego typu badaniach autorzy częściej sięgają po metody porządkowania i grupowania obiektów (Dachin, 2015; Zoroja i Pejić Bach, 2016).

Istota analizy składowych głównych polega na liniowym przekształceniu wyjściowego zbioru skorelowanych cech (zmiennych) $X_1, \ldots, X_p$ w zbiór nowych nieskorelowanych zmiennych $Y_1, \ldots, Y_p$, zwanych składowymi głównymi (ang. *principal components*) (Jolliffe, 2002). To przekształcenie można zapisać w postaci układu równań:

$$\begin{aligned} Y_1 &= w_{11}X_1 + w_{21}X_2 + \ \ldots + w_{P1}X_p \\ Y_2 &= w_{12}X_1 + w_{22}X_2 + \ \ldots + w_{p2}X_p \\ &\vdots \\ Y_p &= w_{1p}X_1 + w_{2p}X_2 + \ \ldots + w_{pp}X_p \end{aligned} \tag{1}$$

W wyniku obliczeń otrzymuje się p głównych składowych. Współczynniki w_{ij}, $i, j = 1, 2, \ldots, p$ nazywa się ładunkami czynnikowymi (ang. *factor loadings*). Do dalszych analiz wybiera się te spośród nich, które zapewniają wystarczająco duży udział w wyjaśnieniu ogólnej zmienności zmiennych wyjściowych. Zazwyczaj wybiera się taką liczbę pierwszych składowych, żeby łączna wariancja była nie mniejsza niż 75% ogólnej zmienności (Morrison, 1990).

Do wyboru liczby składowych głównych stosowane są również bardziej obiektywne kryteria, takie jak kryterium Kaisera (Kaiser i Ruser, 2000), polegające na pozostawieniu do dalszej analizy tylko tych czynników, których wartości własne są większe od 1, i kryterium osypiska.

Przekształcenie zmiennych wyjściowych w nieskorelowane składowe główne jest niezaprzeczalną zaletą. To, że pierwsze kilka głównych składowych zawiera zwykle większość informacji zawartych w danych wyjściowych, można wykorzystać do uzyskania optymalnego rzutowania wielowymiarowej przestrzeni danych na płaszczyznę dwuwymiarową lub przestrzeń trójwymiarową. Takie rzuty stanowią wygodną formę graficznej prezentacji danych i są jednym z podstawowych powodów stosowania analizy składowych głównych.

Tę metodę statystyczną można wykorzystać również do grupowania badanych obiektów. Jednak w tym przypadku należy zachować ostrożność, ponieważ analiza składowych głównych nie informuje o podziale na grupy badanych obiektów. Może być użyta jedynie do wstępnej analizy przed dalszymi badaniami, takimi jak analiza skupień. W omawianym badaniu w celu pogrupowania obiektów (krajów UE) na jednorodne skupienia (grupy) przeprowadzono analizę skupień na wartościach składowych metodą k-średnich dla $k = 3$. Współrzędne wskazujące na położenie każdego obiektu w nowej przestrzeni zmiennych obliczono, mnożąc macierz standaryzowanych zmiennych wyjściowych przez wartość współczynników głównych składowych.

Interpretacja składowych polega najczęściej na analizie wkładu zmiennych pierwotnych w budowę składowej głównej. Odbywa się to poprzez porównanie modułów współczynników stojących przy danej zmiennej pierwotnej. Maksymalna wartość modułu współczynnika stojącego przy zmiennej pierwotnej wskazuje na maksymalny wkład tej zmiennej w budowę składowej głównej.

Punktem wyjścia do analizy składowych głównych jest macierz korelacji lub macierz kowariancji, ponieważ macierze zawierają całą informację niezbędną do wyzna-

czenia tych składowych. W omawianym badaniu do analizy składowych głównych użyto macierzy korelacji. Obliczenia potrzebne do przeprowadzenia analiz statystycznych oraz umieszczone w pracy wykresy zostały wykonane przy użyciu programu Statistica PL.

3. Wyniki badania

3.1. Opis i charakterystyka zmiennych

Wartości podstawowych statystyk opisowych analizowanych zmiennych, które dostarczają wielu istotnych informacji o zróżnicowaniu wykorzystania ICT w krajach UE, przedstawiono w tabl. 1.

Tabl. 1. Wartości wskaźników wykorzystania ICT w krajach UE w 2017 r.

Zmienne	Średnia	Minimum	Maksimum	Dla Polski	Współczynnik zmienności
	w %				
X_1	84,4	67 (BG)	98 (NL, DK)	82	9,6
X_2	83,2	67 (BG)	98 (NL)	78	9,7
X_3	66,4	32 (IT)	87 (NL)	40	21,0
X_4	82,2	63 (BG)	98 (NL)	82	10,6
X_5	80,1	61 (RO)	96 (LU)	73	12,6
X_6	65,3	39 (IT)	85 (LU)	60	18,5
X_7	70,6	45 (RO)	94 (DK)	60	21,4
X_8	57,9	43 (FR)	75 (DK)	48	16,7
X_9	52,6	16 (RO)	83 (GB)	45	37,3
X_{10}	52,6	5 (BG)	90 (DK)	40	44,2
X_{11}	38,5	11 (BG)	71 (LU)	23	42,5
X_{12}	16,2	5 (CZ)	29 (FI)	12	36,3
X_{13}	46,5	7 (RO)	87 (DK)	21	40,9
X_{14}	33,9	4 (RO)	72 (SE)	21	58,5
X_{15}	56,9	29 (BG)	85 (LU)	46	24,5

Uwaga. BG – Bułgaria, CZ – Czechy, DK – Dania, FI – Finlandia, FR – Francja, GB – Wielka Brytania, IT – Włochy, LU – Luksemburg, NL – Holandia, RO – Rumunia, SE – Szwecja.
Źródło: opracowanie własne na podstawie danych Eurostatu, https://ec.europa.eu/eurostat/data/database (dostęp: 8.08.2019).

W 2017 r. ponad 84% Europejczyków miało dostęp do internetu w domu, a w ponad 82% gospodarstw domowych znajdował się przynajmniej jeden komputer. W Polsce wskaźniki te były nieznacznie niższe od średniej w UE. Różnica dzieląca Polskę od przodujących pod tym względem Holandii i Danii wyniosła 16 p.proc. W Holandii odnotowano również największy odsetek gospodarstw domowych posiadających dostęp do szerokopasmowego internetu (98%). Warto podkreślić, że w tych krajach wartości omawianych zmiennych były najmniej zróżnicowane – współczynnik zmienności kształtował się na poziomie 10%. Nieco

większymi dysproporcjami charakteryzował się odsetek użytkowników, którzy regularnie (co najmniej raz w tygodniu) korzystają z internetu. Przeciętna wartość tej cechy wynosiła 80%. Z urządzeń przenośnych w celu łączenia się z internetem poza domem lub miejscem pracy w 2017 r. korzystało ponad 66% użytkowników.

W celach prywatnych internet najczęściej wykorzystywano do obsługi poczty elektronicznej (niemal 71%) oraz lektury internetowych serwisów informacyjnych (65% osób w wieku 16–74 lat). Najwyraźniejsze zróżnicowanie uwidoczniło się w przypadku odsetka osób korzystających z usług bankowych. Największe wartości tego wskaźnika odnotowano w Danii (90%), a najmniejsze – w Bułgarii (5%). W Polsce było to 40%, o niespełna 13 p.proc. mniej niż średnia w UE. Znaczne różnice można było zaobserwować pod względem popularności zakupów dokonywanych przez internet. W 2017 r. najczęściej z tej formy zakupów korzystali mieszkańcy Wielkiej Brytanii (83%), a najrzadziej – mieszkańcy Rumunii (16%). Odsetek osób robiących zakupy przez internet w Polsce był niższy od średniej unijnej i wynosił 45%.

Jednym z ważniejszych badanych zagadnień było wykorzystywanie internetu do kontaktu z organami administracji publicznej. W 2017 r. z możliwości wyszukiwania informacji na stronach urzędów skorzystało prawie 47% mieszkańców UE. Największy ich odsetek obserwowano w Danii (87%), a najmniejszy – w Rumunii (7%). Najbardziej zauważalne dysproporcje dotyczyły wysyłania przez internet wypełnionych formularzy urzędowych. Z tej formy usług korzystało przeciętnie niespełna 34% internautów w krajach UE. Zdecydowanym liderem pod tym względem jest Szwecja (72%), a ostatnie miejsce zajmuje Rumunia (4%).

Ważnym wskaźnikiem są umiejętności cyfrowe. Prawie 60% użytkowników deklarowało podstawowe lub powyżej podstawowych umiejętności cyfrowe. Najlepiej w tym zakresie radzą sobie mieszkańcy Luksemburga (85%), a najgorzej – mieszkańcy Bułgarii (29%). W Polsce co najmniej podstawowy poziom umiejętności cyfrowych deklaruje 46% osób w wieku 16–74 lat korzystających z internetu.

Z przedstawionej charakterystyki wyłania się zróżnicowany obraz wykorzystania ICT w krajach UE. Należy zauważyć, że największe wartości wskaźników osiągają Holandia, Dania, Luksemburg i Finlandia. W grupie krajów o najniższych wskaźnikach aktywności internetowej znajdują się Bułgaria i Rumunia. Pozycja Polski na tle krajów UE, pomimo dynamicznego rozwoju ICT, jest relatywnie niska. W przypadku wszystkich rozpatrywanych wskaźników dotyczących celów i obszarów wykorzystywania ICT wyniki Polski są niższe od średniej unijnej.

3.2. Wyniki analizy składowych głównych

Na kolejnym etapie analizy zweryfikowano podobieństwa między krajami w zakresie wykorzystania ICT. Odczytanie stopnia wzajemnych związków między poszczególnymi zmiennymi jest możliwe dzięki macierzy korelacji, w której podano wartości współczynnika korelacji Pearsona (tabl. 2).

Tabl. 2. Macierz korelacji między zmiennymi

Zmienne	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
X_1	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.
X_2	0,98	1,00	.	.	.	.	.	.	.	.	.	.	.	.	.
X_3	0,77	0,80	1,00	.	.	.	.	.	.	.	.	.	.	.	.
X_4	0,97	0,95	0,77	1,00	.	.	.	.	.	.	.	.	.	.	.
X_5	0,93	0,91	0,84	0,93	1,00	.	.	.	.	.	.	.	.	.	.
X_6	0,63	0,66	0,63	0,68	0,76	1,00	.	.	.	.	.	.	.	.	.
X_7	0,92	0,89	0,78	0,92	0,96	0,73	1,00	.	.	.	.	.	.	.	.
X_8	0,58	0,60	0,72	0,58	0,66	0,59	0,57	1,00	.	.	.	.	.	.	.
X_9	0,93	0,89	0,79	0,94	0,96	0,68	0,96	0,55	1,00	.	.	.	.	.	.
X_{10}	0,87	0,83	0,71	0,88	0,92	0,77	0,93	0,60	0,91	1,00	.	.	.	.	.
X_{11}	0,89	0,90	0,82	0,88	0,91	0,65	0,87	0,52	0,90	0,77	1,00	.	.	.	.
X_{12}	0,73	0,71	0,68	0,73	0,75	0,65	0,69	0,68	0,73	0,75	0,66	1,00	.	.	.
X_{13}	0,66	0,64	0,66	0,67	0,78	0,80	0,78	0,55	0,71	0,85	0,63	0,73	1,00	.	.
X_{14}	0,66	0,60	0,63	0,63	0,72	0,62	0,69	0,54	0,65	0,83	0,54	0,72	0,82	1,00	.
X_{15}	0,89	0,90	0,78	0,90	0,95	0,79	0,93	0,58	0,93	0,88	0,91	0,74	0,75	0,63	1,00

Źródło: opracowanie własne na podstawie danych Eurostatu, https://ec.europa.eu/eurostat/data/database (dostęp: 8.08.2019).

Wartość bezwzględna współczynnika korelacji Pearsona świadczy o sile zależności między zmiennymi – im wyższa, tym związek jest silniejszy. Wartości współczynników korelacji obliczonych dla poszczególnych par zmiennych wynosiły od 0,58 do 0,97, co wskazuje na występowanie silnej współzależności pomiędzy zmiennymi. Oznacza to, że tylko część informacji wnoszonej przez każdą ze zmiennych jest swoista, a reszta powtarza informację wnoszoną przez inne zmienne. W analizie PCA duże wartości współczynników są pożądaną cechą, ponieważ tylko wtedy analiza składowych głównych może skutecznie zmniejszyć liczbę wskaźników opisujących badane obiekty (tu: kraje UE). W takiej sytuacji zasadne jest przekształcenie zmiennych pierwotnych w nowe, wzajemnie nieskorelowane zmienne, zwane składowymi głównymi.

Wyznaczono wartości własne macierzy korelacji (tabl. 3), które odzwierciedlają istotność składowych głównych w wyjaśnianiu zasobów informacyjnych zmiennych wejściowych (procent wyjaśnianej zmienności zbioru danych).

Tabl. 3. Wartości własne macierzy korelacji

Składowe główne	Wartości własne		Skumulowane wartości własne	
		w % ogółu wariancji		w %
1	11,83	78,83	11,83	78,83
2	1,01	6,70	12,83	85,53
3	0,67	4,48	13,50	90,01
4	0,44	2,94	13,94	92,95
5	0,31	2,05	14,25	95,00
6	0,25	1,67	14,50	96,67
7	0,16	1,06	14,66	97,72
8	0,11	0,72	14,77	98,45
9	0,09	0,60	14,86	99,05
10	0,05	0,34	14,91	99,39
11	0,03	0,21	14,94	99,60
12	0,03	0,17	14,97	99,77
13	0,02	0,12	14,98	99,89
14	0,01	0,08	15,00	99,96
15	0,01	0,04	15,00	100,00

Źródło: opracowanie własne na podstawie danych Eurostatu, https://ec.europa.eu/eurostat/data/database (dostęp: 8.08.2019).

Na podstawie kryterium Keisera do dalszych analiz wybrano pierwsze dwie składowe główne, które przenoszą ponad 85% zmienności pierwotnych danych. Można więc z dobrym przybliżeniem analizować pierwotny zbiór danych jedynie w dwóch wymiarach. Przyjęto, że opis zróżnicowania badanych krajów przedstawiony w układzie dwóch pierwszych składowych głównych będzie czytelnym – a zarazem w miarę dokładnym – przybliżeniem podobieństwa krajów pod względem badanych cech.

Interpretacji składowych głównych dokonuje się na podstawie wartości ich współczynników (tabl. 4), które stanowią jednocześnie współczynniki korelacji liniowej pomiędzy zmiennymi wejściowymi i składowymi głównymi.

Tabl. 4. Wartość ładunków składowych głównych

Zmienne	Składowe główne	
	pierwsza	druga
X_1	–0,94	0,23
X_2	–0,93	0,26
X_3	–0,86	0,01
X_4	–0,94	0,23
X_5	–0,98	0,08
X_6	–0,80	–0,48
X_7	–0,86	–0,41
X_8	–0,69	–0,51
X_9	–0,85	0,20
X_{10}	–0,74	–0,12

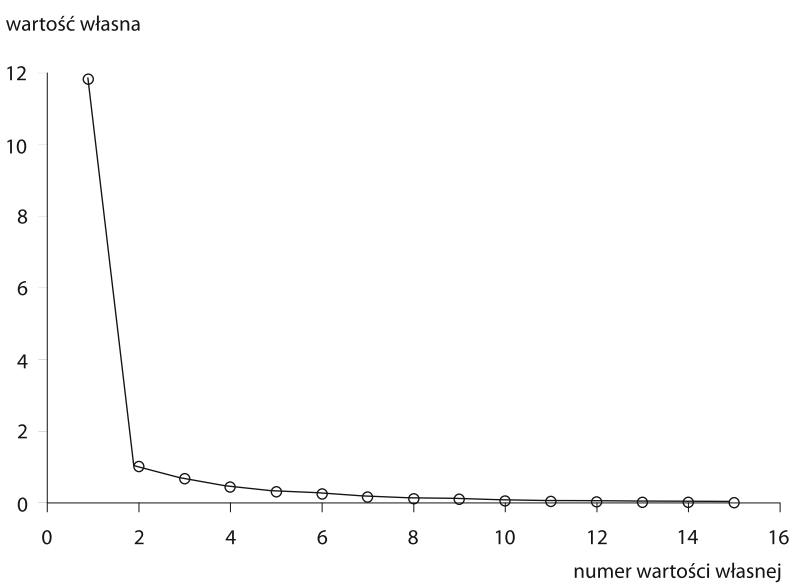
Tabl. 4. Wartość ładunków składowych głównych (dok.)

Zmienne	Składowe główne	
	pierwsza	druga
X_{11}	–0,70	0,31
X_{12}	–0,82	–0,27
X_{13}	–0,82	–0,42
X_{14}	–0,77	–0,36
X_{15}	–0,98	0,13

Źródło: opracowanie własne na podstawie danych Eurostatu, https://ec.europa.eu/eurostat/data/database (dostęp: 8.08.2019).

Decyzję o wyborze pierwszych dwóch składowych głównych potwierdza wykres osypiska (wykr. 1). Miejsce na wykresie, w którym linia malejąca przechodzi w poziomą, to tzw. koniec osypiska.

Wykr. 1. Osypisko dla wartości własnych macierzy korelacji



Źródło: opracowanie własne na podstawie danych Eurostatu, https://ec.europa.eu/eurostat/data/database (dostęp: 8.08.2019).

Pierwsza składowa główna wyjaśnia 78,83% informacji. Ma ona największe ujemne ładunki, związane przede wszystkim ze zmiennymi X_1, X_2, X_4 i X_5, charakteryzującymi dostęp do ICT (istotną rolę odgrywa zarówno wyposażenie w sprzęt informatyczny, jak i dostęp do internetu), oraz zmienną X_{15}, opisującą umiejętności cyfrowe. W skład drugiej składowej wchodzą głównie zmienne X_6, X_7, X_8 i X_{13}, związane z wykorzystaniem ICT. Druga składowa wyjaśnia 6,7% zmienności danych. Ważnym celem jest

nadanie interpretacji składowym poprzez ich powiązania ze zmiennymi wyjściowymi. Pierwszą składową można więc interpretować jako zmienną syntetyczną informującą o dostępie do ICT i umiejętnościach cyfrowych, a drugą – jako wykorzystanie ICT.

Nowe składowe są jednocześnie osiami nowej, dwuwymiarowej płaszczyzny, w której można umieścić punkty reprezentujące kraje UE. Dla każdego analizowanego kraju wyznaczono wartości obu składowych głównych; przedstawiono je na wykr. 2. Na osi poziomej zaznaczono wartości pierwszej składowej głównej, a na osi pionowej – wartości drugiej składowej.

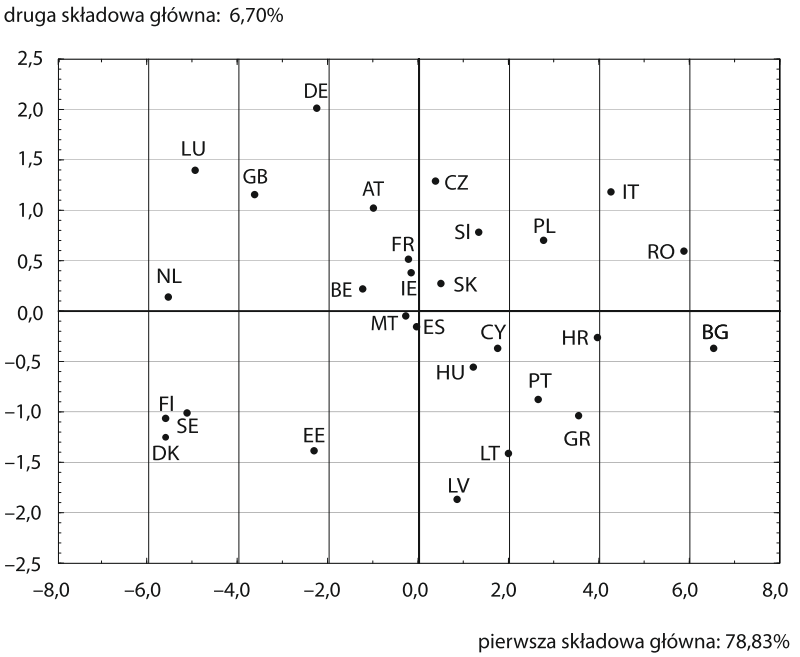
Wykres rozrzutu, pokazujący położenie obiektów względem siebie, obrazuje stopień podobieństwa krajów – im bliżej względem siebie zlokalizowane są punkty, tym bardziej podobne informacje z zakresu rozpatrywanego zestawu cech charakteryzują porównywane kraje. Graficzne przedstawienie tych odległości pozwoliło wychwycić obiekty, które odstają od innych. Na wykr. 2 wyraźnie widać, że są kraje znacznie odstające od pozostałych, jak również takie, które są do siebie podobne ze względu na zmienne reprezentowane przez składowe główne. Kraje położone najbliżej środka układu współrzędnych charakteryzują się wartościami wskaźników na poziomie średniej europejskiej.

Za pomocą analizy składowych głównych nie można jednak ustalić liczby skupień krajów podobnych pod względem rozpatrywanego zestawu cech, a równocześnie różniących się między skupieniami. Nie jest też możliwe przyporządkowanie krajów do odpowiednich skupień. Dlatego przeprowadzono analizę skupień metodą *k*-średnich, w której $k = 3$, odrębnie dla każdej składowej głównej. Uzyskano w ten sposób dwie dość podobne klasyfikacje. Wyniki grupowania dla pierwszej i drugiej składowej głównej przedstawiono na mapie (s. 51).

W miarę jednorodną grupę krajów (grupa I), którą wyróżniono na podstawie pierwszej składowej głównej, tworzą: Finlandia, Szwecja, Wielka Brytania, Dania, Niemcy, Estonia, Luksemburg i Holandia. W klasyfikacji przeprowadzonej na podstawie drugiej składowej do tego skupienia nie zaklasyfikowano Niemiec. Grupa I skupia kraje przodujące w UE pod względem rozwoju społeczeństwa informacyjnego – poziomu dostępu do ICT i ich wykorzystania. Wskaźniki wykorzystania ICT w tych krajach przyjmują maksymalne wartości. To kraje z najmniejszym ładunkiem, tj. najbardziej na lewo oddalone od pierwszej składowej głównej na wykr. 2.

Wśród krajów ocenionych najniżej (zaklasyfikowanych do grupy III) na podstawie pierwszej składowej głównej znalazły się kraje południowej i południowo-wschodniej Europy: Włochy, Chorwacja, Grecja, Portugalia, Bułgaria i Rumunia oraz Polska. Zgodnie z klasyfikacją na podstawie drugiej składowej głównej poza tą grupą znalazły się Grecja i Portugalia. Na wykr. 2 kraje z grupy III leżą po prawej stronie tej składowej. W analizowanym układzie dwóch składowych Polska ma niską pozycję na tle krajów UE.

Wykr. 2. Rzut krajów UE na płaszczyznę pierwszych dwóch składowych głównych

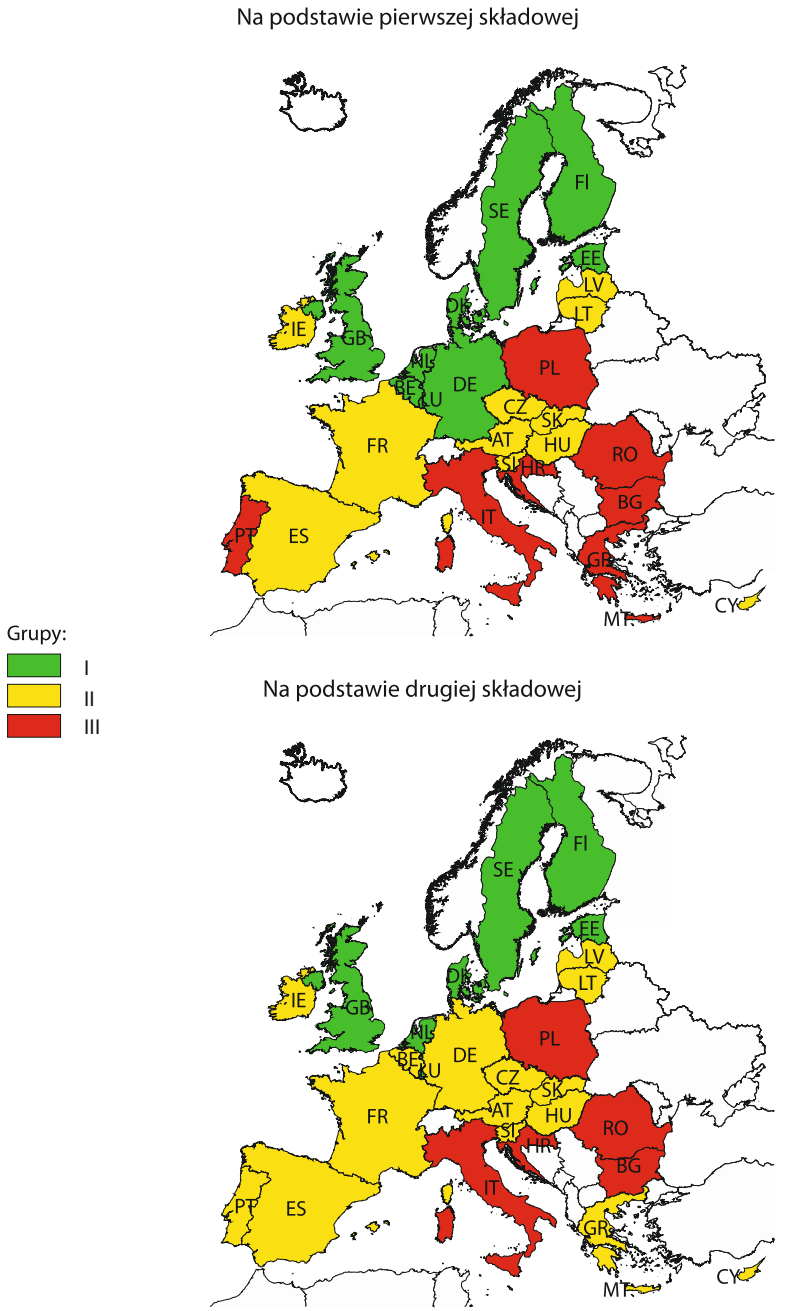


Uwaga. AT – Austria, BE – Belgia, BG – Bułgaria, CY – Cypr, CZ – Czechy, DE – Niemcy, DK – Dania, EE – Estonia, ES – Hiszpania, FI – Finlandia, FR – Francja, GB – Wielka Brytania, GR – Grecja, HR – Chorwacja, HU – Węgry, IE – Irlandia, IT – Włochy, LT – Litwa, LU – Luksemburg, LV – Łotwa, MT – Malta, NL – Holandia, PT – Portugalia, RO – Rumunia, SE – Szwecja, SI – Słowenia, SK – Słowacja, PL – Polska.

Źródło: opracowanie własne na podstawie danych Eurostatu, https://ec.europa.eu/eurostat/data/database (dostęp: 8.08.2019).

Podobne wyniki, świadczące o wysokim poziomie rozwoju krajów północnej Europy w zakresie wykorzystania ICT i niskim poziomie w krajach leżących w południowowschodniej części kontynentu, uzyskał Savulescu (2015). Badacz ten podkreśla, że „wysiłki Unii Europejskiej mają na celu zmniejszenie przepaści cyfrowej w Europie i stworzenie prawdziwego wewnętrznego rynku cyfrowego" (Savulescu, 2015, s. 518). Becker, Becker, Sulikowski i Zdziebko (2018) wykorzystali metodę k-średnich do badania krajów środkowej Europy w zakresie wykorzystania ICT w przedsiębiorstwach. Ustalili, że w badanym okresie liderami były Słowenia i Austria, a najgorzej radziła sobie Polska, wyprzedzona przez Węgry. Metodę składowych głównych oraz grupowanie krajów europejskich pod kątem wykorzystania ICT zastosowali również Zamfir i Iordache (2019), którzy przeanalizowali sytuację 35 krajów europejskich. Wyniki ich badania pokazują rozwarstwienie – Rumunia, Bułgaria, Polska, Grecja, Włochy, Turcja i Portugalia mają najniższy wskaźnik wykorzystania ICT, a najbardziej rozwinięte pod tym względem są: Norwegia, Szwecja, Finlandia, Islandia, Wielka Brytania i Niemcy. W najnowszych badaniach Polska na tle krajów UE plasuje się na jednym z ostatnich miejsc; wyprzedza jedynie Bułgarię i Rumunię.

Mapa grupowania krajów UE na podstawie pierwszych dwóch składowych głównych



Uwaga. Jak przy wykr. 2.
Źródło: opracowanie własne na podstawie danych Eurostatu, https://ec.europa.eu/eurostat/data/database (dostęp: 8.08.2019).

4. Podsumowanie

W społeczeństwie informacyjnym istotną rolę odgrywa zarówno dostęp do technologii informacyjno-komunikacyjnych (ICT), jak i ich wykorzystanie. Można zaryzykować stwierdzenie, że dostęp do ICT jest warunkiem koniecznym kształtowania się społeczeństwa informacyjnego. Nie opracowano dotąd jednolitej, powszechnie akceptowanej metodologii pomiaru rozwoju sektora informacyjnego w gospodarkach narodowych. Rodzi to potrzebę poszukiwania nowych rozwiązań w tym zakresie.

Procedura badawcza wykorzystana w badaniu omawianym w niniejszym artykule umożliwia ocenę wielokryterialnie pojmowanego zjawiska zróżnicowania poziomu wykorzystania ICT. Metoda analizy składowych głównych pozwoliła na redukcję piętnastowymiarowej przestrzeni do dwóch składowych, które wyjaśniają ponad 85% całkowitego zasobu zmienności znajdującego się w wyjściowym zbiorze cech. Pierwsza składowa główna wyjaśnia 78,83% informacji i obejmuje charakterystykę związaną z dostępem do ICT i umiejętnościami cyfrowymi. Druga składowa wyjaśnia 6,70% zmienności danych i obrazuje różne obszary wykorzystania ICT. Uznano zatem, że opis zróżnicowania badanych krajów przedstawiony w układzie pierwszych dwóch składowych głównych jest czytelnym i w miarę dokładnym przybliżeniem podobieństwa krajów pod względem badanych cech.

Przeprowadzona analiza wykazała wyraźne zróżnicowanie krajów UE w zakresie wykorzystania ICT. Badania potwierdziły występowanie przepaści cyfrowej pomiędzy liderami wśród krajów UE (zaklasyfikowanymi do grupy I), do których należą trzy kraje skandynawskie: Finlandia, Szwecja i Dania, a także Estonia, Wielka Brytania i dwa kraje Beneluksu: Luksemburg i Holandia, a najniżej sklasyfikowanymi krajami południowej i południowo-wschodniej Europy: Włochami, Chorwacją, Grecją, Portugalią, Bułgarią i Rumunią oraz Polską.

Pozycja Polski na tle krajów UE jest niska. Jeśli chodzi o wyposażanie w komputery i dostęp do internetu, Polska znajduje się blisko średniej europejskiej, natomiast pod względem zaawansowania technologicznego dysproporcje są już bardzo wyraźne.

Budowanie społeczeństwa informacyjnego to proces długi i wymagający nakładów finansowych. Obecnie zapewne można byłoby przeznaczyć znacznie większe kwoty z funduszy europejskich na skoordynowane, planowe i perspektywiczne przedsięwzięcia w Polsce. Wymaga to jednak aktywności państwa i spójnej polityki w zakresie wykorzystania ICT w zarządzaniu publicznym. Mimo wielu działań i starań Polska pod względem informatyzacji społeczeństwa plasuje się na jednym z ostatnich miejsc w Europie. Wpływa na to wiele czynników. Do najważniejszych należą: brak środków na wprowadzanie kosztownych systemów, brak odpowiednich usług elektronicznych, ograniczony dostęp do szerokopasmowego internetu oraz

mała świadomość obywateli dotycząca wykorzystania informatyki w różnych dziedzinach życia. Na podstawie przeprowadzonego badania można stwierdzić, że rozwój infrastruktury ICT powinien być priorytetem w polityce rządowej, przyczynia się on bowiem do promowania wzrostu gospodarczego kraju.

W ramach dalszych badań wskazane byłoby porównanie zmian wartości badanych wskaźników w wielu punktach czasowych, co pozwoliłoby na zobrazowanie postępów każdego kraju w budowaniu społeczeństwa informacyjnego oraz wskazanie, jaki wpływ ICT wywierają na rozwój społeczno-gospodarczy.

Bibliografia

Bajdor, P. (2015). The Use of Information and Communication Technologies in Polish Companies in Comparison to Companies from European Union. *Procedia Economics and Finance*, *27*, 702–712. DOI: 10.1016/S2212-5671(15)01051-5.

Baller, S., Dutta, S., Lanvin, B. (red.). (2016). *The Global Information Technology Report 2016*. Pobrane z: https://www.weforum.org/reports/the-global-information-technology-report-2016.

Becker, J., Becker, A., Sulikowski, P., Zdziebko, T. (2018). ANP-based analysis of ICT usage in Central European enterprises. *Procedia Computer Science*, *126*, 2173–2183. DOI: 10.1016 /j.procs.2018.07.231.

Campisi, D., Nicola, A., Farhadi, M., Mancuso, P. (2013). Discovering the impact of ICT, FDI and human capital on GDP: a cross-sectional analysis. *International Journal of Engineering Business Management*, *5*, 5–46. DOI: 10.5772/56922.

Czernyszewicz, E. (2008). Zastosowanie analizy głównych składowych do opisu konsumenckiej struktury jakości jabłek. *Żywność. Nauka. Technologia. Jakość*, *2*(57), 119–127.

Dachin, A. (2015). IT clusters in the European Union and the location significance. *Romania Journal of Regional Science, 9*(1), 54–63.

Doong, S. H., Ho, S. (2012). The Impact of ICT Development on the Global Digital Divide. *Electronic Commerce Research and Applications*, *11*(5), 518–533. DOI: 10.1016/j.elerap.2012.02.002.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S. (1999). *Introduction to Multi- and Megavariate Data Analysis Using Projection Methods* (PCA & PLS). Umea: Umetrics AB.

ESPON. (2007). *Identyfikacja istotnych przestrzennie aspektów społeczeństwa informacyjnego, Projekt ESPON 1.2.3. Raport końcowy*. Pobrane z: www.espon.eu.

European Commission. (2018). *Digital Economy and Society Index 2018 Report*. Pobrane z: https://ec.europa.eu/digital-single-market/en/news/digital-economy-and-society-index-2018-report.

GUS. (2017). *Społeczeństwo informacyjne w Polsce w 2017 r*. Pobrane z: https://stat.gov.pl /obszary-tematyczne/nauka-i-technika-spoleczenstwo-informacyjne/spoleczenstwo-informacyjne /spoleczenstwo-informacyjne-w-polsce-w-2017-roku,2,7.html.

GUS. (2018). *Społeczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2014–2018*. Pobrane z: https://stat.gov.pl/obszary-tematyczne/nauka-i-technika-spoleczenstwo-informacyjne /spoleczenstwo-informacyjne/spoleczenstwo-informacyjne-w-polsce-wyniki-badan-statystycznych -z-lat-2014-2018,1,12.html.

Hofman, A., Aravena, C., Aliaga, V. (2016). Information and Communication Technologies and Their Impact in the Economic Growth of Latin America, 1990–2013. *Telecommunications Policy*, *40*(5), 485–501. DOI: 10.1016/j.telpol.2016.02.002.

Hotteling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, *24*(6), 417–441. DOI: 10.1037/h0071325.

Hotteling, H. (1936). Simplified Computation of Principal Components. *Psychometrica, 1*, 27–35. DOI: 10.1007/BF02287921.

ITU. (2017). *Measuring the Information Society Report 2017*. Pobrane z: https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2017.aspx.

Jolliffe, I. T. (2002). *Principal component analysis* New York: Springer.

Kaiser, E. A., Ruser, R. (2000). Nitrous oxide emissions from arable soils in Germany – An evaluation of six long-term field experiments. *Journal of Plant Nutrition and Soil Science*, *163*, 249–260. DOI: 10.1002/1522-2624(200006)163:3%3C249::AID-JPLN249%3E3.0.CO;2-Z.

Kamińska, A., Andrejko, D. (2006). Zastosowanie analizy składowych głównych w badaniu wpływu nawilżania nasion łubinu żółtego odmiany radames na ich wielkość. *Inżynieria Rolnicza*, *7*(82), 241–246.

Karczmarczyk, A., Wątróbski, J., Jankowski, J., Ziemba, E. (2019). Comparative study of ICT and SIS measurement in Polish households using a MCDA-based approach. *Procedia Computer Science*, *159*, 2616–2628. DOI: 10.1016/j.procs.2019.09.254.

Kasprzyk, B. (2013). Klasyfikacja krajów UE-27 w zakresie poziomu rozwoju społeczeństwa informacyjnego. *Nierówności Społeczne a Wzrost Gospodarczy*, *35*, 237–251.

Kolasa-Więcek, A. (2012). Wykorzystanie metody PCA w analizie parametrów powiązanych z rolniczymi emisjami gazów cieplarnianych w Europie. *Journal of Research and Applications in Agricultural Engineering*, *57*(1), 77–79.

Kos-Łabędowicz, J. (2017). Wykorzystanie ICT w wybranych państwach zachodniej hemisfery. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, *336*, 10–25.

Kowal, J., Paliwoda-Pękosz, G. (2017). ICT for Global Competitiveness and Economic Growth in Emerging Economies: Economic, Cultural, and Social Innovations for Human Capital in Transition Economies. *Information Systems Management*, *34*(4), 304–307. DOI: 10.1080/10580530.2017.1366215.

Madon, S. (2000). The Internet and socio-economic development: Exploring the interaction. *Information Technology & People*, *13*(2), 85–101. DOI: 10.1108/09593840010339835.

Misuraca, G., Codagnone, C., Rossel, P. (2013). From Practice to Theory and Back to Practice: Reflexivity in Measurement and Evaluation for Evidence-based Policy Making in the Information Society. *Government Information Quarterly*, *30*, 68–82. DOI: 10.1016/j.giq.2012.07.011.

Morrison, D. F. (1990). *Wielowymiarowa analiza statystyczna*. Warszawa: Państwowe Wydawnictwo Naukowe.

Olszak, C., Ziemba, E. (2011). The Use of ICT for Economic Development in the Silesian Region in Poland. *Interdisciplinary Journal of Information, Knowledge, and Management*, *6,* 197–216. DOI: 10.28945/1392.

Palvia, P., Baqir, N., Nemati, H. (2018). ICT for socio-economic development. A citizens' perspective, *Information & Management, 55*(2), 160–176. DOI: 10.1016/j.im.2017.05.003.

Papaioannou, S. K., Dimelis, S. P. (2007). Information technology as a factor of economic development, evidence from developed and developing countries. *Economics of Innovation and New Technology*, *16*(3), 179–194. DOI: 10.1080/10438590600661889.

Pearson, K. (1901). On lines planes of closest fit to a system of points in space. *Philosophical Magazine*, *2*, 557–572. DOI: 10.1080/14786440109462720.

Pejić Bach, M. (2014). Exploring information and communications technology adoption in enterprises and its impact on innovation performance of European countries. *Ekonomickÿ casopis*, *62*(4), 335–362.

Pejić Bach, M., Zoroja, J., Bosilj Vuksic, V. (2013). Review of corporate digital divide research: A decadal analysis (2003–2012). *International Journal of Information Systems and Project Management*, *1*(4), 41–55. DOI: 10.12821/ijispm010403.

Pick, J. B., Nishida, T. (2015). Digital divides in the world and its regions: A spatial and multivariate analysis of technological utilization. *Technological Forecasting and Social Change*, *91*, 1–17. DOI: 10.1016/j.techfore.2013.12.026.

Rao, C. R. (1964). The Use and Interpretation of Principal Components in Applied Research. *Sankhya* (Series A), *26*, 329–358.

Roztocki, N., Soja, P., Weistroffer, H. R. (2019). The role of Information and Communication Technologies in socioeconomic development: towards a multi-dimensional framework. *Information Technology for Development*, *25*, 171–183. DOI: 10.1080/02681102.2019.1596654.

Roztocki, N., Weistroffer, H. R. (2016). Conceptualizing and researching the adoption of ICT and the impact on socioeconomic development. *Information Technology for Development*, *22*(4), 541–549. DOI: 10.1080/02681102.2016.1196097.

Savulescu, C. (2015). Dynamics of ICT Development in the EU. *Procedia Economics and Finance*, *23*, 513–520. DOI: 10.1016/S2212-5671(15)00552-3.

Schlichter, B. R., Danylchenko, L. (2014). Measuring ICT usage quality for information society building. *Government Information Quarterly*, *31*, 170–184. DOI: 10.1016/j.giq.2013.09.003.

Szkudlarek, P., Milczarek, A. (2014). Rola społeczeństwa informacyjnego w kreowaniu zrównoważonego rozwoju. *Ekonomia i Środowisko*, *3*(50), 231–242.

Talar, S., Kos-Łabędowicz, J. (2015). Unia Europejska wobec wyzwań rewolucji internetowej. *Przegląd Zachodni*, *1*(354), 237–249.

Toader, E., Firtescu, B. N., Roman, A., Sorin, G. A. (2018). Impact of Information and Communication Technology Infrastructure on Economic Growth: An Empirical Assessment for the EU Countries. *Sustainability*, *10,* 2–22. DOI: 10.3390/su10103750.

Tomaszewska, A. (2013). Dostęp do technologii informacyjno-komunikacyjnych w społeczeństwie informacyjnym. Przykład polskich regionów. *Acta Universitatis Lodziensis, Folia Oeconomica*, *290*, 23–37.

Vicente, M. R., López, A. J. (2011). Assessing the Regional Digital Divide across the European Union-27, *Telecommunications Policy*, *35*(3), 220–237. DOI: 10.1016/j.telpol.2010.12.013.

Wojnar, J. (2015). Tempo rozwoju ICT w Polsce oraz syntetyczna ocena dystansu Polski od krajów Unii Europejskiej w zakresie wykorzystania technologii informacyjnych. *Nierówności Społeczne a Wzrost Gospodarczy*, *44*(4), cz. 2, 372–382. DOI: 10.15584/nsawg.2015.4.2.32.

Zamfir, I. C., Iordache, A. M. (2019). A view of ICT in Europe. *Journal of Information Systems & Operations Management*, *13*(1), 170–178.

Zecevic, A., Radovic Stojanovic, J., Cudan, A. (2019). The use of Information and Communication Technologies by enterprises in the European Union Member countries. *Economic Horizons*, *21*(3), 273–285. DOI: 10.5937/ekonhor1903281Z.

Ziemba, E. (2018). Synthetic Indexes for a Sustainable Information Society: Measuring ICT Adoption and Sustainability in Polish Enterprises. *Information Technology for Management. Ongoing Research and Development*, *311*, 151–169. DOI: 10.1007/978-3-319-77721-4_9.

Ziemba, E. (2019). Synthetic Indexes for a Sustainable Information Society: Measuring ICT Adoption and Sustainability in Polish Government Units. *Information Technology for Management: Emerging Research and Applications*, *346*, 214–234. DOI: 10.1007/978-3-030-15154-6_12.

Zoroja, J. (2015). Fostering Competitiveness in European Countries with ICT: GCI Agenda. *International Journal of Engineering Business Management*, *7*, 1–8. DOI: 10.5772/60122.

Zoroja, J., Pejić Bach, M. (2016). Impact of Information and Communication Technology to the Competitiveness of European Countries – Cluster Analysis Approach. *Journal of Theoretical and Applied Electronic Commerce Research*, *11*(2), 1–10. DOI: 10.4067/S0718 -18762016000100001.

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 8, 57–66
Spisy powszechne – problemy i wyzwania / Issues and challenges in census taking

Organizacja Powszechnego Spisu Rolnego w 2020 r.
The organisation of the National Agricultural Census 2020

1. Wprowadzenie

Statystyka publiczna stale prowadzi badania ankietowe w różnych dziedzinach objętych obserwacją statystyczną oraz zbiera i opracowuje szereg sprawozdań składanych przez przedsiębiorców. Jednak najbardziej spektakularnym przedsięwzięciem podejmowanym w ramach statystyki jest przeprowadzanie co 10 lat powszechnych spisów rolnych oraz narodowych spisów powszechnych ludności i mieszkań.

Spisy powszechne są wyjątkowe pod każdym względem. Ich realizacja odbywa się pod egidą organizacji międzynarodowych, takich jak ONZ czy Eurostat, zgodnie z ich rekomendacjami i regulacjami. Regulacje te dotyczą m.in. zakresu tematycznego badania, zestawu metod zbierania danych, terminów referencyjnych ich realizacji oraz terminu, zakresu i formy przekazania wyników. Spisy obejmują całą zbiorowość, a uczestnictwo w nich jest obowiązkowe.

Na ogół spisy przeprowadza się w warunkach względnej stabilizacji społecznej i gospodarczej. Są jednak sytuacje nadzwyczajne i nie do przewidzenia, a stopień ryzyka, jakie się z nimi wiąże, zależy od rodzaju i skali zdarzenia losowego. Wieloletnie przygotowania dają poczucie, że wszystko zostało przewidziane i że opracowano procedury uwzględniające wszelkie możliwe sytuacje. Poczucie to bywa jednak złudne. Obecnie, w czasie pandemii Covid-19, pojawia się pytanie, czy w tych warunkach statystyka publiczna powinna zaniechać spisów powszechnych, czy wręcz przeciwnie – badać zmieniającą się sytuację z zastosowaniem nadzwyczajnych środków organizacyjnych i technicznych, jakich wymaga stan zagrożenia zdrowia i życia obywateli.

W niniejszym artykule opisano planową organizację Powszechnego Spisu Rolnego w 2020 r. (PSR 2020) oraz kluczowe zmiany organizacyjne, techniczne i metodologiczne wprowadzone w obliczu epidemii.

2. Organizacja zbierania danych

PSR 2020 zaplanowano jako spis pełny w indywidualnych gospodarstwach rolnych i gospodarstwach rolnych osób prawnych oraz jednostkach organizacyjnych niemających osobowości prawnej.

Wzorem spisu z 2010 r. PSR 2020 będzie przeprowadzony następującymi metodami:

- wśród osób fizycznych:
  - samospis internetowy (CAWI) jako metoda obligatoryjna w całym okresie trwania spisu, tj. od 1 września do 30 listopada 2020 r., za pośrednictwem interaktywnej aplikacji dostępnej na stronie internetowej Głównego Urzędu Statystycznego (GUS),

  - wywiad telefoniczny wspomagany komputerowo (CATI), przeprowadzany przez rachmistrzów telefonicznych w terminie od 16 września do 30 listopada 2020 r. w ramach kampanii wojewódzkiej (spis realizowany według właściwości miejscowej),
  - wywiad bezpośredni (CAPI), przeprowadzany przez rachmistrzów terenowych[1] przy użyciu urządzeń mobilnych wyposażonych w oprogramowanie przeznaczone do tego celu, w terminie od 1 października do 30 listopada 2020 r.;
- wśród osób prawnych:
  - samospis internetowy (CAWI) jako metoda obligatoryjna, przeprowadzany za pośrednictwem interaktywnej aplikacji dostępnej na Portalu Sprawozdawczym, w terminie od 1 września do 30 listopada 2020 r.,
  - wywiad telefoniczny wspomagany komputerowo (CATI), przeprowadzany przez rachmistrzów telefonicznych z wykorzystaniem zainstalowanego na komputerze oprogramowania przeznaczonego do tego celu, wyłącznie w ramach kampanii wojewódzkiej, w terminie od 1 października do 30 listopada 2020 r.

Poszczególne metody zbierania danych będą stosowane tak, aby umożliwić użytkownikom gospodarstw rolnych spełnienie obowiązku dokonania samospisu internetowego. Na bieżąco będzie monitorowany – za pomocą aplikacji zarządczej CORstat_Rol – spływ danych przez internet z wypełnionych formularzy osadzonych w interaktywnej aplikacji spisowej. Jeżeli formularze będą wypełnione niekompletnie, wymagane będzie uzupełnienie informacji przy wykorzystaniu metod CATI i CAPI.

W związku z pandemią Covid-19 przygotowano nowelizację Ustawy z dnia 31 lipca 2019 r. o powszechnym spisie rolnym w 2020 r. (dalej: ustawa o PSR 2020). Wprowadzono istotne zmiany mające na celu zapewnienie kompletności informacji. Zrównano obligatoryjność metod, co oznacza, że nie tylko samospis internetowy, lecz także wywiady telefoniczny i bezpośredni będą obowiązkowe, jeśli respondent nie dokona samospisu. Oznacza to, że respondent nie będzie mógł odmówić udzielenia informacji rachmistrzowi spisowemu, a po dokonaniu czynności spisowych z udziałem rachmistrza obowiązek udziału w spisie zostanie spełniony.

Dzięki nowoczesnej technologii zrealizowanie samospisu metodą CAWI będzie możliwe przy użyciu różnego rodzaju urządzeń komputerowych. W celu zapewnienia pełnej ochrony danych konieczne będzie jednoznaczne uwierzytelnienie się respondenta, umożliwiające zalogowanie się do interaktywnej aplikacji internetowej.

Uwierzytelnienia będzie można dokonać za pomocą jednej z kilku metod:

- przez Krajowy Węzeł Identyfikacji Elektronicznej – przy wykorzystaniu profilu zaufanego i systemów informatycznych banków;
- przez podanie numeru gospodarstwa rolnego i PESEL – numer gospodarstwa rolnego w operacie do badań rolniczych będzie służył jako login, a PESEL jako ha-

[1] W tym ankieterów statystycznych, którzy podczas spisu będą pełnić funkcję rachmistrzów spisowych.

sło, przy czym numer gospodarstwa zostanie przesłany rolnikowi w spersonalizowanym liście Prezesa GUS;
- przez podanie numeru producenta i PESEL – loginem będzie dostępny w operacie numer producenta służący do uwierzytelniania w e-wniosku ARiMR, a hasłem – PESEL;
- przez identyfikację za pośrednictwem infolinii spisowej.

W celu umożliwienia samospisu osobom fizycznym o niewystarczających umiejętnościach cyfrowych bądź niemających odpowiednich warunków technicznych gminne biuro spisowe zapewni w siedzibie gminy ogólnodostępne stanowisko komputerowe do przekazania danych przez internet, z uwzględnieniem niezbędnych zabezpieczeń wynikających z sytuacji epidemicznej. Na wniosek użytkownika gospodarstwa rolnego skierowany do gminnego komisarza spisowego przy wyznaczonym stanowisku komputerowym zapewniona będzie niezbędna pomoc w zakresie obsługi interaktywnej aplikacji do samospisu.

Dane spisowe pozyskiwane przez rachmistrzów terenowych będą przesyłane do centrali bezpiecznymi łączami teleinformatycznymi. Nad pracą rachmistrzów będą czuwać dyspozytorzy wojewódzcy (pracownicy urzędów statystycznych), którzy za pomocą systemu zarządczego CORstat_Rol będą na bieżąco monitorować postęp prac w zbieraniu danych spisowych.

Gospodarstwa rolne będą wstępnie przydzielane do badania metodami CATI i CAPI w celu zapewnienia jak najpełniejszej kompletności spisu, z uwzględnieniem specyfiki danej metody zbierania danych, m.in. w zależności od posiadania lub braku numeru telefonu, oraz na podstawie znajomości terenu. Rachmistrz terenowy zostanie więc skierowany przede wszystkim do gospodarstw rolnych, w których spis metodami CAWI lub CATI jest niemożliwy (m.in. ze względu na brak dostępu do internetu lub słaby zasięg sieci telefonii mobilnej albo brak numeru telefonu użytkownika gospodarstwa rolnego), oraz tam, gdzie istnieje konieczność uzupełnienia danych w niezakończonych formularzach wypełnianych innymi metodami.

Urządzenia mobilne, którymi będą się posługiwać rachmistrze terenowi, będą zaopatrzone w interaktywną aplikację formularzową. Aplikacja umożliwi także korzystanie z map cyfrowych, które m.in. pokażą aktualne położenie rachmistrza, położenie gospodarstwa rolnego, położenie miejsca zamieszkania użytkownika gospodarstwa rolnego i status realizacji wywiadów. W przypadku utrzymania stanu epidemii rachmistrze terenowi będą wykorzystywać te urządzenia do przeprowadzania wywiadów telefonicznych.

Użytkownicy gospodarstw rolnych przed rozpoczęciem spisu otrzymają list Prezesa GUS – Generalnego Komisarza Spisowego. Na czas zbierania danych spisowych uruchomiona zostanie infolinia spisowa w GUS i urzędach statystycznych. Konsultanci obsługujący infolinię spisową będą dostępni dla respondentów w terminie od 17 do 31

sierpnia 2020 r. przez siedem dni w tygodniu w godz. 8.00–18.00 oraz w terminie od 1 września do 30 listopada 2020 r. przez siedem dni w tygodniu w godz. 8.00–20.00.

Ponadto w ramach infolinii spisowej zostanie uruchomiona ścieżka połączeniowa „Spisz się przez telefon", dostępna przez siedem dni w tygodniu w godz. 8.00–20.00, która umożliwi respondentom natychmiastowe połączenie z dyżurującym rachmistrzem telefonicznym. To rozwiązanie wprowadzono na podstawie doświadczeń ze spisu w 2010 r.[2], a jego funkcjonalność potwierdzono w drugim spisie próbnym przed Narodowym Spisem Powszechnym Ludności i Mieszkań (NSP 2021), który przeprowadzono w kwietniu 2020 r.

3. Rachmistrze spisowi

Rachmistrze spisowi organizacyjnie i funkcjonalnie dzielą się na trzy kategorie:

- rachmistrze zbierający dane metodą CAPI (terenowi) niebędący pracownikami statystyki publicznej. Pełnienie funkcji rachmistrza spisowego wymaga: zgłoszenia się do urzędu gminy (do gminnego biura spisowego), potwierdzenia spełniania warunków określonych w ustawie o PSR 2020, udziału w szkoleniu i zdania kończącego go egzaminu z wynikiem pozwalającym na podpisanie umowy-zlecenia. W związku ze stanem epidemii Covid-19 zarówno szkolenie, jak i egzamin kandydata na rachmistrza terenowego mają formę zdalną;
- rachmistrze zbierający dane metodą CAPI (terenowi) będący ankieterami statystycznymi lub innymi pracownikami statystyki publicznej. Wszyscy ankieterzy statystyczni będą w PSR 2020 pełnili funkcję rachmistrzów spisowych, a w razie konieczności na rachmistrzów zostaną powołani inni pracownicy statystyki;
- rachmistrze zbierający dane metodą CATI (telefoniczni), będący pracownikami statystyki publicznej, wyznaczeni przez dyrektorów urzędów statystycznych – zastępców wojewódzkich komisarzy spisowych. Będą przeprowadzać wywiady telefonicznie za pomocą aplikacji wspierającej metodę CATI, zainstalowanej na wydzielonych stanowiskach pracy w urzędach statystycznych lub w przypadku utrzymania stanu epidemii – zdalnie, z domu.

Rachmistrze terenowi niebędący pracownikami statystyki publicznej zostaną wyposażeni na czas zbierania danych w urządzenia mobilne (smartfony), na których będą rejestrować informacje uzyskane podczas wywiadów. W przypadku utrzymania stanu epdemii i podjęcia przez Generalnego Komisarza Spisowego decyzji o wstrzymaniu terenowych wywiadów bezpośrednich rachmistrze terenowi będą realizować wywiady telefoniczne, z wykorzystaniem urządzeń mobilnych.

[2] Respondenci, którzy mogli uczestniczyć w tym spisie tylko w ściśle określonym czasie (np. z powodu przewidywanej dłuższej nieobecności), dzwonili na infolinię spisową i oczekiwali spisania ich w trybie natychmiastowym.

Rachmistrze terenowi będący ankieterami statystycznymi lub innymi pracownikami statystyki publicznej zostaną wyposażeni w tablety, których funkcjonalność umożliwi prowadzenie wywiadów telefonicznych w przypadku utrzymania stanu epidemii. Jako funkcjonariusze publiczni będą się posługiwać identyfikatorami rachmistrza spisowego; będzie to jeden z elementów uwierzytelniających rachmistrza.

Rachmistrze terenowi zostaną dopuszczeni do wykonywania prac spisowych:

- po przeszkoleniu z zakresu ochrony danych osobowych i otrzymaniu upoważnienia do przetwarzania danych osobowych;
- po przeszkoleniu i pouczeniu o istocie tajemnicy statystycznej oraz po złożeniu pisemnego przyrzeczenia o zachowaniu tajemnicy statystycznej.

Do głównych zadań rachmistrza terenowego będzie należało:

- przeprowadzenie wywiadów bezpośrednich lub telefonicznych, w zależności od aktualnej sytuacji związanej z pandemią Covid-19;
- zebranie danych według ustalonej metodologii i zgodnie z kluczem pytań opracowanym przez Centralne Biuro Spisowe (CBS);
- w sytuacji awaryjnej przejęcie części zadań innych rachmistrzów terenowych, np. gdy zmniejszy się liczba rachmistrzów w gminie lub kompletność zbieranych danych będzie zagrożona.

Rachmistrze telefoniczni muszą sprawnie posługiwać się komputerem, zwłaszcza obsługiwać aplikację, przy użyciu której będą wypełniali formularze spisowe, a ponadto znać funkcjonalność aplikacji CATI od strony technicznej. Powinni też cechować się umiejętnością taktownego i efektywnego przekonywania respondenta do uczestnictwa w telefonicznej formie spisu, mieć łatwość wysławiania się i używać poprawnej polszczyzny.

Na początku rozmowy z użytkownikiem gospodarstwa rolnego rachmistrz telefoniczny musi podać swoje dane identyfikacyjne (imię i nazwisko oraz urząd statystyczny, który reprezentuje). Jeśli użytkownik gospodarstwa wyrazi takie życzenie, rachmistrz musi podać numer identyfikacyjny łącznie ze sposobem weryfikacji swojej tożsamości.

Rachmistrz telefoniczny, tak samo jak terenowy, będzie się posługiwał identyfikatorem rachmistrza spisowego, który otrzyma przed rozpoczęciem prac spisowych. Dane z identyfikatora będą także stanowiły element uwierzytelniający rachmistrza, w przypadku gdy respondent zadzwoni na infolinię spisową w celu upewnienia się, że rachmistrz, który do niego zadzwonił, jest osobą upoważnioną do przeprowadzania wywiadów spisowych. Po uwierzytelnieniu się rachmistrz jest zobowiązany do poinformowania użytkownika gospodarstwa rolnego o ustawowym obowiązku przekazania danych w ramach PSR 2020 oraz o bezpieczeństwie i ochronie jego danych (statystycznych, osobowych). Wszyscy rachmistrze są zobligowani do przestrzegania tajemnicy statystycznej oraz zapisów regulacji prawnych dotyczących ochrony danych osobowych, a także do nienagannego reprezentowania służb statystyki publicznej.

Głównym zadaniem rachmistrza telefonicznego będzie przeprowadzanie wywiadów spisowych metodą CATI oraz obsługa specjalnego kanału infolinii spisowej „Spisz się przez telefon", tj. spisywanie użytkowników gospodarstw rolnych, którzy zadzwonią na infolinię i wyrażą chęć natychmiastowego spisania przez telefon. Zakłada się też, że w sytuacjach awaryjnych może wystąpić konieczność wsparcia prac w terenie przez rachmistrzów telefonicznych.

Ze względów metodologicznych i organizacyjnych, a także z uwagi na wizerunek statystyki publicznej ważne jest poinformowanie rachmistrzów o czynnościach zabronionych. Należą do nich:

- modyfikowanie treści pytań w formularzu lub sugerowanie odpowiedzi użytkownikowi gospodarstwa rolnego;
- zbieranie danych, które nie wchodzą w zakres spisu;
- poprawianie danych na formularzu bez wiedzy użytkownika gospodarstwa rolnego udzielającego odpowiedzi na pytania;
- instalowanie dodatkowego oprogramowania na urządzeniu mobilnym oraz wykorzystywanie tego urządzenia do celów innych niż prace spisowe;
- użyczanie powierzonego sprzętu osobom trzecim;
- kopiowanie, upublicznianie i wykorzystywanie w jakikolwiek sposób zebranych danych oraz udostępnianie ich osobom trzecim;
- kontaktowanie się z mediami bez wcześniejszego powiadomienia komisarza gminnego i zastępcy wojewódzkiego komisarza spisowego (m.in. z prasą, telewizją, portalami społecznościowymi i innymi internetowymi stronami informacyjnymi);
- wykonywanie podczas prac spisowych innych czynności niezwiązanych ze spisem (np. rozdawanie ulotek itp.);
- udzielanie wypowiedzi dla mediów bez wcześniejszego powiadomienia zastępcy wojewódzkiego komisarza spisowego.

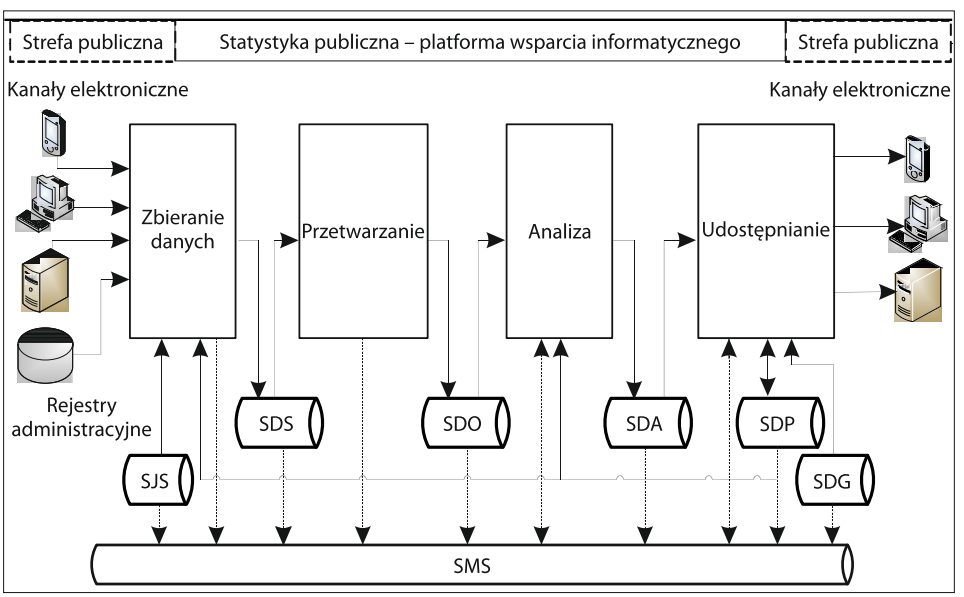
4. Architektura IT

Biorąc pod uwagę nowoczesną technologię, która może znaleźć zastosowanie w realizacji spisów powszechnych, cyfryzację poszczególnych procesów spisowych, konieczność zapewnienia szczególnych warunków w procesie zbierania i przetwarzania danych, a zwłaszcza bezwzględną konieczność ochrony danych, za priorytet uznano budowę niezawodnego, wysoce wydajnego i bezpiecznego informatycznego systemu spisowego. Obecny, tradycyjny proces produkcji statystycznej w zasadzie uniemożliwia podniesienie efektywności działań bez istotnych zmian modernizacyjnych, począwszy od gromadzenia informacji o potrzebach użytkowników, poprzez zbieranie i przetwarzanie danych, a skończywszy na udostępnianiu produktów statystycznych i zaspokojeniu rosnących potrzeb informacyjnych ich odbiorców. Opracowanie i zaimplementowanie nowej architektury wspierającej produkcję statystyczną wyma-

ga więc modernizacji obecnej infrastruktury, co jest realizowane przez pracowników statystyki z myślą o wykorzystaniu jej przy przeprowadzaniu innych badań statystycznych, organizowanych po zakończeniu spisów w 2020 i 2021 r. W ramach realizacji spisów powszechnych w latach 2020 i 2021 powstaje zmodernizowana infrastruktura teleinformatyczna, zaplanowana na bazie środowiska zwirtualizowanej infrastruktury fizycznej. Wirtualizacja infrastruktury, dzięki elastyczności rozwiązania, pozwala na budowę dowolnej architektury aplikacji i danych.

Opracowywane obecnie w GUS modele, ramy i wytyczne architektoniczne można zastosować – jak już wspomniano – zarówno do spisów powszechnych, jak i corocznie realizowanych badań statystycznych. Na schemacie przedstawiono za – sadnicze elementy architektury środowiska informatycznego dla procesowego modelu produkcji statystycznej, które będą częściowo wprowadzane w PSR 2020, a docelowo w NSP 2021. W celu objaśnienia podejścia procesowego i uogólnienia opisu funkcji struktur bazodanowych przechowujących i udostępniających kolejne „stany" danych statystycznych wynikające z tego, jak zaawansowane jest ich przetwarzanie w trakcie procesu produkcji statystycznej, wprowadzono pojęcie składnic danych.

Schemat docelowej platformy wsparcia informatycznego dla spisów powszechnych i badań statystycznych



Uwaga. Linie ciągłe na rysunku dotyczą przepływu danych, a linie przerywane – przepływu metadanych. SDA – Składnica Danych Analitycznych, SDG – Składnica Danych Geoprzestrzennych, SDO – Składnica Danych Operacyjnych, SDP – Składnica Danych Publikacyjnych, SDS – Składnica Danych Surowych, SJS – Składnica Jednostek Statystycznych, SMS – Składnica Metadanych Statystycznych.
Źródło: opracowanie własne.

Na uwagę zasługuje to, że wybuch pandemii spowodował konieczność dalszej rozbudowy infrastruktury teleinformatycznej, m.in. o funkcje pozwalające wymiennie stosować metody wywiadów bezpośrednich i telefonicznych czy świadczyć pracę w formie zdalnej. Rola składnic danych, zakres prowadzonych prac, wykorzystanie najnowszych technologii i rozwiązań architektonicznych oraz realne wsparcie informatyczne w tworzeniu narzędzi niezbędnych do realizacji spisów zasługuje na osobny artykuł poświęcony tej tematyce.

5. Niezbędne zmiany prawne

Aby możliwe było zrealizowanie celów PSR 2020 w sytuacji utrzymania stanu epidemii, konieczne było przygotowanie zmian w ustawie o PSR 2020, służących:

- zwiększeniu elastyczności stosowania metod pozyskiwania danych od użytkowników gospodarstw rolnych, w szczególności gdy stosowanie niektórych metod nie będzie możliwe. Dotyczy to przede wszystkim stworzenia ram prawnych umożliwiających elastyczne zastępowanie metody CAPI innymi metodami niewymagającymi kontaktu osobistego;
- zwiększeniu efektywności pozyskiwania danych od użytkowników gospodarstw rolnych;
- zapewnieniu możliwości zaangażowania pracowników jednostek służb statystyki publicznej do zbierania danych metodą CAPI;
- zapewnieniu możliwości zaangażowania rachmistrzów spoza grupy pracowników jednostek służb statystyki publicznej do zbierania danych metodą CATI, a nie tylko metodą CAPI.

W przepisach prawnych dotyczących spisu rolnego wprowadzony został jednoznaczny obowiązek samospisu internetowego przez cały czas trwania spisu. Równocześnie dopuszczono możliwość zebrania danych od użytkownika gospodarstwa rolnego w formie zarówno CATI, jak i CAPI i zagwarantowano, że przekazanie danych w formie CATI lub CAPI zwolni go z obowiązku samospisu internetowego. Oznacza to, że użytkownik gospodarstwa nie będzie mógł odmówić rachmistrzowi spisowemu udzielenia rzetelnych i wyczerpujących odpowiedzi. Powyższe zmiany zapewnią elastyczność stosowania metod zbierania danych, a także przyczynią się do zwiększenia jego efektywności.

Uporządkowano terminologię dotyczącą rachmistrzów – jednoznacznie zdefiniowano kategorie rachmistrzów telefonicznych i rachmistrzów terenowych. Wprowadzono możliwość zaangażowania do wywiadów przeprowadzanych metodą CATI również rachmistrzów terenowych, a do wywiadów przeprowadzanych metodą CAPI – rachmistrzów telefonicznych, co pozwala na elastyczne angażowanie rachmistrzów obydwu kategorii w zależności od potrzeb wynikających z trudnego do przewidzenia rozwoju pandemii Covid-19 i konieczności stosowania albo niestosowania którejś z metod spisowych.

6. Popularyzacja PSR 2020

Niezależnie od rozwoju pandemii kluczowa dla efektywności PSR 2020 będzie skuteczność dotarcia informacji o spisie do wszystkich respondentów. W stanie epidemii jest to o tyle ważne, że samospis internetowy może się okazać jednym z dwóch (oprócz wywiadu telefonicznego) podstawowych sposobów zbierania danych.

Za realizację ogólnopolskich działań promujących spis rolny odpowiedzialne jest CBS funkcjonujące w ramach GUS. Za koordynację i realizację działań wojewódzkich odpowiadają wojewódzkie biura spisowe, które ponadto wspierają biura gminne w jak najpełniejszej i efektywnej realizacji działań popularyzujących spis rolny na terenie gminy.

Główne hasło PSR 2020 brzmi „Spiszmy się, jak na rolników przystało", hasło uzupełniające – „Liczy się rolnictwo", a obowiązujący hasztag w postach publikowanych w mediach społecznościowych to #LiczySięRolnictwo.

Podstawowymi kanałami popularyzacji PSR 2020 są:

- strona https://spisrolny.gov.pl;
- strony internetowe urzędów statystycznych;
- strony internetowe urzędów miast i gmin;
- strony internetowe innych instytucji, ze szczególnym uwzględnieniem stron resortu rolnictwa i podległych mu agend, jednostek samorządu terytorialnego (JST), służb mundurowych, kościołów i innych związków wyznaniowych, organizacji i stowarzyszeń działających na danym terenie oraz pozostałej administracji publicznej;
- media społecznościowe GUS i urzędów statystycznych, ze szczególnym uwzględnieniem Facebooka i Twittera, oraz profile społecznościowe innych instytucji, ze szczególnym uwzględnieniem profili resortu rolnictwa i podległych mu agend, JST, organizacji i stowarzyszeń działających na danym terenie oraz pozostałej administracji publicznej;
- media informacyjne, branżowe i lifestylowe, ogólnopolskie i lokalne, ze szczególnym uwzględnieniem radia, telewizji i prasy (drukowanej i wydań online);
- kanały komunikacji udostępnione przez inne instytucje poza wymienionymi wyżej, tj. system powiadamiania SMS, e-dziennik, newsletter, mailing czy ogłoszenia duszpasterskie;
- przestrzeń publiczna – billboardy, tablice informacyjne, ekrany informacyjne w środkach transportu, reklamy obwoźne itp.;
- rozsyłanie listów pocztowych oraz mailingów, newsletterów itp.;
- spotkania bezpośrednie i zdalne, w tym spotkania z przedstawicielami JST, władz centralnych, spotkania edukacyjne, lekcje, pogadanki itp.;
- wydarzenia, np. festyny, pikniki, konkursy, webinaria itp.

Bieżące informacje na temat PSR 2020 są zamieszczane na stronie https://spisrolny.gov.pl. Ustalono też, że wszystkie informacje dotyczące spisów publikowane na Portalu

Informacyjnym GUS, w intranecie oraz na stronach urzędów statystycznych mają zawierać zapis: „więcej informacji nt. spisu rolnego na stronie https://spisrolny.gov.pl".

Działania popularyzujące spis rolny są wspierane poprzez rozpowszechnianie audycji w ramach misji publicznej nadawców publicznych TVP SA i Polskie Radio SA, co zapewnia ustawa o PSR 2020. Szczegółowy podział czasu antenowego jest określony w Rozporządzeniu Rady Ministrów z dnia 5 lutego 2020 r. w sprawie szczegółowych warunków i sposobu rozpowszechniania audycji popularyzujących powszechny spis rolny w 2020 r. Za bezpośrednią współpracę z TVP SA i Polskim Radiem SA odpowiada lider grupy ds. promocji spisów w ramach CBS, a za współpracę z ich regionalnymi spółkami odpowiadają dyrektorzy urzędów statystycznych według właściwości wojewódzkiej.

7. Podsumowanie

Realizacja spisów powszechnych niewątpliwie stanowi wyzwanie dla statystyki publicznej. W każdym kraju na świecie oznacza mobilizację służb statystyki i administracji publicznej, ponieważ badania te obejmują całe społeczeństwo. Obecne warunki realizacji spisów są wyjątkowo trudne. Zagrożenia związane z Covid-19 pojawiły się nieoczekiwanie, kiedy koncepcja realizacji PSR 2020 była już gotowa do wdrożenia. Spowodowało to konieczność opracowania na nowo wielu elementów spisu i sprawdzenia ich pod względem wykonalności.

Organizacja spisu rolnego w nowych, trudnych i nieprzewidywalnych warunkach wiąże się z koniecznością zmierzenia się z ryzykiem, a co za tym idzie – opracowania scenariuszy alternatywnych rozwiązań. Paradoksalnie okazuje się, że pandemia Covid-19 jest siłą napędową postępu w realizacji spisów powszechnych, a polska statystyka publiczna dowiodła, że może sprostać nadzwyczajnym wymaganiom.

Janusz Dygaszewicz (zastępca Generalnego Komisarza Spisowego, dyrektor Centralnego Biura Spisowego, Główny Urząd Statystyczny, dyrektor Departamentu Systemów Teleinformatycznych, Geostatystyki i Spisów)
Magdalena Janczur-Knapek (zastępca dyrektora Centralnego Biura Spisowego, Główny Urząd Statystyczny, naczelnik w Departamencie Systemów Teleinformatycznych, Geostatystyki i Spisów)

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 8, 67–69
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

XLIX Ogólnopolski Konkurs Statystyczny
49th Polish Nationwide Statistical Competition

Na początku czerwca 2020 r. zakończył się XLIX Ogólnopolski Konkurs Statystyczny (OKS). To wydarzenie edukacyjne jest organizowane od 1968 r. przez Centralną Bibliotekę Statystyczną im. Stefana Szulca (CBS) pod patronatem Prezesa Głównego Urzędu Statystycznego (GUS). Co roku w ramach konkursu uczestnicy – uczniowie szkół średnich – opracowują jeden z trzech tematów podawanych przez organizatora pod koniec lutego. Pracę należy przygotować pod kierunkiem nauczyciela na podstawie informacji zaczerpniętych z *Małego Rocznika Statystycznego Polski*, wydawanego przez GUS, i przesłać do końca kwietnia.

W tegorocznej edycji, ze względu na odbywający się w tym roku Powszechny Spis Rolny, tematy konkursowe były związane z rolnictwem i leśnictwem i brzmiały następująco:

1. Na podstawie *Małego Rocznika Statystycznego Polski* (edycja 2019 lub 2018) napisz, jakie znaczenie dla Polski mają lasy.
2. Scharakteryzuj produkcję ważniejszych wyrobów rolnych w Polsce w oparciu o *Mały Rocznik Statystyczny Polski* (edycja 2019 lub 2018).
3. Na podstawie *Małego Rocznika Statystycznego Polski* (edycja 2019 lub 2018) oceń opłacalność produkcji rolnej w Polsce.

Mimo trudnej sytuacji związanej z epidemią Covid-19 i pracą szkół w trybie zdalnym nadesłano wiele interesujących prac, charakteryzujących się wysokim poziomem merytorycznym. Uczniowie przysłali je w większości pocztą elektroniczną (z poświadczeniem autentyczności), a nie jak zazwyczaj – w wersji drukowanej. Ogółem wpłynęły 64 prace z 29 szkół w całym kraju.

Autorzy nagrodzonych prac wyróżnili się wnikliwą analizą danych zawartych w *Małym Roczniku Statystycznym*, kreatywnym podejściem do analizy statystycznej i wykorzystaniem dodatkowych narzędzi przekazu (tablice w programie PowerPoint, wywiady), dlatego wiele nagród przyznano w tym roku ex aequo. Najlepsza praca dotyczyła opłacalności produkcji rolnej w Polsce. Jej autor, Jakub Wiśniewski, nie tylko wykorzystał dane zaczerpnięte z rocznika, lecz także przeprowadził wywiady na temat produkcji rolnej z posłami z terenu swojego województwa, rolnikiem oraz ze swoimi dziadkami. Dołączył również kilkadziesiąt wykonanych samodzielnie tablic, przedstawiających analizę omawianego zjawiska statystycznego, a nawet nakręcił film udostępniony na YouTubie.

Pełna lista laureatów i przyznanych nagród, którą zamieszczamy poniżej, została opublikowana 3 czerwca br. na stronie internetowej CBS cbs.stat.gov.pl w zakładce „Konkurs":

- I nagrodę (laptop) zdobył Jakub Wiśniewski, uczeń Zespołu Szkół nr 2 w Stargardzie (praca napisana pod kierunkiem Jolanty Stępczyńskiej);
- II nagrodę (smartfon) otrzymali: Zuzanna Bazyk, uczennica Technikum Ekonomicznego nr 6 w Bydgoszczy (praca napisana pod kierunkiem Małgorzaty Sokali), Aleksandra Chuda, uczennica Zespołu Szkół Zawodowych im. Stefana Bobrowskiego w Rawiczu (praca napisana pod kierunkiem Rafała Jędrzejaka), Anna Grębowska, uczennica Zespołu Szkół Zawodowych im. Stefana Bobrowskiego w Rawiczu (praca napisana pod kierunkiem Rafała Jędrzejaka), Milena Mętel, uczennica Technikum Ekonomicznego nr 6 w Bydgoszczy (praca napisana pod kierunkiem Małgorzaty Sokali), Nikola Ulman, uczennica Zespołu Szkół im. gen. Sylwestra Kaliskiego w Górze (praca napisana pod kierunkiem Janiny Ziombry), Katarzyna Warmińska, uczennica II Liceum Ogólnokształcącego im. gen. Władysława Sikorskiego w Tomaszowie Lubelskim (praca napisana pod kierunkiem Jolanty Kidy), Karolina Witkowska, uczennica Technikum Ekonomicznego nr 6 w Bydgoszczy (praca napisana pod kierunkiem Małgorzaty Sokali);
- III nagrodę (czytnik e-booków) zdobyli: Zuzanna Gabryszak, uczennica Zespołu Szkół Ponadgimnazjalnych nr 1 w Jarocinie (praca napisana pod kierunkiem Małgorzaty Ginter), Jakub Jarosz, uczeń II Liceum Ogólnokształcącego im. gen. Władysława Sikorskiego w Tomaszowie Lubelskim (praca napisana pod kierunkiem Jolanty Kidy), Marta Sobiechowska, uczennica II Liceum Ogólnokształcącego im. Stanisława Wyspiańskiego w Legnicy (praca napisana pod kierunkiem Beaty Kuci i Jolanty Dudzic);
- IV nagrodę (album) wygrali: Kacper Drozdowski, uczeń Liceum Ogólnokształcącego im. Piotra Skargi w Sędziszowie Małopolskim (praca napisana pod kierunkiem Grzegorza Pacha), Patrycja Jędras, uczennica Zespołu Szkół Zawodowych im. Stanisława Staszica w Wysokiem Mazowieckiem (praca napisana pod kierunkiem Agnieszki Grabowskiej), Jakub Klawikowski, Maciej Kordas – uczniowie Zespołu Szkół Ekonomicznych im. Stanisława Staszica nr 5 w Słupsku (prace napisane pod kierunkiem Anny Jachurskiej), Dominika Kostro, uczennica Zespołu Szkół Zawodowych im. Stanisława Staszica w Wysokiem Mazowieckiem (praca napisana pod kierunkiem Agnieszki Grabowskiej), Dawid Królicki, uczeń Zespołu Szkół Powiatowych im. Władysława Stanisława Reymonta w Chorzelach (praca napisana pod kierunkiem Rafała Burczyńskiego), Patrycja Mozler, uczennica Zespołu Szkół Ponadpodstawowych nr 4 im. Królowej Jadwigi w Jaworznie (praca napisana pod kierunkiem Haliny Migacz), Patrycja Pawłowska, uczennica Zespołu Szkół Ekonomicznych im. Stanisława Staszica nr 5 w Słupsku (praca napisana pod kierunkiem Anny Jachurskiej), Wiktoria Pierzchalska, uczennica Centrum Kształcenia Zawodowego i Ustawicznego nr 3 „Ekonomik" w Zielonej Górze (praca napisana pod kierunkiem Aleksandry Kozikowskiej), Kinga Sobkowiak,

uczennica Zespołu Szkół Ponadgimnazjalnych nr 1 w Jarocinie (praca napisana pod kierunkiem Małgorzaty Ginter), Natalia Sobór, uczennica Zespołu Szkół Ponadpodstawowych nr 4 im. Królowej Jadwigi w Jaworznie (praca napisana pod kierunkiem Haliny Migacz);

- V nagrodę – album – otrzymali: Aleksandra Biała, Marcin Bochenek, Daria Dąbrowska, Julia Godzik, Karina Kleczek – uczeniowie Zespołu Szkół Ponadpodstawowych nr 4 im. Królowej Jadwigi w Jaworznie (prace napisane pod kierunkiem Haliny Migacz), Bartosz Kolwicz, uczeń Zespołu Szkół Powiatowych im. Władysława Stanisława Reymonta w Chorzelach (praca napisana pod kierunkiem Rafała Burczyńskiego), Weronika Kulesza, uczennica Zespołu Szkół Zawodowych im. Stanisława Staszica w Wysokiem Mazowieckiem (praca napisana pod kierunkiem Agnieszki Grabowskiej), Daria Nowak, Otylia Przepióra, Paula Sówka – uczniowie Zespołu Szkół Ponadpodstawowych nr 4 im. Królowej Jadwigi w Jaworznie (prace napisane pod kierunkiem Haliny Migacz).

Poza tym jury przyznało nagrodę specjalną (albumy i wydawnictwa GUS) Zespołowi Szkół Ponadpodstawowych nr 4 im. Królowej Jadwigi w Jaworznie za największą liczbę prac nadesłanych na tegoroczny konkurs.

Dziękujemy serdecznie wszystkim uczestnikom, gratulujemy zwycięzcy i nagrodzonym uczniom oraz zapraszamy do udziału w kolejnym, jubileuszowym L OKS, który odbędzie się w przyszłym roku.

Bożena Łazowska (Centralna Biblioteka Statystyczna im. S. Szulca)

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 8, 70–71
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

WYDAWNICTWA GUS. LIPIEC 2020
PUBLICATIONS OF STATISTICS POLAND. JULY 2020

W ofercie wydawniczej Głównego Urzędu Statystycznego z ubiegłego miesiąca warto zwrócić uwagę na następującą publikację:
Among Statistics Poland's last month's publications, we would like to recommend:

Tytuł: *Jakość życia i kapitał społeczny w Polsce. Wyniki Badania spójności społecznej 2018*
Title: *Quality of life and social capital in Poland. Results of the Social Cohesion Survey 2018*
Język: polski (dodatkowo przedmowa, spis treści i aneks wojewódzki w j. angielskim)
Language: Polish (additionally preface, contents and voivodship annex in the English version)
Dodatkowe informacje: opracowanie dostępne w wersji elektronicznej
Additional information: publication available in the electronic version

Raport analityczny, wydawany co trzy lata, poświęcony jest problematyce pomiaru jakości życia i kapitału społecznego. Opracowanie przedstawia wyniki trzeciej edycji Badania spójności społecznej, która odbyła się w 2018 r. Badanie umożliwia identyfikację poszczególnych obszarów dobrobytu społecznego i współzależności pomiędzy nimi oraz stanowi ważne źródło informacji o dobrobycie subiektywnym i postrzeganiu niektórych zjawisk społecznych. W publikacji omówiono m.in.: zróżnicowanie sytuacji materialnej gospodarstw domowych, społeczną percepcję ubóstwa, nierówności dochodowych oraz instrumentów polityki rodzinnej, przejawy religijności mieszkańców Polski, różne wymiary kapitału społecznego, postrzeganie wykluczenia społecznego i zjawiska dyskryminacji, a także wpływ niektórych czynników na poziom dobrobytu subiektywnego. Opracowanie zawiera szczegółowy opis metodologii i organizacji badania. W aneksie zamieszczono informacje o niektórych wymiarach jakości życia i kapitału społecznego na poziomie wojewódzkim.

W lipcu br. ukazały się ponadto:

- *Aktywność ekonomiczna ludności Polski I kwartał 2020 r.*;
- „Biuletyn statystyczny" nr 6/2020;
- *Ceny robót budowlano-montażowych i obiektów budowlanych (maj 2020 r.)*;

- *Koniunktura w przetwórstwie przemysłowym, budownictwie, handlu i usługach 2000–2020 (lipiec 2020). Z pogłębioną prezentacją wyników dla sekcji zakwaterowanie i gastronomia*;
- *Mały Rocznik Statystyczny Polski 2020*;
- *Polska w Unii Europejskiej* [folder];
- *Powierzchnia i ludność w przekroju terytorialnym w 2020 r.*;
- *Produkcja ważniejszych wyrobów przemysłowych w czerwcu 2020 r.*;
- *Rocznik Demograficzny 2020*;
- *Rolnictwo w 2019 r.*;
- *Sytuacja społeczno-gospodarcza kraju w I półroczu 2020 r.*;
- *Trwanie życia w 2019 r.;*
- „Wiadomości Statystyczne. The Polish Statistician" nr 7/2020;
- *Zwierzęta gospodarskie w 2019 r.*

Wersje elektroniczne wszystkich publikacji GUS są dostępne na stronie stat.gov.pl/publikacje/publikacje-a-z.

Electronic versions of all the publications by Statistics Poland are available at stat.gov.pl/en/publications.

Justyna Gustyn (Główny Urząd Statystyczny, Departament Opracowań Statystycznych)

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 polskich punktowanych czasopism naukowych Ministerstwa Nauki i Szkolnictwa Wyższego. Zgodnie z komunikatem MNiSW z dnia 31 lipca 2019 r. w sprawie wykazu czasopism naukowych i recenzowanych materiałów z konferencji międzynarodowych wraz z przypisaną liczbą punktów „WS" otrzymały 20 punktów.

„Wiadomości Statystyczne. The Polish Statistician" są udostępniane w następujących bazach indeksacyjnych i repozytoriach: Agro, BazEkon, Central and Eastern European Online Library (CEEOL), Central European Journal of Social Sciences and Humanities (CEJSH), ICI Journals Master List, ICI World of Journals, Norwegian Register for Scientific Journals and Publishers (The Nordic List) oraz POL-index.

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

1. Zgłaszanie artykułów

Prace należy przesyłać na adres: **redakcja.ws@stat.gov.pl**.

Artykuł powinien być utrzymany w formie bezosobowej i zawierać streszczenie, słowa kluczowe, kod/kody JEL oraz ORCID i afiliację autora. Tytuł, streszczenie i słowa kluczowe powinny być podane w językach polskim i angielskim.

Jeżeli w pracy występują tablice, wykresy lub mapy, powinny być umieszczone w treści artykułu. W osobnym pliku (najlepiej w formacie Excel) należy podać dane do wykresów.

Autor jest zobowiązany do podania w artykule wszelkich źródeł finansowania badań będących podstawą pracy. Jeżeli doszło do zaprezentowania podobnych materiałów podczas konferencji lub sympozjum naukowego, to podczas składania tekstu do publikacji w „WS" należy poinformować o tym redakcję.

Prosimy o niestosowanie stylów i ograniczenie formatowania do wymogów redakcyjnych. Więcej informacji w rozdziale *Wymogi redakcyjne*.

Razem z artykułem należy przesłać skan oświadczenia (do pobrania ze strony internetowej czasopisma) o oryginalności pracy i niezłożeniu jej w innym wydawnictwie, zawierającego zgodę na przeniesienie autorskich praw majątkowych, numer ORCID, afiliację lub afiliacje oraz dane kontaktowe autora, wraz ze wskazaniem proponowanego działu czasopisma. Oryginał oświadczenia należy wysłać na adres: Redakcja „Wiadomości Statystycznych. The Polish Statistician", Główny Urząd Statystyczny, al. Niepodległości 208, 00-925 Warszawa.

Załączenie skanu oświadczenia jest warunkiem poddania pracy ocenie wstępnej i skierowania do recenzji.

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. **Razem z poprawionym artykułem autor przesyła w osobnym pliku zanonimizowaną wersję pracy, przeznaczoną do recenzji.** Anonimizacja polega na utajnieniu nazwiska autora (także we właściwościach pliku), usunięciu podziękowań i informacji o źródłach finansowania, a także innych informacji wskazujących na afiliację lub umożliwiających zidentyfikowanie autora. Warunkiem skierowania pracy do recenzji jest potwierdzenie oryginalności tekstu uzyskane za pomocą systemu antyplagiatowego Similarity Check. 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.

   Autorzy artykułów, które otrzymały pozytywne oceny, wprowadzają poprawki zalecane przez recenzentów i przesyłają do redakcji 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. Znajdują się tam 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 danych przekazanych przez autora i poddawane korekcie i rewizji. Autor dokonuje ich akceptacji na etapie rewizji.

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

3. Zasady etyki publikacyjnej COPE

Redakcja „WS" dokłada wszelkich starań, aby utrzymać najwyższe standardy etyczne, 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, zespół redakcyjny, 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 zespołu redakcyjnego

1. Redakcja „WS" odpowiada za zorganizowanie i sprawny przebieg procesu wydawniczego, na który składają się: wstępna ocena zgłoszonego maszynopisu, ocena recenzentów (w przypadku artykułów naukowych), ocena KR, redakcja językowa, redakcja techniczna, skład i łamanie oraz korekta.
2. Redakcja „WS" ustala zasady obowiązujące w procesie wydawniczym, informuje jego uczestników o konieczności ich przestrzegania i egzekwuje je na każdym z jego etapów oraz dba o stałą aktualizację informacji na temat przyjętych zasad na stronie internetowej i na łamach czasopisma.
3. Redakcja nie może pozostawać w jakimkolwiek konflikcie interesów w odniesieniu do przyjmowanych artykułów. Przez konflikt interesów należy rozumieć sytuację, w której jakiekolwiek interesy lub związki (służbowe, finansowe lub inne) mogą mieć wpływ na obiektywną ocenę zgłoszonego maszynopisu lub decyzję o jego publikacji.
4. W celu przeciwdziałania nierzetelności naukowej redakcja wymaga od autorów złożenia oświadczenia, w którym deklarują oni, że zgłaszany artykuł nie narusza praw autorskich osób trzecich, nie był dotychczas publikowany i jest ich oryginalnym dziełem, oraz określają swój wkład w opracowanie artykułu.
5. Redaktorzy weryfikują zgłoszony maszynopis pod względem zgodności z celem i zakresem tematycznym czasopisma oraz spełniania wymogów redakcyjnych „WS", a także ewentual-

nych przejawów nierzetelności naukowej i możliwości wystąpienia konfliktu interesów. Obiektywną ocenę oryginalności tekstu zapewnia system antyplagiatowy stosowany przez redakcję.

6. Redakcja jest odpowiedzialna za ustalenie spójnych kryteriów oceny artykułu oraz wybór niezależnych recenzentów, którzy są zobligowani do złożenia oświadczenia o przestrzeganiu zasad etyki recenzowania COPE (publicationethics.org/resources/guidelines-new /cope-ethical-guidelines-peer-reviewers) i niewystępowaniu konfliktu interesów. Informacje dotyczące maszynopisu mogą być przekazywane przez redakcję wyłącznie autorom, recenzentom, wydawcy lub doradcom redakcyjnym.
7. W przypadku podejrzenia nadużyć redakcja postępuje zgodnie z procedurami COPE.
8. Redakcja zapewnia, że zmiany dokonane w tekście na etapie prac redakcyjnych nie naruszają zasadniczej myśli autorów.
9. Kolegium Redakcyjne, podejmując decyzję o publikacji artykułu, kieruje się wyłącznie wynikiem dyskusji dotyczącej zgłoszonego artykułu, w której uwzględniane są oceny recenzentów oraz opinie redaktorów tematycznego i merytorycznego. Rezultat ten zależy od merytorycznej oceny wartości artykułu, jego oryginalności i jasności przekazu, a także od ścisłego związku z celem i zakresem tematycznym miesięcznika.
10. W przypadku podjęcia decyzji o niepublikowaniu przesłanego materiału redakcja nie może go w żaden sposób wykorzystać bez pisemnej zgody autora.

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 na zasadach otwartego dostępu, tj. bezpłatnie i bez technicznych ograniczeń. Użytkownicy mogą czytać, pobierać, kopiować, drukować i wykorzystywać do innych celów artykuły zamieszczone online, zgodnie

z właściwymi przepisami o dozwolonym użytku, pod warunkiem wskazania źródła pochodzenia artykułu. Inne sposoby wykorzystania treści artykułów z „WS" wymagają zgody wydawcy.
3. Wydawca deklaruje gotowość do opublikowania poprawek, wyjaśnień oraz przeprosin.

4. Wymogi redakcyjne

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

Zachęcamy do przygotowania pracy z wykorzystaniem szablonu artykułu „WS" – do pobrania ze strony ws.stat.gov.pl/ForAuthors.

4.1. Struktura i zawartość artykułu

Wymagane elementy artykułu:
1. Tytuł.
2. Dane autora: imię i nazwisko, ORCID, afiliacja.
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, 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, zawierająca: przedmiot i okres badania, źródła danych i zastosowane metody badawcze, w tym uzasadnienie ich wyboru;
   - wyniki badania wraz z wnioskami;
   - 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, streszczenia, 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, tytuły rozdziałów;
   - 12 pkt – tekst główny, tytuły podrozdziałów;
   - 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). Dane, na podstawie których opracowano wykresy, należy przekazać osobno w pliku programu Excel (lub innym edytowalnym w pakiecie Microsoft Office), ewentualnie wykresy powinny dawać możliwość odczytania z nich danych.
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. Oznaczenia literowe należy zapisywać następująco: liczby i inne wielkości niezłożone – małe lub duże litery, kursywa, bez pogrubienia (np. a, A, $y(x)$, a_i); wektory – małe litery,

kursywa, pogrubione (np. $\boldsymbol{a}$, $\boldsymbol{w}$, $\boldsymbol{y}(x)$, $\boldsymbol{w}_i$); macierze – duże litery, proste, pogrubione (np. $\mathbf{A}$, $\mathbf{M}$, $\mathbf{Y}(x)$, $\mathbf{M}_i$).

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

4.3. Zasady przywoływania publikacji w treści artykułu

1. Jeden autor: bez względu na to, ile razy przywoływana jest praca, zawsze należy podać nazwisko autora i datę publikacji pracy, a w przypadku więcej niż jednej pracy danego autora opublikowanej w tym samym roku należy dodać kolejne litery alfabetu przy dacie (np. 2001a). Przykład zapisu: Jak stwierdza Iksiński (2001)... Badania wskazują, że... (Iksiński, 2001).
2. Dwóch autorów: bez względu na to, ile razy przywoływana jest praca, zawsze należy podać nazwiska obu autorów i datę publikacji pracy, a w przypadku więcej niż jednej pracy tych autorów opublikowanej w tym samym roku należy dodać kolejne litery alfabetu przy dacie. Nazwiska autorów zawsze należy łączyć spójnikiem „i", nawet w przypadku przywoływania publikacji obcojęzycznej. Przykład zapisu: Jak sugerują Iksiński i Nowak (1999)... Badania wskazują, że... (Iksiński i Nowak, 1999).
3. Od trzech do pięciu autorów: przywołanie po raz pierwszy – należy wymienić nazwiska wszystkich autorów, rozdzielając je przecinkami i stawiając spójnik „i" pomiędzy dwoma ostatnimi nazwiskami. Przy kolejnych powołaniach na tę samą pracę należy podać nazwisko pierwszego autora, a nazwiska pozostałych autorów zastąpić określeniem „i współpracownicy" (gdy wchodzą one w skład zasadniczej części zdania) lub „i in." (gdy stanowią element przypisu podanego w nawiasie). Przykład zapisu z przywołaniem po raz pierwszy: Jak sugerują Nowak, Iksiński i Jankiewicz (2003)... Badania (Nowak, Iksiński i Jankiewicz, 2003) wskazują, że... Przykład zapisu z kolejnymi przywołaniami: Badania Nowaka i współpracowników (2003)... Badania te wskazują, że... (Nowak i in., 2003).
4. Sześciu i więcej autorów: należy wymienić tylko nazwisko pierwszego autora, zarówno gdy praca przywoływana jest po raz pierwszy, jak i w późniejszych przywołaniach, a nazwiska pozostałych autorów zastąpić określeniem „i współpracownicy" (gdy wchodzą one w skład zasadniczej części zdania) lub „i in." (gdy stanowią element przypisu podanego w nawiasie). W bibliografii załącznikowej należy umieścić nazwiska wszystkich autorów pracy. Przykład zapisu: Nowakowski i współpracownicy (1997) twierdzą, że... Pierwsze badania na ten temat sugerują... (Nowakowski i in., 1997).

5. Przywoływanie jednocześnie kilku prac: należy wymienić je alfabetycznie według nazwiska pierwszego autora. Przywołania kolejnych prac muszą być oddzielone średnikiem. Lata wydania prac tego samego autora / tych samych autorów muszą być oddzielone przecinkiem. Przykład zapisu: Iksiński (2001)... Nowak i Iksiński (1999, 2005)... (Iksiński, 1997, 1999, 2004a, 2004b; Nowak i Iksiński, 1999).
6. Przywoływanie pracy za innym autorem: stosuje się w tekście, natomiast w bibliografii należy umieścić tylko pracę czytaną. Przykład zapisu: Jak wykazał Nowakowski (1990; za: Zieniecka, 2007)... Badania sugerują, że... (Nowakowski, 1990; za: Zieniecka, 2007).

4.4. Przykłady opisu bibliograficznego

1. Bibliografia powinna być zamieszczona na końcu opracowania. Prace należy podać alfabetycznie według nazwiska pierwszego autora. W przypadku dwóch lub więcej prac tego samego autora / tych samych autorów należy je uporządkować chronologicznie według roku publikacji. Jeśli kilka prac tego samego autora / tych samych autorów zostało opublikowanych w tym samym roku, należy ułożyć je alfabetycznie według tytułu i odpowiednio oznaczyć, dopisując przy roku publikacji litery a, b, c itd.
2. Artykuł w czasopiśmie, w którym każdy kolejny numer/zeszyt (issue) w ramach jednego rocznika ma osobną numerację stron (w każdym zeszycie pierwsza strona opatrzona jest numerem 1): Nazwisko, X., Nazwisko 2, X. Y., Nazwisko 3, Z. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*(zeszyt), strona początku–strona końca.
3. Artykuł w czasopiśmie, w którym kolejne numery/zeszyty (issues) w ramach jednego rocznika nie mają osobnej numeracji stron (pierwsza strona w kolejnym zeszycie opatrzona jest numerem kolejnym po ostatniej stronie w zeszycie poprzednim): Nazwisko, X., Nazwisko 2, X. Y., Nazwisko 3, Z. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*, strona początku–strona końca. Jeśli artykuł ma numer DOI (Digital Object Identifier), należy podać go na końcu opisu bibliograficznego: Nazwisko, X., Nazwisko 2, X. Y. (rok). Tytuł artykułu. *Tytuł czasopisma*, *rocznik*, strona początku–strona końca. DOI: xxxxx.
4. Książka: Nazwisko, X., Nazwisko 2, X. Y. (rok). *Tytuł książki*. Miejsce wydania: Wydawnictwo.
5. Książka napisana pod redakcją: Nazwisko, X. (red.). (rok). *Tytuł książki*. Miejsce wydania: Wydawnictwo.
6. Rozdział w pracy zbiorowej: Nazwisko, X. (rok). Tytuł rozdziału. W: Y. Nazwisko, B. Nazwisko 2 (red.), *Tytuł książki* (s. strona początku–strona końca). Miejsce wydania: Wydawnictwo.
7. Jeśli dany tekst znajduje się na stronie internetowej i nie jest artykułem w czasopiśmie, książką ani rozdziałem w książce, należy podać autora, datę publikacji (jeśli jest znana), tytuł, a następnie zamieścić informację o stronie, z której został pobrany, oraz – jeśli są to materiały informacyjne – datę dostępu. Tekst: Nazwisko, X. (rok). *Tytuł tekstu*. Pobrane z: adres strony internetowej (dostęp: DD.MM.RRRR).

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

ZAKRES TEMATYCZNY DZIAŁÓW
THEMATIC SCOPE OF SECTIONS

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

Studia metodologiczne

W tym dziale zamieszczane są artykuły naukowe przedstawiające teoretyczne rozwiązania metodologiczne ze wskazaniem ich praktycznej użyteczności, w tym prace przeglądowe i porównawcze oraz dotyczące etyki w statystyce. Poruszane w nich zagadnienia obejmują różne dziedziny statystyki, ekonomii matematycznej i ekonometrii. Omawiane rezultaty badawcze mogą znaleźć efektywne zastosowanie w badaniach empirycznych oraz analizach statystycznych i służyć podnoszeniu ich jakości, jak również powiększeniu zasobu informacyjnego.

Statystyka w praktyce

Dział ten zawiera artykuły poświęcone nowatorskim zastosowaniom w praktyce znanych narzędzi i modeli statystycznych oraz analizie i ocenie statystycznej zjawisk społeczno-ekonomicznych i innych; zamieszczane tu prace opierają się w szczególności na danych pochodzących z zasobów statystyki publicznej. Zastosowania w praktyce obejmują również wykorzystanie narzędzi informatycznych do uzyskiwania i przetwarzania informacji statystycznych, naliczania danych wynikowych, ich prezentacji i rozpowszechniania. Może to też dotyczyć opracowań stosujących nowoczesne techniki programistyczne pozwalające na efektywną komunikację z systemami informacyjnymi oraz ułatwiające wykorzystanie danych wynikowych. Publikowane są także artykuły sygnalizujące problemy związane z projektowaniem badań statystycznych, uzyskiwaniem, integracją i przetwarzaniem danych oraz generowaniem wynikowych informacji statystycznych i kontrolą ich ujawniania wraz z propozycjami efektywnych rozwiązań w tym zakresie.

Studia interdyscyplinarne. Wyzwania badawcze

To blok tematyczny zawierający artykuły wskazujące i podejmujące wyzwania badawcze, które są szczególnie istotne ze względu na rosnące potrzeby współczesnych użytkowników danych statystycznych i wymagają zaangażowania znacznych nakładów pracy, środków oraz rozwiązań z różnych dziedzin nauki i techniki. W dziale tym publikowane są również opracowania dotyczące: wykorzystania technologii informacyjnych i komunikacyjnych (ICT), gospodarki opartej na wiedzy, problematyki innowacyjności, przepływu informacji we współczesnym społeczeństwie oraz przetwarzania i analizy zagadnień związanych z data science i big data, a zatem problematyki bardzo często powiązanej z działaniami interdyscyplinarnymi.

Edukacja statystyczna

W tym dziale zamieszczane są artykuły dotyczące metod i efektów nauczania statystyki oraz popularyzacji myślenia statystycznego. Odnosi się to zwłaszcza do problemów związanych z kształceniem w zakresie umiejętności stosowania statystyki na wszystkich poziomach edukacji, a także do wykorzystywania nowoczesnych koncepcji i metod dydaktycznych oraz pomocy naukowych w nauczaniu statystyki. Uwaga skoncentrowana jest na rozumieniu prawdopodobieństwa i statystyki, badaniach z zakresu nauczania statystyki, postaw i zachowań społecznych w odniesieniu do tej dziedziny wiedzy, jak również na rozumieniu informacji statystycznych. Ponadto ukazywane są problemy związane z prezentacją danych statystycznych oraz ich interpretacją w powszechnym obiegu informacyjnym, np. w środkach społecznego przekazu.

Z dziejów statystyki

Prace publikowane w tym dziale poświęcone są historii prowadzenia obserwacji statystycznych oraz rozwoju ich metodologii i narzędzi. Ponadto zamieszczane są tu informacje dotyczące życia i osiągnięć zawodowych wybitnych statystyków, jak również najważniejszych instytucji i organizacji statystycznych w Polsce i za granicą.

Dyskusje. Recenzje. Informacje

Jedyny dział zawierający teksty nierecenzowane i niemające charakteru artykułów naukowych. Obejmuje informacje o najważniejszych wydarzeniach dotyczących statystyki polskiej i międzynarodowej, a także sprawozdania z konferencji naukowych, recenzje książek i opracowań z zakresu statystyki i jej zastosowań, rekomendacje nowych, istotnych i ciekawych pozycji wydawniczych z tego obszaru wiedzy, jak również odpowiedzi autorów na recenzje oraz polemiki, dyskusje i sprostowania dotyczące artykułów zamieszczonych na łamach czasopisma.