Cena 13,00 zł
(VAT 8%)

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

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

STYCZEŃ / JANUARY
ROK / VOLUME 65

2020 | 1

GŁÓWNY URZĄD STATYSTYCZNY
STATISTICS POLAND

POLSKIE TOWARZYSTWO STATYSTYCZNE
POLISH STATISTICAL ASSOCIATION

GUS

TPS

WIADOMOŚCI STATYSTYCZNE

THE POLISH STATISTICIAN

STYCZEŃ / JANUARY
ROK / VOLUME 65

2020 | 1 (704)

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 edition: 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
An electronic edition of the journal is an original one. It is 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 realised by RUCH S.A.
Zamówienia na prenumeratę można składać na stronie / Orders at www.prenumerata.ruch.com.pl

Zam. 515/2019 – nakład 340 egz.

LISTA RECENZENTÓW OCENIAJĄCYCH ARTYKUŁY W 2019 ROKU
LIST OF REVIEWERS WHO REFEREED MANUSCRIPTS IN 2019

Składamy serdeczne podziękowania Recenzentom, którzy służyli nam wiedzą i doświadczeniem i poświęcili czas na wykonanie recenzji artykułów zgłoszonych do naszego czasopisma. W 2019 r. w tym gronie znaleźli się:

We would like to thank the following Reviewers, who in 2019 shared their expertise and knowledge with us and devoted their valuable time to assess the articles submitted for publication in our journal:

Marta Anacka
Andrzej Bąk
Maciej Beręsewicz
Katarzyna Bień-Barkowska
Tadeusz Borys
Marek Bryx
Dariusz Chojecki
Beata Czarnacka-Chrobot
Elżbieta Czarny
Barbara Dańska-Borsiak
Wiesław Dębski
Czesław Domański
Piotr Drygaś
Andrzej Dudek
Grzegorz Dudek
Hanna Dudek
Tadeusz Dudycz
Włodzimierz Dzun
Józefa Famielec
Stanisław Flejterski
Małgorzata Fronczek
Eugeniusz Gatnar
Jan Gawełko
Ewa Giermanowska
Michał Goliński
Wiesław Golnau
Piotr Guzowski
Mariusz Hamulczuk
Magdalena Jaciow
Alina Jędrzejczak
Witold Jurek
Paweł Kaczmarczyk
Ewa Kaznowska
Magdalena Knapińska
Ryszard Kokoszczyński
Adam Kubów
Leszek Kucharski
Janusz Kudła
Emanuel Kulczycki
Sławomir Kurek
Ireneusz Kuropka
Joanna Landmesser
Janusz Łyko
Grzegorz Maciejewski
Sebastian Majewski
Anna Malina
Grażyna Marciniak
Małgorzata Markowska
Ewa Mazurek-Krasodomska
Grzegorz Mentel
Wawrzyniec Michalczyk
Ewa Miklaszewska
Andrzej Miszczuk
Jacek Mizerka
Ruslan Motoryn
Krzysztof Najman
Joanicjusz Nazarko
Wojciech Niemiro
Andrzej Ochocki
Józef Oleński
Walenty Ostasiewicz
Izabela Ostoj
Anna Pajor
Andrzej Paliński
Tomasz Panek
Jan Paradysz
Robert Perdał
Eliza Przeździecka
Grażyna Rosa
Anna Sączewska-Piotrowska
Sven Schreiber
Teresa Słaby
Beata Stachowiak
Iveta Stankovičová
Monika Stanny
Małgorzata Stec
Bogdan Stefanowicz
Mirosław Szreder
Urszula Sztandar-Sztanderska
Izabela Sztangret
Piotr Szukalski
Tomasz Przemysław Śleszyński
Dominik Śliwicki
Grzegorz Ślusarz
Michał Twardochleb
Joanna Tyrowicz
Piotr Tyszko
Marek Walesiak
Agnieszka Wałęga
Stanisław Wanat
Zenon Wiśniewski
Dorota Witkowska
Bogdan Wojtyniak
Feliks Wysocki
Dorota Wyszkowska
Artur Zaborski
Anna Zalcewicz
Małgorzata Zaleska
Wojciech Zieliński
Wojciech Ziętara
Maciej Żukowski

SPIS TREŚCI
CONTENTS

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 1, 6–7
Od redakcji / From the editorial team

OD REDAKCJI

W styczniowym numerze „Wiadomości Statystycznych. The Polish Statistician", który otwiera 65. rok wydawania czasopisma, wprowadzamy nową szatę typograficzną. Żywimy nadzieję, że odświeżenie layoutu, mające na celu uzyskanie większej przejrzystości przekazu, przypadnie Czytelnikom do gustu.

Zamieszczamy artykuły w trzech działach: Studia metodologiczne, Statystyka w praktyce oraz Dyskusje. Recenzje. Informacje.

Przedmiotem pracy dr. hab. inż. Jacka Białka, prof. UŁ, pt. *Wykorzystanie danych skanowanych do pomiaru inflacji – doświadczenia międzynarodowe i wyzwania metodologiczne* są szczegółowe informacje o dobrach konsumpcyjnych uzyskane dzięki skanowaniu ich kodów kreskowych w punktach sprzedaży. Autor wskazuje na zalety, jak również wady i ograniczenia badań wykorzystujących tego rodzaju dane, zwracając uwagę na wyzwania metodologiczne związane z uzyskiwaniem, przetwarzaniem i agregacją danych skanowanych.

Dr inż. Tomasz Śmiałowski w artykule *Demograficzne i terytorialne uwarunkowania zróżnicowania wykluczenia cyfrowego* porusza bardzo istotną kwestię wykluczenia cyfrowego w społeczeństwie informacyjnym. Celem badania opisanego w pracy jest ustalenie wpływu czynników demograficznych i terytorialnych na poziom oraz zróżnicowanie wykluczenia cyfrowego. Autor, wykorzystując skonstruowany przez siebie wskaźnik, wykazuje, że poziom i zróżnicowanie wykluczenia cyfrowego w latach 2003–2015 zwiększały się wraz z wiekiem mieszkańców i spadkiem ich liczby.

Opracowanie prof. dr. hab. Jana Kordosa *Innowacje i badania innowacyjności* dotyczy złożonych zagadnień innowacji i innowacyjności. Autor wskazuje na brak standardowej metody badań innowacyjności w sektorze publicznym. W artykule podkreślono, że organizacje statystyczne coraz częściej dążą do rozwijania współpracy z dostawcami danych, aby badać rozwój innowacji, niemniej jednak skuteczność takich przedsięwzięć nadal pozostaje ograniczona.

Dr hab. Aleksandra Baszczyńska szczegółowo omawia wystąpienia prezentowane na XXXVIII Międzynarodowej Konferencji Naukowej Wielowymiarowa Analiza Statystyczna WAS 2019, która odbyła się w dniach 4–6 listopada 2019 r. w Łodzi. Konferencję zorganizowały Uniwersytet Łódzki, Polskie Towarzystwo Statystyczne – Oddział w Łodzi oraz Komitet Statystyki i Ekonometrii Polskiej Akademii Nauk.

Justyna Gustyn przedstawia nowości wydawnicze GUS z grudnia 2019 r.

Bieżący numer zamyka nowa rubryka *Zapowiedzi wydarzeń w statystyce*. Zamieszczamy w niej informacje o nadchodzących wybranych konferencjach międzynarodowych i innych ważnych wydarzeniach w statystyce, w których wezmą udział przedstawiciele GUS i urzędów statystycznych.

Zapraszamy do lektury.

FROM THE EDITORIAL TEAM

With the January issue of *Wiadomości Statystyczne. The Polish Statistician*, which opens the 65th year of the journal's existence, we are introducing a new graphic design publication, hoping that the changed layout, intended to improve the readability, will appeal to our Readers.

The issue features articles from three sections: Methodological studies, Statistics in practice and Discussions. Reviews. Information.

The subject of the opening paper by Jacek Białek BEng, PhD, DSc, Assoc. Prof, entitled, *Utilisation of scanner data in the measurement of inflation: international experiences and methodological challenges*, are detailed data concerning consumer goods obtained by means of scanning their bar codes in points of sales. The author shows the advantages as well as disadvantages and limitations of research utilising this kind of data, drawing attention especially to research challenges connected to the acquisition, processing and aggregation of scanner data.

Tomasz Śmiałowski BEng, PhD, in his paper entitled *Demographic and territorial determinants of the variability in the degree of digital divide* raises a topical issue of digital divide in the information society. The aim of his research is to determine the influence of demographic and territorial factors on the level of digital divide and its variability. Utilising his original, self-created indicator, the author demonstrates that in 2013–2015, the level of digital divide and its variability were growing when the age of members of the households examined in the study was, as well as when their number was falling.

The complex issues of innovation and innovativeness are discussed in Jan Kordos's (PhD, DSc, Prof. Tit) paper *Innovations and research on innovativeness*. The author indicates that there is no standard method for measuring innovativeness in the public sector. He emphasizes that even though statistical institutions have increasingly often sought collaboration with data providers in order to monitor the progress of innovation, the effectiveness of such endeavours remains limited.

Aleksandra Baszczyńska, PhD, DSc, proposes a detailed description of the 38th International Scientific Conference on Multivariate Statistical Analysis MSA 2019, which was held on 4–6 November 2019 in Łódź. The event was organised by the University of Łódź, Łódź-based branch of the Polish Statistical Association and the Statistics and Econometrics Committee of the Polish Academy of Sciences.

Justyna Gustyn customarily offers the review of Statistics Poland's publications – presently from December 2019.

The issue concludes with *Upcoming events in official statistics* new column, where we publish information on selected upcoming international conferences as well as other important events in official statistics which will be attended by the representatives of Statistics Poland and statistical offices.

We wish you a pleasant reading.

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 1, 9–33 DOI: 10.5604/01.3001.0013.9021
Studia metodologiczne / Methodological studies

Wykorzystanie danych skanowanych do pomiaru inflacji – doświadczenia międzynarodowe i wyzwania metodologiczne[1]

Jacek Białek[a]

Streszczenie. Zgodnie z definicją Organizacji Współpracy Gospodarczej i Rozwoju (OECD) dane skanowane to szczegółowe informacje o dobrach konsumpcyjnych uzyskane dzięki skanowaniu ich kodów kreskowych w punktach sprzedaży. Zaletami tego rodzaju danych są: kompletność już na najniższym poziomie agregacji, relatywnie niski koszt ich uzyskania oraz mnogość obserwacji. Niemniej jednak dane skanowane mają też wady i ograniczenia. Celem artykułu jest wskazanie problemów i wyzwań metodologicznych związanych z uzyskiwaniem, przetwarzaniem i agregacją danych skanowanych wykorzystywanych do szacowania wskaźnika towarów i usług konsumpcyjnych (CPI). Jedna z kluczowych decyzji polega na wyborze formuły indeksowej przeznaczonej dla elementarnych, homogenicznych grup produktów. Istotę problemu wraz z rekomendacjami zademonstrowano na przykładzie dwóch zbiorów danych z portalu Allegro za okres 4.12.2015–28.12.2018, uzyskanych za pomocą narzędzia TradeWatch. Badanie wrażliwości wyników pomiaru dynamiki cen ze względu na wybór formuły indeksu objęło dwie grupy produktów: zegarek męski sportowy oraz fotel biurowy.

Najważniejsze spostrzeżenia są następujące: po pierwsze, różnice między indeksami bilateralnymi a ich łańcuchowymi wersjami mogą być znaczne, co wynika zapewne z dynamicznego charakteru danych skanowanych; po drugie, różnice między wskazaniami indeksów multilateralnych mogą wynosić nawet parę punktów procentowych dla rocznego okna obserwacji; po trzecie, różnice pomiędzy wartościami indeksów GEKS i CCDI są nieznaczne, a różnice między indeksem Geary'ego-Khamisa dla pełnego okna czasowego i okna bieżącego (*real time index*) przestają być znaczące już po upływie kilku miesięcy; po czwarte, ceny produktów sprzedawanych na platformie elektronicznej, a także wartość i wielkość ich sprzedaży mogą zależeć od dnia tygodnia, a nawet godziny pomiaru.

Słowa kluczowe: wskaźnik cen towarów i usług konsumpcyjnych (CPI), dane skanowane, indeksy cen

JEL: C43

Utilisation of scanner data in the measurement of inflation – international experiences and methodological challenges

Abstract. According to the Organisation for Economic Co-operation and Development's definition, scanner data are detailed data on consumer goods obtained by scanning their bar codes at points of sale. The advantages of this kind of data include completeness at the lowest level of aggregation, relatively low cost of acquisition and the fact that they enable a multiplicity of observations. Nevertheless, scanner data also have drawbacks and limitations. The aim of the paper is to identify the problems and methodological challenges related to the acquisition, processing and aggregation of scanner data, which are then used to estimate the Consumer Price Index (CPI). One of the most important decisions in the whole process is the right choice

[1] Artykuł przedstawia wyniki badań sfinansowanych przez Narodowe Centrum Nauki w ramach grantu nr 2017/25/B/HS4/00387.

[a] Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny. ORCID: https://orcid.org/0000-0002-0952-5327.

of an index formula dedicated to elementary (homogeneous) product groups. The essence of the problem, along with some recommendations, has been presented on the example of two sets of data from Allegro e-commerce platform for the period of 4 Dec, 2015 to 28 Dec, 2018, obtained through a special tool – TradeWatch. The evaluation of the sensitiveness of the results of the price dynamics measurement according to the chosen index formula has been carried out on the basis of two groups of products: men's sports watches and office chairs.

The most important observations are as follows: firstly, the differences between bilateral price indices and their chained versions are likely to be significant because of the dynamic character of scanner data sets; secondly, the differences between multilateral price indices might amount to several percentage points for a yearly time window; thirdly, the differences between the values of the GEKS and CCDI indices are slight, and the Geary-Khamis index for the full-time window ceases to differ significantly from the real time index just after a few months; and fourthly, the prices of products sold via electronic platforms as well as their quantities and sales volumes might differ depending on which particular day of a week they were sold, and even at which hour.

Keywords: Consumer Price Index (CPI), scanner data, price indices

1. Wprowadzenie

Zgodnie z definicją Organizacji Współpracy Gospodarczej i Rozwoju (OECD) przez dane skanowane (ang. *scanner data*) rozumie się szczegółowe dane o dobrach konsumpcyjnych uzyskane dzięki skanowaniu ich kodów kreskowych w punktach sprzedaży (ILO, 2004). Technologię użytkowania kodów kreskowych produktów opracowano w latach 70. XX w., a ich wykorzystanie do analizy dynamiki cen i poprawy szacunków wskaźnika cen towarów i usług konsumpcyjnych (Consumer Price Index, CPI) w ciągu ostatnich 20 lat nabierało rozpędu. W tym czasie nie tylko ewoluowały zakres i techniki uzyskiwania danych skanowanych, lecz także poszerzyły się możliwości zdobywania takich informacji i powstała nowa metodologia z zakresu indeksów cen (nowe formuły i nowe sposoby aktualizacji wag).

Pionierem w stosowaniu danych skanowanych są Stany Zjednoczone, zaś niedoścignionym poziom zaawansowania metodologicznego reprezentują Australia i Japonia. Kraje europejskie w niewielkim stopniu korzystają z informacji tego rodzaju, choć ostatnio sytuacja zaczyna się zmieniać. Przykładowo do 2015 r. w Unii Europejskiej (UE) tylko Holandia, Norwegia i Szwecja wykorzystywały dane skanowane do obliczeń indeksów cen detalicznych, a zaledwie rok później dołączyły do nich Belgia, Dania i Islandia. Z literatury przedmiotu opublikowanej do roku 2018 (Guerreiro, Walzer i Lamboray, 2018; Leonard, Sillard, Varlet i Zoyem, 2017; Saraiva dos Santos, Lidonio i Cardoso, 2012) wynika, że obecnie również Luksemburg, Portugalia i Francja eksperymentują z danymi skanowanymi dla wybranych podgrup koszyka CPI. W Polsce konsorcjum Głównego Urzędu Statystycznego (GUS), Instytutu Podstaw Informatyki Polskiej Akademii Nauk (IPI PAN) i Szkoły Głównej Handlowej w Warszawie (SGH) rozpoczyna właśnie projekt *InstatCeny*, ukierunkowany na

wykorzystanie alternatywnych źródeł danych przy kalkulacji CPI, w tym danych skanowanych.

Korzystanie z danych skanowanych jest relatywnie tanie i automatyczne, a do tego operuje się na ich ogromnych wolumenach. Główną zaletą tych danych (czy szerzej: elektronicznych danych transakcyjnych) jest ich kompletność już na najniższym poziomie agregacji. Oznacza to, że dostarczają one informacji zarówno o cenach produktów, jak i o wartości ich sprzedaży na poziomie elementarnym (najbardziej zdezagregowanym). Niemniej jednak dane skanowane mają też wady i ograniczenia, a ich wykorzystanie do pomiaru CPI rodzi wiele trudności metodologicznych. Celem artykułu jest wskazanie problemów i wyzwań związanych z uzyskiwaniem, przetwarzaniem oraz agregacją danych skanowanych. Niezmiernie ważny jest wybór odpowiedniej formuły indeksu cenowego dla elementarnych, homogenicznych grup produktów. Empiryczną ilustrację konsekwencji tego wyboru wraz z rekomendacjami zademonstrowano na przykładzie dwóch dużych zbiorów danych pochodzących z portalu Allegro.

2. Zawartość danych skanowanych

Elektroniczne terminale w punktach sprzedaży obsługują najczęściej następujące kody kreskowe: GTIN (Global Trade Item Number) lub jego europejską wersję EAN (European Article Number), PLU (Price Look-Up) i SKU (Stock Keeping Unit). Najbardziej rozpowszechniony jest kod GTIN (EAN), choć funkcjonują też bardzo specyficzne jego wersje, np. UPC (Universal Product Code) czy lokalny APN (Australian Product Number). Przykładowo GTIN zawiera 8, 12, 13 lub 14 cyfr. Najpopularniejsza jest jego pełna wersja – 13- i 14-cyfrowa. Kod GTIN składa się z: 1 cyfry wskazującej poziom pakowania, 3-cyfrowego kodu organizacji krajowej GS1 (potocznie: kodu kraju, np. 590 – Polska), 4–7 cyfr numeru jednostki kodującej GS1, 2–5 cyfr kodu produktu i 1 cyfry kontrolnej. Kod PLU jest krótszy, a SKU – ogólniejszy niż GTIN i dostarcza mniej detalicznych informacji. Paradoksalnie jednak to właśnie zbyt szczegółowy poziom informacji o produkcie GTIN sprawia, że posługując się tym kodem, trudno jest wyłonić homogeniczne grupy produktów (np. ten sam produkt, ale w innym opakowaniu może mieć dwa różne numery GTIN). Niektóre kraje korzystające z danych skanowanych przy szacowaniu CPI używają więc kodu SKU, a nawet podkreśla się, że stosowanie kodów GTIN lub EAN może nie dawać dobrych rezultatów przy pomiarze inflacji (Dalen, 2017).

Poza kodem kreskowym produktu dane skanowane zawierają wiele innych cennych informacji. Ich zasób jest różny w poszczególnych krajach, także w zależności od dostawcy (sieci supermarketów) lub rodzaju produktu. Obejmuje on: kod sprze-

dawcy (określa grupę towarową według indywidualnej klasyfikacji danej sieci), kod identyfikujący punkt sprzedaży w obrębie danej sieci, etykietę produktu (dodatkowy opis produktu i jego charakterystykę), jednostkę sprzedaży (optymalnie według ujednoliconego formatu – np. „szt", „kg", „paczka", „gr" i „litr"), wartość sprzedaży, liczbę sprzedanych jednostek produktu, flagę (np. oznacza się produkty z przecen i promocji) oraz informację o podatku VAT.

3. Dostawcy danych skanowanych

Wyróżniamy kilka podstawowych źródeł danych skanowanych. Najcenniejszym wydają się bezpośredni dostawcy, a więc punkty sprzedaży, zwłaszcza sieci supermarketów. Supermarkety to potężni potencjalni dostawcy danych – typowy supermarket posiada bazę od 10 tys. do 25 tys. kodów kreskowych sprzedawanych produktów, z których większość stanowią żywność i napoje. Podobnymi dostawcami mogą być mniejsze markety, drobni sprzedawcy, apteki, biura turystyczne, a nawet sklepy internetowe, jeśli tylko archiwizują dane o sprzedaży z uwzględnieniem kodowania produktów.

Drugim źródłem są firmy wyspecjalizowane w badaniu rynku. Niektóre kraje korzystają z danych skanowanych dostarczanych przez firmę A. C. Nielsen lub GfK i włączają je do szacunków krajowego CPI (Krsinich, 2014). Zaletą tego sposobu uzyskiwania danych skanowanych jest możliwość ich szczegółowego filtrowania ze względu na bezpośrednich dostawców, rejony czy uzgodnione z firmą dostawcą charakterystyki homogenicznych grup produktów. Daje to większą swobodę w wyborze dóbr do koszyka CPI i może zwiększyć reprezentatywność próby produktów. Główną wadą tego rozwiązania jest jednak wysoki koszt uzyskiwania danych.

Trzecim, najmniej rozpoznanym w literaturze źródłem danych skanowanych są potężne elektroniczne platformy handlowe (np. działające na polskim rynku serwisy typu Allegro, OLX czy Shumee). Platformy te zrzeszają ogromną liczbę sprzedawców detalicznych i hurtowników, a niektóre z nich (np. Allegro czy Shumee) archiwizują dane o sprzedaży na poziomie kodów EAN. Największą zaletą tego rodzaju źródła danych skanowanych jest ogromny wolumen obrotów dla bardzo wielu produktów. Nie mniej ważny jest bardzo szeroki asortyment sprzedawanych produktów. Za uwzględnieniem tego źródła danych przemawia również możliwość bardzo szczegółowego filtrowania danych o sprzedaży. Na przykład Allegro prowadzi serwis TradeWatch, dzięki któremu można filtrować dane o sprzedaży nie tylko ze względu na kod EAN, lecz także charakterystyki opisowe, kategorie produktów, sprzedawców oraz moment dokonania zakupu (można analizować rozkłady cen danego produktu nawet w ciągu dnia).

4. Techniki uzyskiwania i przetwarzania danych skanowanych

Dane skanowane dostarczane są przez dostawców w formacie csv lub rzadziej, ze względu na rozmiar, w formacie xls. Niektóre kraje dokonują też transferu danych przez serwisy sieciowe (format xml) lub stosują swój wewnętrzny format zapisu. Niekiedy niezbędne jest przeformatowanie otrzymanych danych zgodnie z wymaganiami środowiska IT, w którym dokonywane są dalsze analizy lub realizowany jest konkretny skrypt danego środowiska (może to być środowisko R, Python, SAS czy inne akceptujące wymienione formaty zapisu danych). W zależności od kraju częstotliwość uzyskiwania danych skanowanych może być dzienna, tygodniowa lub miesięczna. Dane najczęściej już na wstępie są filtrowane (usunięcie duplikatów i brakujących lub podejrzanych rekordów) oraz agregowane. Stosownie do potrzeb dane skanowane agreguje się do dnia, tygodnia lub (najczęściej) do miesiąca. W tym ostatnim przypadku z reguły uwzględnia się transakcje z trzech środkowych tygodni każdego miesiąca.

Tak przygotowane dane klasyfikuje się do grup COICOP (Classification of Individual Consumption by Purpose); minimalnym wymogiem jest COICOP 5. W tym celu wykorzystuje się kody produktów (EAN, kod dostawcy), a w przypadku ich braku lub niejednoznaczności korzysta się dodatkowo z etykiet produktów za pomocą metod uczenia maszynowego – machine learningu, w tym metody analizy tekstu – text miningu.

Po etapie klasyfikacji następuje dopasowywanie produktów z porównywanych momentów – matching. Chodzi tutaj o obserwowanie ceny tego samego homogenicznego produktu, nawet jeśli zmieni on kolor opakowania i wagę, a co za tym idzie – zostanie mu nadany inny kod, np. EAN. Dysponując kodami kreskowymi (GTIN, EAN, SKU itd.), kodami wewnętrznymi sprzedawców oraz dostarczanymi przez nich charakterystykami ilościowymi i jakościowymi, ustala się zatem zbiór dopasowanych homogenicznych produktów dla porównywanych okresów jednostkowych, np. kolejnych miesięcy.

Zbiory poprawnie sklasyfikowanych i dopasowanych produktów analizuje się ze względu na zachowania ekstremalne. Ten etap filtrowania (czyszczenia) danych najczęściej uwzględnia udział danego produktu w łącznej sprzedaży w przynależnej mu grupie produktów w porównywanych miesiącach czy też zmianę ceny produktu z miesiąca na miesiąc. Produkty o zbyt małym udziale w rynku (np. w porównywanych miesiącach) najczęściej są usuwane z próby, a w przypadku ekstremalnie małej lub dużej zmiany ceny produktu są flagowane i poddawane dalszej analizie (niektóre kraje usuwają je z próby).

Dalsza analiza ma również wiele wariantów w zależności od kraju. Niekiedy ekstremalną cenę traktuje się jako brakującą informację i dokonuje się jej imputacji. Często

także sprawdza się, czy np. nietypowo niskiej cenie towarzyszy nietypowy spadek wolumenu sprzedaży. Jeśli tak jest, to taką cenę uznaje się za zaniżoną (ang. *dump price*) i usuwa z analizy, przyjmując, że reprezentuje produkt wycofywany z rynku.

Tak przygotowane dane stanowią podstawę do wyznaczenia indeksu cenowego. Poszczególne kraje stosują różne rozwiązania – dynamikę cen na podstawie danych skanowanych wyznacza się za pomocą metod bilateralnych (w tym indeksów nieważonych), jak również multilateralnych (Chessa, 2017).

5. Doświadczenia międzynarodowe w zakresie wykorzystania danych skanowanych

Skala i zakres uzyskiwanych danych skanowanych, techniki ich filtrowania i imputowania oraz metodologia pomiaru dynamiki cen są odmienne w poszczególnych krajach europejskich, a także poza Europą. Poniżej przedstawiono krótką charakterystykę wybranych krajów ze względu na doświadczenie w wykorzystywaniu danych skanowanych do szacowania inflacji (CPI). Została ona opracowana na podstawie literatury przedmiotu podanej w bibliografii załącznikowej oraz referatów wygłoszonych podczas 16. spotkania grupy ottawskiej (Ottawa Group on Price Indices) w Rio de Janeiro w 2019 r.[2]

5.1. Holandia

W Holandii dane skanowane wykorzystano po raz pierwszy w 2001 r. W 2002 r. uzyskiwano dane z dwóch supermarketów, w 2010 r. – z sześciu, a w 2016 r. – z dziesięciu. Od 2017 r. dostawcami danych skanowanych oprócz supermarketów są m.in. sklepy samoobsługowe i biura podróży. Od 2009 r. holenderski urząd statystyczny (Centraal Bureau voor de Statistiek, CBS) otrzymuje dane z częstotliwością tygodniową. Obecnie ponad 20% CPI w Holandii bazuje na danych skanowanych. Od stycznia 2010 r. holenderskie CBS implementuje łańcuchowy indeks Jevonsa (dla miesięcznych podokresów) w modelu dopasowanym (ang. *matched model*).

Holenderscy statystycy uznają kod GTIN za zbyt szczegółowy i preferują SKU. Homogeniczne grupy produktów, poniżej najniższego dostępnego poziomu agregacji ECOICOP, uzyskuje się dzięki holenderskiemu systemowi klasyfikacji produktów według kodów GTIN i ich charakterystykom (Externe Scannerdata Berichtgever Aggregaat, ESBA). W przypadku odzieży uwzględnia się dodatkowe atrybuty produktu: typ ubrania, markę, sezonowość i kolor. Indeksem multilateralnym, który może zastąpić indeks Jevonsa przy kalkulacji holenderskiego CPI, jest indeks Gea-

[2] Elektroniczna wersja wystąpień jest dostępna pod adresem https://eventos.fgv.br/en/ottawa-group-meeting/publications.

ry'ego-Khamisa (Chessa, Verburg i Willenborg, 2017) oraz jego bardzo specyficzna implementacja – indeks czasu rzeczywistego (ang. *real time index*), w której sukcesywnie miesiąc po miesiącu poszerza się okno aktualizacji danych, traktując grudzień poprzedniego roku jako okres bazowy (Chessa, 2016). Holenderscy statystycy eksperymentują ponadto z rolowaną (rolowanie polega na przesuwaniu całego okna czasowego w prawo o miesiąc wraz z każdym nowym miesiącem obserwacji) rocznie wersją indeksu GEKS (Rolling Year GEKS – RYGEKS), który również jest indeksem multilateralnym, ale jego implementacja jest odraczana.

5.2. Belgia

W 2015 r. belgijski urząd statystyczny (Direction générale Statistique) zaczął wykorzystywać dane skanowane (otrzymywane z supermarketów). Zaimplementowano wówczas jedną z metod dynamicznych bazującą na kodach SKU i eliminującą produkty powracające na rynek ze zmienionym kodem (ang. *relaunches*). Naczelną formułą przeznaczoną dla danych skanowanych jest łańcuchowy indeks Jevonsa, przy czym od początku testuje się również indeksy multilateralne: GEKS, oparty na indeksie Törnqvista, TPD (Time Product Dummy), Geary'ego-Khamisa oraz quasi-multilateralny indeks Lehra (Loon i Roels, 2018). Zakłada się, że do końca roku 2020 zostanie wybrana docelowa formuła indeksu cen.

Obliczanie formuły indeksu poprzedza się podstawowym filtrowaniem danych. Belgowie zaimplementowali następujące filtry: (a) sprawdzający, czy dostępność produktu w homogenicznej grupie jest na zadowalającym poziomie; (b) sprawdzający, czy wartość sprzedaży danego produktu jest powyżej ustalonego poziomu procentowego; (c) eliminujący z homogenicznej grupy produktów te, które charakteryzują się dużym spadkiem ceny i ilości; (d) eliminujący z grupy produkty o dużym spadku wolumenu sprzedaży przy niezmienionej cenie; (e) eliminujący produkty charakteryzujące się ekstremalnymi zmianami cen z miesiąca na miesiąc. Ceny brakujące lub eliminowane imputuje się na podstawie ewolucji cenowej pozostałych produktów danej grupy homogenicznej.

Obecnie prowadzone eksperymenty bazują na danych skanowanych dotyczących odzieży, obuwia, usług hotelarskich i sprzętu elektronicznego. Na koniec 2018 r. dane otrzymywano z trzech największych supermarketów, których udział w rynku wynosił 75%–80%.

Dynamikę cen obliczoną na podstawie danych z supermarketów wyznacza się na dzień bieżący dla następujących grup produktów: żywność i napoje bezalkoholowe (waga 16,9%), napoje alkoholowe i wyroby tytoniowe (2,3%), drobne narzędzia i akcesoria (0,3%), nietrwałe dobra domowe (0,8%), produkty dla zwierząt (0,7%), produkty papierowe (0,1%), produkty do rysowania (0,2%) oraz produkty higieny osobistej (1,4%).

5.3. Szwecja

Początki szwedzkich eksperymentów z danymi skanowanymi sięgają lat 90. XX w. W jednej z pierwszych prac na ten temat (Dalen, 1997) można przeczytać, że najwcześniejszym dostawcą tego rodzaju danych dla Szwecji była firma A. C. Nielsen, która stworzyła bazę danych otrzymywanych ze szwedzkich supermarketów obejmującą jedynie cztery homogeniczne grupy produktów: mrożone ryby, płatki śniadaniowe, tłuszcze i detergenty. Początkowo brano pod uwagę jedynie formuły cenowych indeksów Laspeyresa i Fishera wraz z ich wersjami łańcuchowymi (okresem podstawowym był miesiąc). Obecnie dane skanowane w Szwecji pochodzą z: aptek i sklepów farmaceutycznych (od 2010 r.), sklepów monopolowych z alkoholami (od 2016 r.) oraz małych marketów, supermarketów i hipermarketów (których udziały pokrywają 80% rynku – od 2011 r.). Co ciekawe, w przypadku supermarketów losowana jest próba 60 punktów sprzedaży (według schematu losowania proporcjonalnego do udziałów w rynku), a w przypadku aptek oraz sklepów z alkoholami przeprowadzane jest pełne badanie. Generalnie sześć źródeł danych skanowanych daje podstawę do szacowania 19% szwedzkiego CPI. Statystycy skrupulatnie wyliczają, że w Szwecji automatyzacja pobierania i przetwarzania danych skanowanych bez manualnej pomocy ankietera prowadzi do oszczędności rzędu 80 tys. euro rocznie, mimo przynajmniej dwukrotnie większej próby produktów.

Obecnie danych skanowanych używa się w przypadku następujących grup produktów: owoce, warzywa, mięso, wyroby tytoniowe, napoje alkoholowe, lekarstwa, lampy i baterie, środki czystości, żywność dla zwierząt oraz środki higieny osobistej. Od 2001 r. szwedzki urząd statystyczny (Statistika centralbyrån) posługuje się rynkowym indeksem sprzedaży żywności (Food Sales in the Trade), wykorzystującym kody GTIN oraz automatyczne, wewnętrzne kodowanie produktów do grup homogenicznych. Szwedzi szacują, że każdego miesiąca analizuje się w ten sposób 100 tys. cen produktów, przy czym informacje o cenach i ilościach agreguje się do tygodnia. W przypadku supermarketów dane pochodzą z trzech środkowych tygodni miesiąca, a w przypadku aptek ich zbieranie odbywa się między 25. dniem poprzedniego a 24. dniem bieżącego miesiąca. Na najbardziej elementarnym poziomie agregacji danych skanowanych Szwedzi stosują indeks Jevonsa.

5.4. Luksemburg

W Luksemburgu współpraca urzędu statystycznego STATEC[3] z dostawcami danych skanowanych trwa od niedawna. W styczniu 2018 r. po opracowaniu metodologii korzystania z tego rodzaju danych rozpoczęła się ich regularna transmisja do syste-

[3] Narodowy Instytut Statystyki i Studiów Ekonomicznych Wielkiego Księstwa Luksemburga (L'Institut national de la statistique et des études économiques du Grand-Duché de Luxembourg).

mu STATEC. W fazie wdrożeniowej założono analizowanie produktów żywnościowych i napojów bezalkoholowych z wyłączeniem produktów sezonowych (świeże owoce i warzywa) i stopniowe poszerzanie asortymentu. Obecnie dane skanowane dostarczane do luksemburskiego urzędu statystycznego zawierają: kod EAN (13-cyfrowy), kod dostawcy, etykietę produktu, kod produktu pochodzący z jego wewnętrznej kategoryzacji, etykietę produktu nadaną przez dostawcę oraz wartość, wielkość i okres sprzedaży z dokładnością do miesiąca. Dane są dostarczane 18. dnia każdego miesiąca, a zamiast cen stosuje się wartości jednostkowe (jest to wartość sprzedaży z danego miesiąca podzielona przez wielkość sprzedaży).

Pierwszym etapem analizy danych skanowanych jest ich klasyfikacja do podgrup znajdujących się poniżej poziomu COICOP 5 (np. żywność i napoje bezalkoholowe z wyłączeniem owoców i warzyw ma w Luksemburgu aż 72 podgrupy). Klasyfikacja odbywa się automatycznie, ale pierwszy zbiór uczący bazował jeszcze na klasyfikacji manualnej. Obecnie system luksemburski uznaje za niedopasowane te produkty, które różnią się kodem EAN lub etykietą, czyli jest wrażliwy na powtórne wprowadzanie na rynek tego samego produktu, ale np. o zmienionej wadze czy w opakowaniu o innym kolorze. Dla wszystkich dopasowanych produktów wyznacza się miesięczną zmianę ceny (zwykły iloraz ceny z bieżącego miesiąca i ceny z poprzedniego miesiąca).

5.5. Australia

Australijskie Biuro Statystyczne (Australian Bureau of Statistics, ABS) zaczęło regularnie uzyskiwać dane skanowane w 2014 r. Dotyczyło to zaledwie kilku produktów (m.in. paliw) i kilku dostawców. W ostatnich latach ABS podjęło współpracę z największymi australijskimi sieciami handlowymi i obecnie produkty, w przypadku których wykorzystuje się dane skanowane, stanowią ok. 25% koszyka CPI; są to m.in.: owoce i warzywa, mięso i owoce morza, chleb i produkty zbożowe oraz wyroby tytoniowe.

W 2015 r. Australia rozpoczęła realizację projektu badawczego *Enhancing the Australian CPI: a roadmap*, którego jednym z priorytetowych celów było rozwinięcie metodologii uwzględnienia danych skanowanych w CPI. W pierwszych próbach wyznaczania dynamiki cen dla grup elementarnych na podstawie danych skanowanych posługiwano się bilateralnymi, superlatywnymi formułami Törnqvista i Fishera. Ostatecznie ABS rozpoczęło na szeroką skalę eksperymenty z metodami multilateralnymi. Pod uwagę wzięto m.in. indeks GEKS oparty na formule Törnqvista oraz indeks TPD (Time Product Dummy).

5.6. Stany Zjednoczone

W Stanach Zjednoczonych zaczęto korzystać z danych skanowanych ponad 25 lat temu, jeszcze przed erą Internetu. Pierwsze wzmianki na temat współpracy amery-

kańskich urzędów statystycznych z dostawcami danych skanowanych pochodzą z 1993 r., natomiast informacje o kalkulacji indeksów cen bazujących na danych skanowanych – z 1995 r., gdy firma A. C. Nielsen rozpoczęła dostarczanie danych do amerykańskiego Bureau of Labor Statistics (BLS). Początkowo dane te dotyczyły spożycia kawy w Waszyngtonie i Chicago; w 1996 r. Marshall Reinsdorf skonstruował nawet specjalne indeksy cenowe dla kawy oparte na danych skanowanych. Podobne analizy przeprowadzano później z wykorzystaniem danych dotyczących spożycia płatków śniadaniowych sprzedawanych w Nowym Jorku (1998). Dostawcą danych również w tym przypadku była firma A. C. Nielsen. Dane skanowane z amerykańskiego rynku są dostarczane także przez Information Resources, Inc.

Obecnie w przypadku wyrobów medycznych przy każdorazowej kalkulacji indeksu cenowego (aktualizacja miesięczna) bierze się pod uwagę ok. 600 mln rekordów (2,5 TB danych). To znacznie więcej niż przeciętna w Europie, ale trzeba uwzględnić wielkość i liczbę ludności Stanów Zjednoczonych. Obliczenia są wykonywane na specjalnym, przeznaczonym tylko do tego zadania serwerze (Unix Sun server), a mimo to analizy na potrzeby publikacji miesięcznego CPI trwają parę dni.

Metodologia dotycząca danych skanowanych bazuje na kodach UPC, a głównymi formułami indeksów są cenowe indeksy superlatywne, przede wszystkim łańcuchowy indeks Törnqvista.

5.7. Japonia

Japonia jest przykładem kraju, który w pełni wykorzystuje swój potencjał technologiczny i podejmuje współpracę z potężnymi instytucjami, takimi jak np. Bank of Japan czy Nikkei Digital Media, aby jak najdokładniej szacować inflację, opierając się na możliwie najszerszej gamie danych skanowanych. Od 1998 r. gromadzi dane skanowane pochodzące z 300 sieci supermarketów z całego kraju, przy czym jest to celowa próba dostawców (w przeciwieństwie np. do Stanów Zjednoczonych, gdzie pobiera się próbę losową).

Obecnie realizowany jest projekt rządowy *UTokyo Daily Price Index*, którego wymierny efekt, jak nazwa wskazuje, stanowi publikacja rocznej stopy inflacji każdego dnia (w Japonii poziom CPI jest podawany do publicznej wiadomości raz na miesiąc). Skala przedsięwzięcia jest ogromna: zbiór danych skanowanych od 1998 r. obejmuje 6 bln transakcji dotyczących ok. 2 mln produktów sklasyfikowanych w 213 homogenicznych grupach produktów. W 2013 r. dane skanowane zebrano z 16008 punktów sprzedaży na łączną kwotę transakcji 0,42 tln jenów. Kodem kreskowym służącym do identyfikacji produktów jest japoński odpowiednik EAN – 13-cyfrowy JAN (Japanese Article Number). Japońskie dane skanowane dotyczą żywności, napojów i artykułów gospodarstwa domowego, a nie obejmują np. artykułów elektronicznych oraz usług, jak w niektórych krajach europejskich. Udział produktów uwzględnionych w *UTokyo Daily Price Index* wynosi 17% CPI.

Wspomniany indeks jest publikowany codziennie i ma tylko trzydniowe opóźnienie. Wskazuje na roczną inflację, co oznacza, że okresem bazowym jest odpowiadający badanemu dniu dzień z roku poprzedniego. Do jego wyznaczenia stosuje się cenowy indeks Törnqvista.

Japonia publikuje również *UTokyo Monthly Price Index*, który obrazuje roczną stopę inflacji i jest aktualizowany raz na miesiąc według takiej samej metodologii co *UTokyo Daily Price Index*.

5.8. Polska

Główny Urząd Statystyczny w Polsce ponad trzy lata temu rozpoczął współpracę z kilkoma sieciami supermarketów. Chociaż dane skanowane nie są jeszcze dostarczane regularnie, to jednak ich fragmentaryczny zbiór pozwolił na budowanie pewnych doświadczeń w zakresie importu i przetwarzania danych. W ośrodkach urzędów statystycznych w Poznaniu i Opolu trwają prace nad doskonaleniem metod poprawnej klasyfikacji produktów do odpowiednich grup COICOP (m.in. na podstawie machine learningu). Z początkiem roku 2019 w Departamencie Handlu i Usług GUS podjęto prace eksperymentalne nad wyborem optymalnej formuły multilateralnego indeksu cenowego odpowiedniego dla danych skanowanych.

W tym samym czasie zainicjowano także projekt *InstatCeny* (GUS, IPI PAN, SGH), ukierunkowany na wykorzystanie nowych źródeł danych w pomiarze CPI.

6. Problemy i wyzwania metodologiczne

Wykorzystanie danych skanowanych do pomiaru CPI rodzi pewne problemy i przynosi wyzwania metodologiczne. Głównymi są:

- **Wybór dostawcy danych i współpraca z nim.** Doświadczenia Polski i innych krajów wskazują, że nie jest łatwo pozyskać nowego dostawcę danych skanowanych. Sieci marketów są niekiedy niechętnie nastawione do współpracy z urzędami statystycznymi, ponieważ nie widzą w takim działaniu korzyści finansowych i obawiają się osłabienia swojej konkurencyjności (obawa przed niepełną poufnością danych). Doświadczenia różnych krajów pokazują, że od nakłonienia do współpracy supermarketu do sfinalizowania umowy legalizującej jej realizację upływa najczęściej około pół roku.
- **Dobór próby do analizy.** Polityka krajów wykorzystujących dane skanowane w zakresie poboru próby jest różna. Część krajów stosuje próby losowe, a część – celowe, przy czym dużo zależy od tego, z iloma sieciami handlowymi podpisano umowy o współpracy, oraz od zróżnicowania geograficznego i demograficznego danego kraju. Jeżeli sieć marketów różnicuje swoją ofertę i ceny w zależności od regionu, to przy doborze próby niezbędne jest uwzględnienie rejoni-

zacji. Podobnie jeśli w danym rejonie poszczególne punkty notowań należące do tej samej sieci mają np. różne godziny otwarcia czy inne promocje, to dobór próby również powinien to uwzględniać i traktować takie punkty notowań osobno (agregacja do całej sieci jest tu niewskazana).

Problematyczny może być także wybór momentów czasowych dla poboru próby. Rozkład cen w ciągu tygodnia nie jest równomierny, np. blisko weekendu ceny często są zawyżane. Również agregacja danych do tygodnia może budzić wątpliwości, które tygodnie (i ile) z danego miesiąca są najbardziej reprezentacyjne.

- **Budowa środowiska IT.** Środowisko IT odpowiadające potrzebom analizy danych skanowanych musi korzystać z zaawansowanych metod i technik uczenia maszynowego (klasyfikacja, rozpoznawanie tekstu). Dynamiczny charakter zbiorów danych skanowanych w połączeniu z ich ogromnym wolumenem i różnorodnością sprawia, że niezbędne procesy: filtrowanie danych, dopasowywanie produktów, klasyfikacja do grup homogenicznych i obliczenia indeksów cen są dość czasochłonne i wymagają nie tylko szybkich algorytmów, lecz często także uruchomienia dodatkowych serwerów w celu przeprowadzenia obliczeń.
- **Opracowanie metodologii przygotowania i analizy danych skanowanych.** Mimo bogatych doświadczeń urzędów statystycznych w zakresie stosowania danych skanowanych wiele problemów metodologicznych nadal czeka na rozwiązanie. Wśród krajów korzystających z danych skanowanych są zarówno zwolennicy, jak i przeciwnicy filtrowania danych; ci drudzy uważają, że filtrowanie nadmiernie redukuje próbę, co powoduje utratę dynamicznego charakteru zbioru danych). Kraje, które zdecydowały się na filtrowanie i imputację danych, stoją przed wyborem parametrów progowych filtrów (np. jaką zmianę ceny uznać już za nietypową albo jaki udział w rynku uznać za wystarczająco duży, aby włączyć produkt do próby) lub samej techniki imputacji (w tym zakresie dostępne są rekomendacje Eurostatu). Metodologia powinna np. określać, za pomocą jakich technik uczenia maszynowego dokonuje się dopasowywania produktów w porównywanych miesiącach oraz ich klasyfikacji do grup/podgrup COICOP, ponieważ właściwe mapowanie produktów nadal jest dużym wyzwaniem dla większości krajów.

Kolejne wyzwanie stanowi uwzględnienie dóbr sezonowych. Niektóre kraje wykluczają takie produkty z próby (zwłaszcza w przypadku tzw. mocnej sezonowości, polegającej na tym, że w pewnych okresach roku produkty są całkowicie niedostępne). Choć metodologia konstruowania indeksów cen bazuje na pojęciu homogenicznej grupy produktów, to problem właściwego zdefiniowania tego pojęcia oraz stworzenia technik do ich selekcji (np. metoda MARS – zob. Chessa, 2018) jest nadal szeroko dyskutowany w Eurostacie.

Ogromnym wyzwaniem metodologicznym jest wybór formuły indeksu cenowego, który spełniałby określone wymogi formalne i praktyczne (np. powinien mieć dobre

własności aksjomatyczne i redukować zjawisko tzw. dryfu łańcuchowego (ang. *chain drift* – zob. Chessa, 2015).

Problemem, z jakim musi się zmierzyć każdy urząd statystyczny korzystający z danych skanowanych, jest wreszcie wybór systemu wag, dzięki któremu możliwe będzie przechodzenie do wyższych poziomów agregacji i włączanie wyliczonych indeksów cen do publikowanej informacji o poziomie inflacji.

7. Formuły indeksów cen

W przypadku danych skanowanych zastosowanie mają indeksy bilateralne (bezpośrednie) i multilateralne. Pierwsza grupa indeksów uwzględnia jedynie okres badany (t) oraz bazowy (0), a druga grupa porównuje te dwa momenty i bierze pod uwagę informacje o cenach i ilościach z ustalonego okna czasowego $[0, T] \supset [0, t]$. Najczęściej przyjmuje się 13-miesięczne okno czasowe (Diewert i Fox, 2017), choć niektóre kraje stosują znacznie dłuższe (np. Australia rekomenduje 27-miesięczne). Ponieważ zbiory danych skanowanych są bardzo dynamiczne, chociażby ze względu na sezonowość produktów oraz dobra nowe i znikające z rynku, oznaczmy przez $G_{0,t}$ te produkty z rozważanej homogenicznej grupy produktów, które są dostępne jednocześnie w okresie 0 i t. Niech $N_{0,t} = \text{card}\,(G_{0,t})$; przez p_i^τ, q_i^τ i s_i^τ oznaczmy odpowiednio cenę, ilość i relatywny udział w sprzedaży i-tego produktu w okresie τ. Zakładając, że w momencie τ dostępnych jest N_τ produktów, otrzymujemy:

$$s_i^\tau = \frac{p_i^\tau q_i^\tau}{\sum_{k=1}^{N_\tau} p_k^\tau q_k^\tau} \tag{1}$$

Artykuł prezentuje najpopularniejsze formuły indeksów cen używanych w przypadku danych skanowanych, które zostały wykorzystane w badaniu empirycznym. Ich listę otwiera najczęściej stosowany, bilateralny i przy tym nieważony indeks Jevonsa (1865):

$$P_J^{0,t} = \prod_{i \in G_{0,t}} \left(\frac{p_i^t}{p_i^0} \right)^{\frac{1}{N_{0,t}}} \tag{2}$$

przy czym najczęściej korzysta się z jego łańcuchowej wersji (hained Jevons index) w postaci:

$$P_{CH-J}^{0,t} = \prod_{\tau=0}^{t-1} P_J^{\tau,\tau+1} \tag{3}$$

gdzie $P_J^{\tau,\tau+1} = \prod_{i\in G_{0,t}} \left(\frac{p_i^{\tau}}{p_i^{\tau-1}}\right)^{\frac{1}{N_{\tau,\tau-1}}}$, $N_{\tau,\tau-1} = \text{card}\ (G_{\tau,\tau-1})$, gdzie $G_{\tau,\tau-1}$ – produkty z rozważanej homogenicznej grupy produktów, które są dostępne jednocześnie w okresie τ i τ–1.

Ważonymi indeksami bilateralnymi są tzw. indeksy superlatywne (Diewert, 1976) – najczęściej indeks Törnqvista (1936):

$$P_T^{0,t} = \prod_{i\in G_{0,t}} \left(\frac{p_i^t}{p_i^0}\right)^{\frac{s_i^0+s_i^t}{2}} \tag{4}$$

lub indeks Fishera (1922), stanowiący średnią geometryczną z indeksów Laspeyresa (1871) i Paaschego (1874), oznaczonych odpowiednio przez $P_{La}^{0,t}$ i $P_{Pa}^{0,t}$:

$$P_F^{0,t} = \sqrt{P_{La}^{0,t} \cdot P_{Pa}^{0,t}} \tag{5}$$

gdzie:

$$P_{La}^{0,t} = \frac{\sum_{i\in G_{0,t}} p_i^t q_i^0}{\sum_{i\in G_{0,t}} p_i^0 q_i^0} \tag{6}$$

$$P_{Pa}^{0,t} = \frac{\sum_{i\in G_{0,t}} p_i^t q_i^t}{\sum_{i\in G_{0,t}} p_i^0 q_i^t} \tag{7}$$

W ostatnich latach szczególną uwagę w literaturze przedmiotu poświęca się indeksom multilateralnym, m.in. dlatego, że są one tranzytywne, czyli niewrażliwe na wybór referencyjnego okresu, oraz niwelują zjawisko dryfu łańcuchowego. Problem z dryfem łańcuchowym polega na tym, że w przypadku gdy ceny i ilości powracają do wartości wyjściowych, wartość indeksu cenowego nie wraca do 1 (chociaż powinna). Jest to charakterystyczne dla dóbr sezonowych i występuje nawet w przypadku łańcuchowych wersji indeksów superlatywnych.

W niniejszej pracy wykorzystano następujące indeksy multilateralne:
a) indeks GEKS (zob. Gini, 1931; Eltetö i Köves, 1964; Szulc, 1964):

$$P_{GEKS}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{P_F^{\tau,t}}{P_F^{\tau,0}}\right)^{\frac{1}{T+1}} \tag{8}$$

gdzie $P_F^{\tau,t} = \sqrt{P_{La}^{\tau,t} \cdot P_{Pa}^{\tau,t}}$, $P_{La}^{\tau,t} = \frac{\sum_{i \in G_{0,t}} p_i^t q_i^\tau}{\sum_{i \in G_{0,t}} p_i^\tau q_i^\tau}$, $P_{Pa}^{\tau,t} = \frac{\sum_{i \in G_{0,t}} p_i^t q_i^t}{\sum_{i \in G_{0,t}} p_i^\tau q_i^t}$

b) indeks CCDI (Caves, Christensen i Diewert, 1982; Inklaar i Diewert, 2016):

$$P_{CCDI}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{P_T^{\tau,t}}{P_T^{\tau,0}} \right)^{\frac{1}{T+1}} \tag{9}$$

gdzie $P_T^{\tau,t} = \prod_{i \in G_{\tau,t}} \left(\frac{p_i^t}{p_i^\tau} \right)^{\frac{s_i^\tau + s_i^t}{2}}$

c) uogólniony indeks Lehra, traktowany niekiedy jako quasi-multilateralny (Loon i Roels, 2018):

$$P_{Lehr}^{0,t} = \frac{\sum_{i \in G_t} p_i^t q_i^t \,/\, \sum_{i \in G_0} p_i^0 q_i^0}{\sum_{i \in G_t} v_i^L q_i^t \,/\, \sum_{i \in G_0} v_i^L q_i^0} \tag{10}$$

gdzie:

$$v_i^L = \frac{\sum_{\tau=0}^{T} p_i^\tau q_i^\tau}{\sum_{\tau=0}^{T} p_i^\tau} \tag{11}$$

d) indeks Geary'ego-Khamisa (Geary, 1958; Khamis, 1972):

$$P_{GK}^{0,t} = \frac{\sum_{i \in G_t} p_i^t q_i^t \,/\, \sum_{i \in G_0} p_i^0 q_i^0}{\sum_{i \in G_t} v_i^{GK} q_i^t \,/\, \sum_{i \in G_0} v_i^{GK} q_i^0} \tag{12}$$

gdzie:

$$v_i^{GK} = \sum_{z=0}^{T} \varphi_{i,GK}^z \frac{p_i^z}{P_{GK}^{0,z}} \tag{13}$$

$$\varphi_{i,GK}^z = \frac{q_i^z}{\sum_{\tau=0}^{T} q_i^\tau} \tag{14}$$

Formuły (12), (13) i (14) wyznaczają układ wzajemnie powiązanych równań nieliniowych, którego rozwiązanie uzyskuje się jedynie symultanicznie (pewne iteracyjne metody można znaleźć w pracach: Chessa, 2016; Diewert i Fox, 2017; Maddison

i Rao, 1996). Z technicznego punktu widzenia jest to indeks najtrudniejszy do wyznaczenia. Cen i ilości ze wszystkich okresów T+1 można użyć dopiero w ostatnim okresie (przyjmijmy, że w miesiącu T) rozpatrywanego okna czasowego. Ta niedogodność została wyeliminowana w indeksie czasu rzeczywistego[4] (*real time index* $P_{RT}^{0,t}$). Jego konstrukcja bazuje na koncepcji indeksu GEKS, przy czym za pomocą specjalnego algorytmu (omawia go szczegółowo np. praca Chessy, 2016) aktualizuje się system wag występujących w formule (12) dla bieżącego, analizowanego momentu t, tzn.:

$$v_i^{RT} = \sum_{z=0}^{t} \varphi_{i,RT}^{z} \frac{p_i^z}{P_{GK}^{0,z}} \tag{15}$$

$$\varphi_{i,RT}^{z} = \frac{q_i^z}{\sum_{\tau=0}^{t} q_i^{\tau}} \tag{16}$$

Dla ostatniego okresu z okna czasowego $[0, T]$ zachodzi $P_{GK}^{0,T} = P_{RT}^{0,T}$. W ilustracji empirycznej wykorzystano także indeks JGEKS, którego formuła odpowiada wzorowi (8), lecz zamiast bazowego indeksu Fishera stosuje się indeks Jevonsa (zob. (2)).

8. Przykład empiryczny

Departament Handlu i Usług GUS prowadzi prace eksperymentalne nad formułami indeksów cen i sposobami agregacji danych skanowanych, korzystając m.in. z danych otrzymywanych z sieci supermarketów. Pierwsze wyniki omawiają m.in. Białek i Bobel (2019) oraz Białek i Roszko-Wójtowicz (2019). W niniejszym artykule wykorzystano jednak dane pochodzące z portalu Allegro, który może stanowić alternatywne źródło elektronicznych danych transakcyjnych przy pomiarze inflacji (podobnie jak OLX, Shumee i in.). Dane tego rodzaju mają odmienną naturę niż dane skanowane pochodzące z supermarketów. Po pierwsze, charakteryzuje je bardzo duża liczba podmiotów wystawiających dany produkt, często pojedynczo (wiele osób wystawia bardzo dużo produktów; czasem są to pojedyncze sztuki danego modelu, a czasem jest to sprzedaż hurtowa). Wolumen sprzedaży produktów jest ogromny (tablica).

[4] Autor artykułu nie spotkał się z polskim tłumaczeniem nazwy tego indeksu. Nazwa *real time index*, wprowadzona przez jego pomysłodawcę (Chessa, 2015, 2016), jest powszechnie stosowana przez ekspertów od indeksów multilateralnych (np. należących do tzw. grupy ottawskiej), choć nie została oficjalnie wprowadzona do literatury przedmiotu. Koncepcji, zgodnie z którą szacowany jest ten indeks, nadano nazwę FBEW (Fixed Base Expanding Window).

Po drugie, fluktuacje cenowe na elektronicznych platformach handlowych są znacznie dynamiczniejsze niż w supermarketach. Ceny produktów na Allegro często zmieniają się w ciągu doby, a histogram rozkładu ceny (także wielkości sprzedaży) jest najczęściej powiązany z dniem tygodnia (wykr. 1 i 2). W przypadku tego rodzaju danych ważne jest zatem, aby do kalkulacji indeksów używać nie cen z konkretnego momentu czasowego, lecz wartości jednostkowych (ang. *unit values*), stanowiących uśrednienie ceny zagregowane do ustalonego okresu (np. tygodnia czy miesiąca).

W niniejszej analizie wykorzystano dane dotyczące fotela biurowego oraz męskiego zegarka sportowego (pobrane za pomocą wyspecjalizowanego narzędzia, jakim jest TradeWatch), a więc dane z bardzo niskiego szczebla agregacji, znacznie poniżej COICOP 5. Te dwie homogeniczne grupy produktów mogłyby uzupełniać listę reprezentantów w grupach odpowiednio COICOP 05111 (meble do mieszkania i domu) oraz COICOP 12312 (zegary i zegarki). Tablica przedstawia ogólną charakterystykę sprzedaży tych produktów na Allegro w okresie 4.12.2015–28.12.2018 oraz przykładowe kody EAN w tych grupach. Wykresy 1 i 2 ukazują fluktuacje cen i wielkości sprzedaży tych produktów w różnych ujęciach (m.in. rozkład według godzin i dni tygodnia).

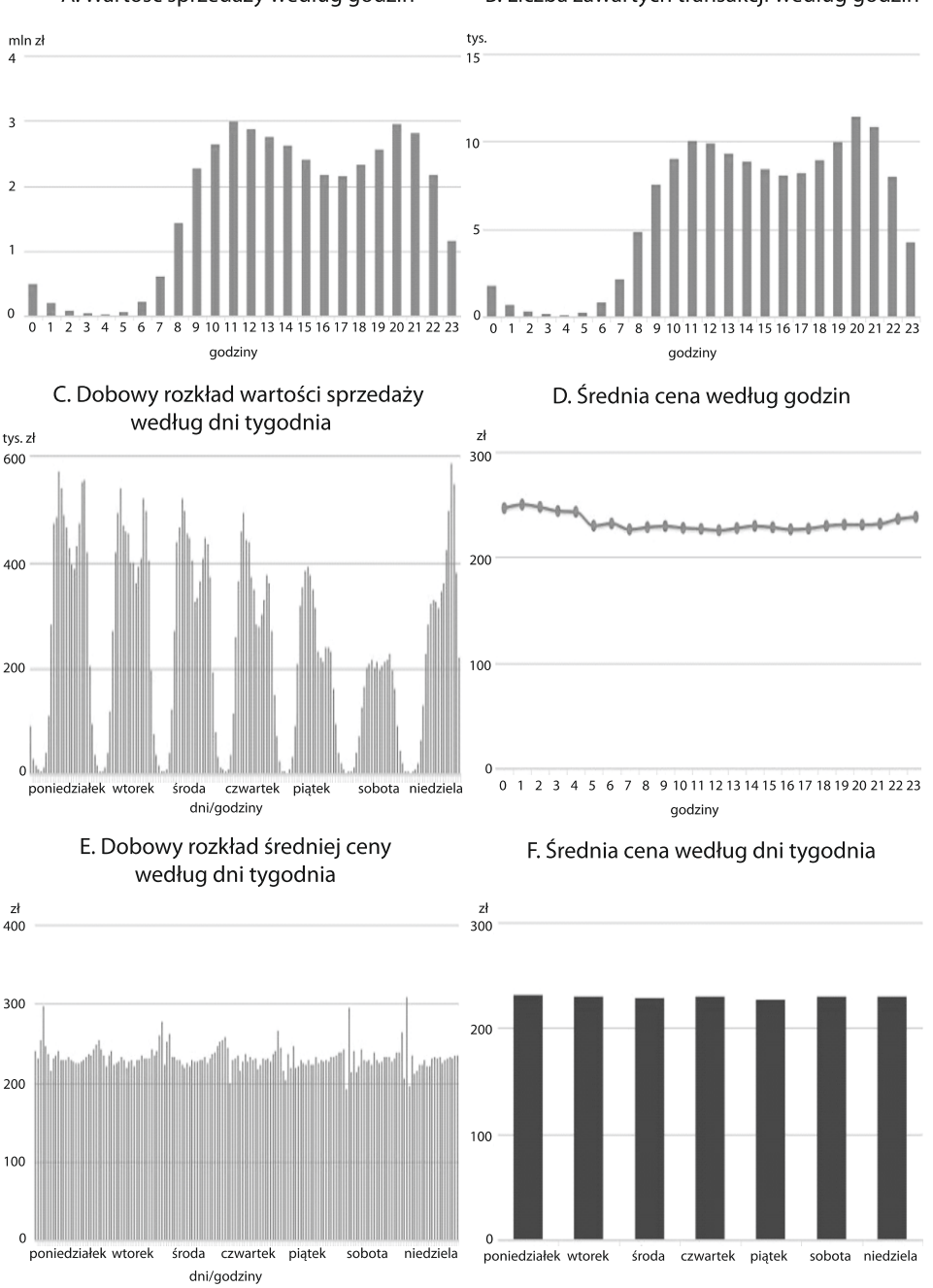
Tablica. Charakterystyka sprzedaży fotela biurowego i męskiego zegarka sportowego na Allegro w okresie 4.12.2015–28.12.2018

Wyszczególnienie	Fotel biurowy	Męski zegarek sportowy
Łączna sprzedaż w zł	40588787,48	6829768,26
Średnia dzienna sprzedaż w zł	36207,66	6092,57
Średnia cena 1 sztuki w zł	229,91	56,19
Średnia wartość 1 transakcji w zł	278,47	63,43
Liczba sprzedanych sztuk	176541	121551
Liczba zawartych transakcji	145758	107666
Przykładowe kody EAN	5902759970007	4971850907152
	5902767100175	4971850948377
	5908239692667	4971850984191

Źródło: TradeWatch.pl.

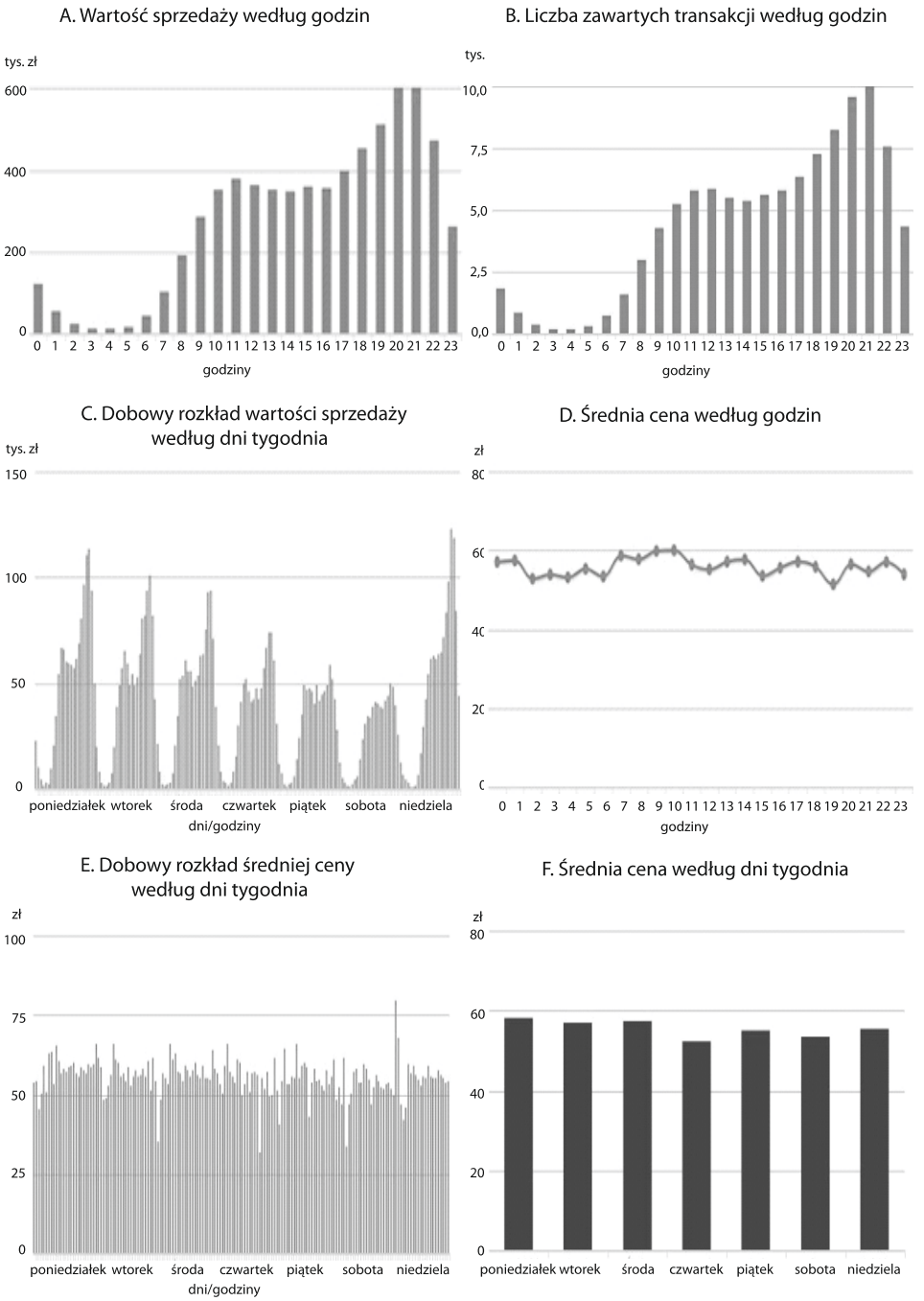
W dalszej analizie ograniczono się do ostatnich 12 miesięcy rozpatrywanego okresu w celu wyznaczenia rocznej dynamiki cen. Dane zagregowano do miesiąca, a następnie przefiltrowano w celu usunięcia ekstremalnych zmian cen (np. ponad 50-procentowy wzrost ceny z miesiąca na miesiąc) oraz produktów o relatywnie niskiej wartości sprzedaży (udział poniżej 1% w rynku). W ten sposób usunięto ok. 30% produktów w obu grupach. Wyniki dla indeksów cen prezentują wykr. 3 i 4.

Wykr. 1. Fluktuacje cen i wielkości sprzedaży fotela biurowego w ujęciu godzinowym i dziennym: transakcje na Allegro w okresie 4.12.2015–28.12.2018



Źródło: TradeWatch.pl.

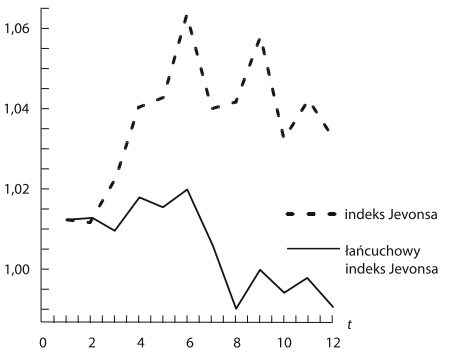
Wykr. 2. Fluktuacje cen i wielkości sprzedaży męskiego zegarka sportowego w ujęciu godzinowym i dziennym: transakcje na Allegro w okresie 4.12.2015–28.12.2018



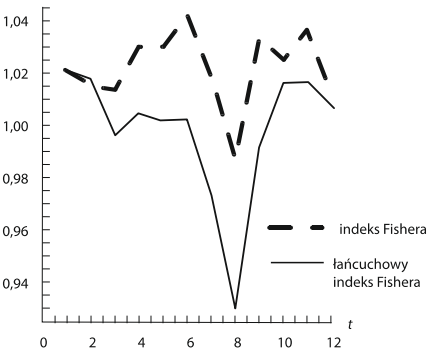
Źródło: jak przy wykr. 1.

Wykr. 3. Porównanie indeksów cen dla fotela biurowego: transakcje na Allegro, okno: 13 miesięcy ($t = 0, 1, ...,12$) w okresie grudzień 2017 – grudzień 2018

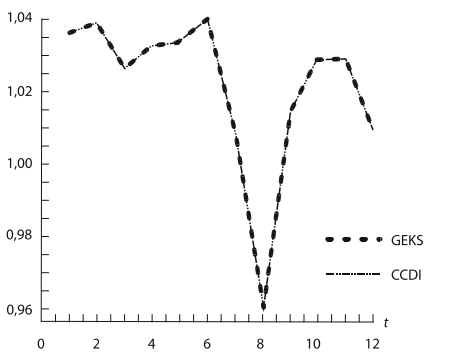
A. Indeks Jevonsa i łańcuchowy indeks Jevonsa



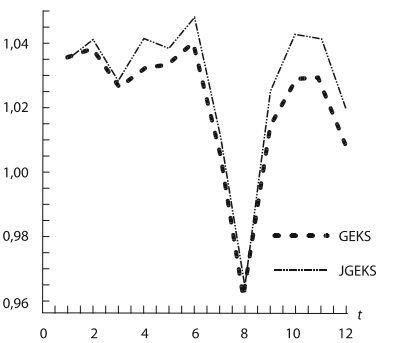
B. Indeks Fishera i łańcuchowy indeks Fishera



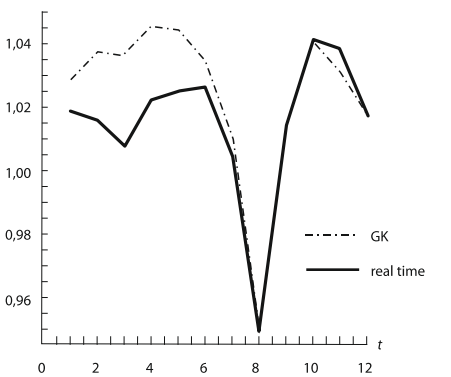
C. Indeksy GEKS i CCDI



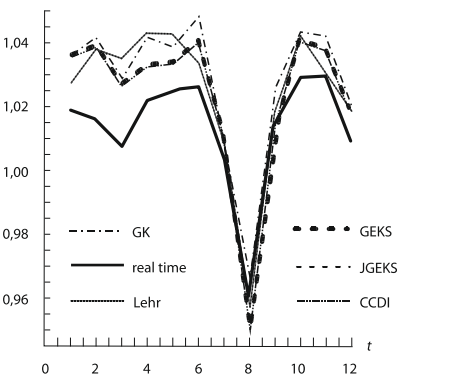
D. Indeks GEKS i jego wersja oparta na indeksie Jevonsa (JGEKS)



E. Indeksy Geary'ego-Khamisa i real time

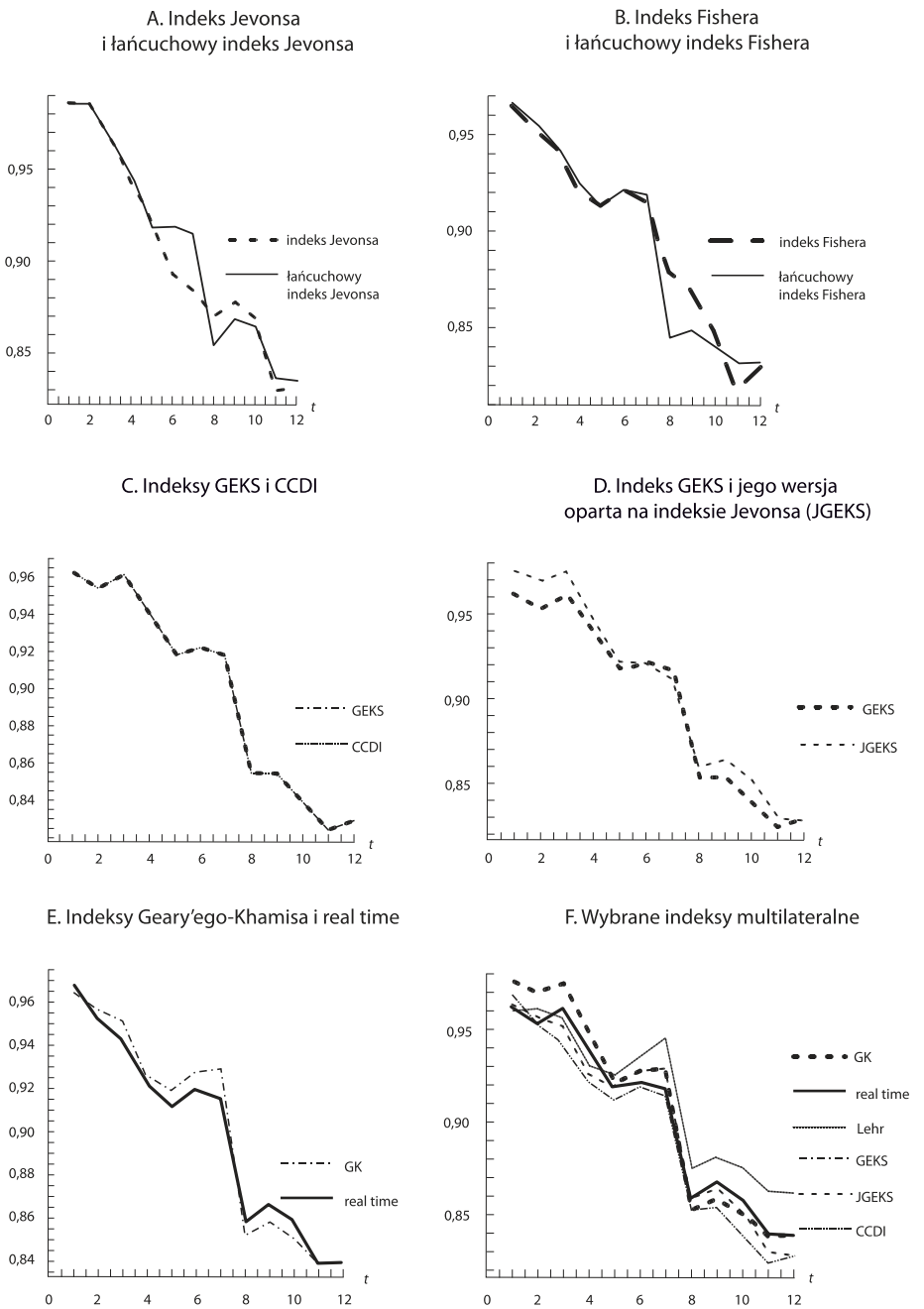


F. Wybrane indeksy multilateralne



Źródło: jak przy wykr. 1.

Wykr. 4. Porównanie indeksów cen dla zegarka męskiego sportowego: transakcje na Allegro, okno: 13 miesięcy (t = 0,1,...12) w okresie grudzień 2017 – grudzień 2018



Źródło: jak przy wykr. 1.

Na podstawie analizy produktów sprzedawanych na platformie elektronicznej można zaobserwować, że ich ceny i wielkość sprzedaży mogą się zmieniać nie tylko ze względu na dzień tygodnia (zob. wykr. 1E, 1F, 2E i 2F), lecz także w ciągu doby (wykr. 1A–D, 2A–D). Dlatego – na co zwrócono uwagę w artykule i co jest podkreślane w literaturze przedmiotu – przy kalkulacji indeksów cen nie powinno się używać cen dla konkretnego momentu czasowego, lecz cen uśrednionych (ang. *unit values*) za dany okres, najczęściej miesiąc.

Dane z Allegro charakteryzują się ogromnym wolumenem zawieranych transakcji (tablica), jest to zatem cenne źródło informacji o zachowaniach konsumentów. Co jednak za tym idzie, różnice wskazań nieważonych indeksów cen w stosunku do wersji ważonych mogą być znaczne, gdyż obserwuje się duże różnice w udziałach w sprzedaży nawet wśród najlepiej sprzedawanych produktów w danej grupie (wykr. 3).

Zarówno dane skanowane pochodzące z supermarketów, jak i dane transakcyjne z elektronicznych platform handlowych cechuje bardzo duża dynamika z uwagi na sezonowość produktów, trendy sprzedaży oraz politykę dostawców. Na Allegro występuje znaczna rotacja produktów i sprzedawców, nawet po odfiltrowaniu sprzedaży incydentalnych, z czego wynikają zauważalne różnice wskazań indeksów bilateralnych (ważonych lub nie) oraz ich wersji łańcuchowych (wykr. 3A i 4A oraz wykr. 3B i 4B). W przeprowadzonym badaniu różnice te sięgały nawet kilku punktów procentowych w ciągu roku.

Jako alternatywę dla indeksów bilateralnych w badaniu uwzględniono indeksy multilateralne, szczególnie rekomendowane w literaturze do analizy danych skanowanych. W przypadku indeksu GEKS zastąpienie bazowej formuły, jaką jest indeks Fishera, indeksem Törnqvista (w ten sposób uzyskuje się indeks CCDI) zdaje się nie mieć znaczenia (wykr. 3C i 4C), ale już powrót do bazowego indeksu Jevonsa powoduje zauważalne różnice wartości (wykr. 3D i 4D). Z kolei różnice między indeksem Geary'ego-Khamisa szacowanym dla 13-miesięcznego okna czasowego i indeksem czasu rzeczywistego stają się nieznaczne już po upływie 6–7 miesięcy (wykr. 3E i 4E). Jest to dobra wiadomość dla krajów, które dopiero rozpoczynają uzyskiwanie i gromadzenie danych skanowanych i nie posiadają jeszcze informacji dla pełnego (np. rocznego) okna analizy.

Podsumowując, nawet wybór pomiędzy indeksami multilateralnymi nie jest oczywisty, a różnica ich wskazań dla jednej homogenicznej grupy produktów może wynosić kilka punktów procentowych (wykr. 3F i 4F).

9. Podsumowanie

Dane skanowane powinny być traktowane w szczególny sposób. Po pierwsze, już na najniższych poziomach agregacji dostarczają informacji o konsumpcji produktów. Jest to podstawowa różnica w stosunku do tradycyjnego zbioru danych, a także da-

nych scrapowanych (zbieranych ze stron internetowych). Jakość danych skanowanych jest zatem nieporównywalnie lepsza niż danych scrapowanych, tym bardziej że zawierają informacje o cenach transakcyjnych, a nie ofertowych. Elektroniczne platformy handlowe są potencjalnie bardzo dobrym źródłem danych elektronicznych, które noszą wszelkie znamiona danych skanowanych (wolumen jest duży, ceny są cenami transakcyjnymi, wielkość sprzedaży i kod produktu są znane).

Po drugie, gromadzenie, przetwarzanie, filtrowanie, klasyfikowanie i dopasowywanie (matching) danych skanowanych wymaga zaawansowanych technik – machine learningu i text miningu – oraz budowy odpowiedniego środowiska IT w celu automatyzacji tych procesów.

Po trzecie, mimo wieloletniego doświadczenia niektórych krajów w posługiwaniu się danymi skanowanymi, wciąż istnieje wiele wyzwań metodologicznych. W artykule wskazano na potencjalne problemy i utrudnienia w stosowaniu danych skanowanych, związane m.in. z wyborem ich dostawcy, doborem próby produktów czy procedurą postępowania z danymi. Jak nadmieniono, nawet właściwe zdefiniowanie i zlokalizowanie homogenicznych grup produktów nadal rodzi wiele problemów, takich jak np. wybór właściwej formuły indeksu cen. Wiąże się to z ogromną dynamicznością zbiorów danych skanowanych, wynikającą z sezonowości produktów, trendów sprzedaży i polityki dostawców.

W pracy pominięto analizę głównych problemów, takich jak dryf łańcuchowy, silna sezonowość czy efekt substytucji. Skoncentrowano się na określeniu skali różnic, jakie może spowodować zmiana sposobu obliczania dynamiki cen. Wybór nieważonej formuły Jevonsa, jakkolwiek zasadny przy klasycznie uzyskiwanych cenach z najniższych poziomów agregacji, wydaje się tu mniej racjonalny, gdyż nie wykorzystuje się wówczas informacji o wielkości konsumpcji. Nie zmienia to jednak faktu, że większość krajów stosuje łańcuchowy indeks Jevonsa do analizy danych skanowanych, ponieważ indeks ten stwarza znacznie mniej trudności aplikacyjnych niż indeksy multilateralne. Niektóre kraje, np. Stany Zjednoczone czy Japonia, sięgnęły po ważone indeksy superlatywne (Fishera, Törnqvista), ale i w ich wypadku nie ma zgodności co do tego, czy lepiej przyjąć podejście z ustaloną podstawą (ang. *fixed base approach*), czy też stosować łańcuchowe wersje tych indeksów. Wiadomo, że to drugie rozwiązanie może prowadzić do efektu dryfu łańcuchowego, czyli braku powrotu wartości indeksu do 1 w przypadku dóbr sezonowych, których ceny wracają do poziomu wyjściowego.

Rozsądnym wyborem w przypadku danych skanowanych wydają się zatem indeksy multilateralne, które „nadążają" za dynamizmem homogenicznych grup produktów i cechują się tranzytywnością oraz brakiem dryfu łańcuchowego. Do sformułowania ostatecznej rekomendacji indeksu cen odpowiedniego dla danych skanowanych potrzebne są jednak kolejne doświadczenia krajów w tym zakresie oraz pogłębione badania własności indeksów zarówno na gruncie teoretycznym, jak i empirycznym.

Bibliografia

Białek, J., Bobel, A. (2019). Comparison of Price Index Methods for CPI Measurement using Scanner Data. W: *Paper presented at the 16th Meeting of the Ottawa Group on Price Indices* (s. 1–40). Rio de Janeiro, Brazil, 8–10 May 2019. Rio de Janeiro: FGV.

Białek, J., Roszko-Wójtowicz, E. (2019). The Impact of the Price Index Formula on the Consumer Price Index Measurement. *Statistika – Statistics and Economy Journal*, *99*(3), 246–258.

Caves, D. W., Christensen, L. R., Diewert, W. E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. *Economic Journal*, *92*(365), 73–86. DOI: 10.2307/2232257.

Chessa, A. G. (2015). Towards a generic price index method for scanner data in the Dutch CPI. *Room document for Ottawa Group Meeting*. Urayasu City, Japan.

Chessa, A. G. (2016). A new methodology for processing scanner data in the Dutch CPI. *Eurona, 1*, 49–69.

Chessa, A. G. (2017). Comparisons of QU-GK Indices for Different Lengths of the Time Window and Updating Methods. *Paper prepared for the second meeting on multilateral methods organised by Eurostat*. Luxembourg: Statistics Netherlands.

Chessa, A. G. (2018). Product definition and index calculation with MARS-QU: Applications to consumer electronics. *Report Statistics Netherlands.*

Chessa, A. G., Verburg, J., Willenborg, L. (2017). A comparison of price index methods for scanner data. *Paper presented at the 15th Meeting of the Ottawa Group on Price Indices*. Eltville am Rhein, Germany.

Dalen, J. (1997). Experiments with Swedish Scanner Data. *Proceedings of the Third Meeting of the International Working Group on Price Indexes*. Research Paper no. 9806. Statistics Netherlands, Division Research and Development, Department of Statistical Methods.

Dalen, J. (2017). Unit values in scanner data and some operational issues. *Paper presented at the fifteenth Ottawa Group Meeting*. Eltville am Rhein, Germany.

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, *4*(2), 115–145. DOI: 10.1016/0304-4076(76)90009-9.

Diewert, W. E., Fox, K. J. (2017). Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data. *Discussion paper*, *17*(2), 1–62. DOI: 10.2139/ssrn.3276457.

Eltetö, Ö., Köves, P. (1964). Egy nemzetközi összehasonlításoknál fellépő indexszámítási problémáról. On a Problem of Index Number Computation Relating to International Comparisons (in Hungarian). *Statisztikai Szemle, 42*, 507–518.

Fisher, I. (1922). *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability*. Boston, New York: Houghton Mifflin Company.

Geary, R. C. (1958). A Note on the Comparisons of Exchange Rates and Purchasing Power Between Countries. *Journal of the Royal Statistical Society. Series A (General)*, *121*(1), 97–99. DOI: 10.2307/2342991.

Gini, C. (1931). On the Circular Test of Index Numbers. *Metron*, *9*, 3–24.

Guerreiro, V., Walzer, M., Lamboray, C. (2018). The use of Supermarket Scanner data in the Luxembourg Consumer Price Index. *Working papers du STATEC*, *Economie et Statistiques*, *97*, 1–18.

ILO. (2004). *Consumer Price Index Manual. Theory and practice*. Geneva: International Labour Office.

Inklaar, R., Diewert, W. E. (2016). Measuring industry productivity and cross-country convergence. *Journal of Econometrics*, *191*(2), 426–433. DOI: 10.1016/j.jeconom.2015.12.013.

Jevons, W. S. (1865). On the Variation of Prices and the Value of the Currency since 1782. *Journal Statistical Society of London*, *28*(2), 294–320. DOI: 10.2307/2338419.

Khamis, S. H. (1972). A New System of Index Numbers for National and International Purposes. *Journal of the Royal Statistical Society. Series A (General), 135*(1), 96–121. DOI: 10.2307 /2345041.

Krsinich, F. (2014). The FEWS Index: Fixed Effects with a Window Splice – Non-Revisable Quality-Adjusted Price Indices with no Characteristic Information. *Paper presented at the meeting of the group of experts on consumer price indices* (s. 26–28). Geneva, Switzerland.

Laspeyres, E. (1871). Die Berechnung einer mittleren Waaren-preissteigerung. *Jahrbücher für Nationalöko-nomie und Statistik*, *16*, 296–314.

Leonard, I., Sillard, P., Varlet, G., Zoyem, J. P. (2017). Scanner data and quality adjustment. *Serie des Documents de Travail*. Working Paper No. F1704, INSEE, 1–31.

Loon, K. V., Roels, D. (2018). Integrating big data in the Belgian CPI. *Paper presented at the meeting of the group of experts on consumer price indices* (s. 8–9). Geneva, Switzerland.

Maddison, A., Rao, D. S. P. (1996). A Generalized Approach to International Comparison of Agricultural Output and Productivity. Research memorandum GD-27. *Groningen Growth and Development Centre*. Groningen, The Netherlands.

Paasche, H. (1874). Über die Preisentwicklung der letzten Jahre nach den Hamburger Börsennotirungen. *Jahrbücher für Nationalökonomie und Statistik*, *23*(2/4), 168–178. DOI: 10.1515/jbnst-1874-0113.

Saraiva dos Santos, P., Lidonio, F., Cardoso, C. (2012). Scanner Data Project: the experience of Statistics Portugal. *Paper presented at the Workshop on Scanner Data* (s. 1–13). Stockholm.

Szulc, B. (1964). Indices for Multiregional Comparisons. *Przegląd Statystyczny*, *3*, 239–254.

Törnqvist, L. (1936). The Bank of Finland's Consumption Price Index. *Bank of Finland Monthly Bulletin*, *16*(10), 27–34.

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 1, 34–45 DOI: 10.5604/01.3001.0013.9027
Statystyka w praktyce / Statistics in practice

Demograficzne i terytorialne uwarunkowania zróżnicowania wykluczenia cyfrowego

Tomasz Śmiałowski[a]

Streszczenie. Rozwój społeczeństwa informacyjnego nieuchronnie wiąże się ze zjawiskiem wykluczenia cyfrowego. Nierówności w dostępie do technologii informacyjno-komunikacyjnych (ICT) i korzystaniu z nich wpływają na procesy gospodarcze, społeczne i demograficzne. Celem badania jest ustalenie wpływu czynników demograficznych i terytorialnych na poziom i zróżnicowanie wykluczenia cyfrowego polskich gospodarstw domowych. Badanie obejmowało lata 2003–2015. Wykorzystano dwa źródła danych: raporty Diagnozy społecznej oraz zasoby danych Głównego Urzędu Statystycznego dotyczące budżetów gospodarstw domowych. Analizy oparto na autorskim wskaźniku wykluczenia cyfrowego. Uzyskane wyniki wskazują, że poziom i zróżnicowanie wykluczenia cyfrowego zwiększały się wraz ze wzrostem wieku mieszkańców oraz spadkiem ich liczby. Nie zaobserwowano wpływu płci i regionu na badane zjawisko.

Słowa kluczowe: technologie informacyjno-komunikacyjne, wykluczenie cyfrowe, wskaźnik wykluczenia cyfrowego

JEL: O33

Demographic and territorial determinants of the variability of the degree of digital divide

Abstract. One of the inevitable outcomes of the development of information society is the phenomenon of digital divide. Unequal access to and unequal use of information and communication technologies (ICT) by Polish households affect the economic, social and demographic processes. The aim of the article is to determine the influence of demographic and territorial factors on the level of digital divide and its variability among Polish households in the years 2003–2015. The study was conducted on the basis of data from the reports of the Social Diagnosis and Statistics Poland's data on households' budgets, and utilised an original, author-created indicator of digital divide. The results demonstrated that in the analysed period, the level of digital divide and its variability were growing when, on the one hand, the age of the members of the households taken into account in the research was, and on the other, when their number was decreasing. No influence of these persons' sex or the region of residence was observed.

Keywords: information and communication technologies, digital divide, indicator of digital divide

[a] Szkoła Główna Gospodarstwa Wiejskiego, Wydział Ekonomiczny. ORCID: https://orcid.org/0000-0002-7559-9180.

1. Wprowadzenie

Rozwój społeczeństwa informacyjnego, w którym dość szybko upowszechnia się digitalizacja danych – a w konsekwencji informatyzacja administracji publicznej – powoduje, że wykluczenie cyfrowe staje się odczuwalne nie tylko na poziomie jednostkowym, lecz także w skali całego kraju. Różnice w dostępie do technologii informacyjno-komunikacyjnych (Information and Communication Technologies, ICT) oraz ich wykorzystaniu są obecnie jednymi z najważniejszych rodzajów nierówności. W badaniach nad wykluczeniem cyfrowym oprócz czynników ekonomicznych (Śmiałowski i Ochnio, 2019) i społecznych (Śmiałowski, 2019) analizuje się również procesy demograficzne, a przede wszystkim zmiany struktury ludności według płci i wieku (Bylok, 2012; Witkowska i Matuszewska-Janica, 2013). Van Dijk (2012), badając ICT, zauważa, że w kolejnych grupach wieku stopień wykorzystania rozwiązań informacyjno-komunikacyjnych się zmniejszał, jak również że nieznacznie częściej tymi technologiami posługiwali się mężczyźni. Badania przeprowadzone przez Dixona i współpracowników (2014) potwierdzają występowanie powyższych zależności, zaobserwowali oni jednak, że mężczyźni znacznie częściej korzystają z ICT. Dodatkowo autorzy wskazują na duże znaczenie przynależności rasowej – częściej z technologii korzystały osoby rasy białej. Zbliżone stanowisko prezentują także Brendesha i Kimberly (2013), Witte, Kiss i Lynn (2015) oraz Zillien i Marr (2015).

W badaniach nad wykluczeniem cyfrowym jako determinanta zróżnicowania równie często wskazywany jest czynnik terytorialny. W zależności od charakterystyki prowadzonych analiz badacze posługują się różnymi klasyfikacjami jednostek terytorialnych (NUTS). Śmiałowski i Jałowiecki (2012) wykazali, że we wszystkich województwach widoczny jest systematyczny wzrost liczby gospodarstw domowych wyposażonych w różne technologie. Pomimo że dynamika tego procesu była intensywniejsza we wschodniej części kraju, pomiędzy wschodnią i zachodnią Polską nadal występowały różnice. Ten podział potwierdzają również badania rozwoju społeczeństwa informacyjnego w Polsce przeprowadzone przez Saramę (2011). Drugim bardzo często wykorzystywanym czynnikiem terytorialnym jest klasa miejscowości zamieszkania. Turczak i Zwiech (2013) wskazują na klasę miejscowości, obok grupy społeczno-ekonomicznej i poziomu wykształcenia, jako czynnik wpływający na strukturę wydatków konsumpcyjnych gospodarstw domowych, w tym wydatków na ICT.

Celem badania jest ustalenie wpływu czynników demograficznych i terytorialnych na poziom i zróżnicowanie wykluczenia cyfrowego polskich gospodarstw domowych.

2. Metoda badania

W prowadzonym badaniu do grupy czynników demograficznych zaliczono grupę wieku (GWK) i płeć (PŁĆ), natomiast do grupy czynników terytorialnych – region (RGN) i klasę miejscowości zamieszkania (KLM). W badaniu posłużono się klasyfi-

kacją opierającą się na czterech grupach wieku (Rószkiewicz, 2006). W pierwszej (GWK 1) znalazły się osoby, które w swojej grupie społeczno-ekonomicznej należały do pierwszej grupy kwartylowej wyznaczonej na podstawie wieku. W analogiczny sposób została utworzona druga (GWK 2), trzecia (GWK 3) i czwarta (GWK 4) grupa wieku. W przypadku regionów podstawę analiz stanowiła klasyfikacja zgodna z podziałem jednostek terytorialnych do celów statystycznych (NUTS 1). W badaniu wyróżniono cztery klasy miejscowości: miasta powyżej 500 tys. mieszkańców (KLM 1), miasta od 100 tys. do 500 tys. mieszkańców (KLM 2), miasta poniżej 100 tys. mieszkańców (KLM 3) i wsie (KLM 4) (Podolec, 2008).

Badane czynniki wybrano na podstawie analizy literatury przedmiotu oraz dwóch źródeł danych – raportów z Diagnozy społecznej oraz wyników badań Głównego Urzędu Statystycznego (GUS) dotyczących budżetów gospodarstw domowych.

Wyznaczanie poziomu i zróżnicowania wykluczenia cyfrowego w polskich gospodarstwach domowych w latach 2003–2015 składało się z czterech etapów (Śmiałowski, 2019; Śmiałowski i Ochnio, 2019):

1. Na podstawie wskaźnika wykluczenia cyfrowego zaklasyfikowano każdą dorosłą osobę do jednej z czterech grup wyodrębnionych ze względu na nierówności cyfrowe (GWC[1]): osób wykluczonych cyfrowo (GWC I), osób zagrożonych wykluczeniem cyfrowym (GWC II), osób korzystających z najnowszych rozwiązań ICT (GWC III) oraz osób w pełni korzystających z najnowszych rozwiązań ICT (GWC IV) i charakteryzujących się wysokimi kompetencjami cyfrowymi zarówno w obszarze usług internetowych (takich jak komunikacja, zakupy, rozrywka, e-administracja), jak i obsługi komputera czy urządzeń mobilnych.
2. Za pomocą testu niezależności χ^2 zweryfikowano występowanie statystycznie istotnej zależności pomiędzy stopniem wykluczenia cyfrowego a badanym czynnikiem oraz określono jej siłę na podstawie wartości współczynnika kontyngencji C Pearsona (Koronacki i Mielniczuk, 2006).
3. Na podstawie współczynnika Giniego (Anand, 1983) oraz indeksu Theila (Theil, 1967) dokonano oceny stopnia zróżnicowania wykluczenia cyfrowego dla poszczególnych czynników oraz określono jego strukturę.
4. Za pomocą gradacyjnej analizy danych (Grade Correspondence Analysis, GCA; Grade Correspondence-cluster Analysis, GCCA) (Kowalczyk, Pleszczyńska i Ruland, 2004) zidentyfikowano współzależności pomiędzy stopniem wykluczenia cyfrowego a badanym czynnikiem.

Na pierwszym etapie badań dla analizowanych czynników różnicujących w kolejnych latach przygotowano tabele wielodzielcze. Przedstawiały one rozkłady ze względu na grupę nierówności cyfrowych oraz na kategorie danego czynnika różnicujące-

[1] W artykule zachowano oznaczenie „GWC" stosowane we wcześniejszych pracach autora.

go. Na podstawie otrzymanych wyników wyznaczono procentowe udziały grup nierówności cyfrowych w poszczególnych kategoriach danego czynnika różnicującego zgodnie ze wzorem:

$$y_j = \frac{x_i}{\sum_{i=1}^{4} x_i} \tag{1}$$

gdzie:

x_i – liczba osób w *i*-tej grupie nierówności cyfrowych w *j*-tej kategorii,

$\sum_{i=1}^{4} x_i$ – liczba osób ogółem w *j*-tej kategorii.

Wyznaczone udziały procentowe pogrupowano według kategorii danego czynnika różnicującego w taki sposób, że dla każdej kategorii została stworzona tabela zawierająca strukturę grup nierówności cyfrowych w kolejnych latach badania. Następnie dokonano oceny średnich wartości i dynamiki zmian dla badanego okresu. Do wyznaczenia przeciętnej wartości badanej cechy obserwowanej w różnych momentach wykorzystano średnią chronologiczną obliczoną na podstawie wzoru:

$$\bar{y} = \frac{1}{n+1} \sum_{t=1}^{n} y_t \tag{2}$$

Otrzymane wyniki posłużyły do stworzenia profilu wykluczenia cyfrowego. Wykorzystanie powyższego wzoru powoduje, że wartości w tabl. 1 i 2 nie sumują się do 100%. Szczegółowy opis metodologii wyznaczenia wskaźnika wykluczenia cyfrowego oraz jego oceny przedstawiono w pracy Śmiałowskiego (2019).

Ze względu na dwuletnie odstępy w dostarczaniu raportów z Diagnozy społecznej przez Radę Monitoringu Społecznego oceny poziomu i zróżnicowania wykluczenia cyfrowego dokonano w interwałach dwuletnich.

3. Wyniki badania

Spośród czterech analizowanych czynników wpływ na poziom zróżnicowania wykluczenia cyfrowego miały grupa wieku oraz klasa miejscowości. W przypadku dwóch pozostałych czynników (płci i regionu) profil wykluczenia cyfrowego oraz zróżnicowanie były zbliżone do ogólnopolskich. Powyższe wyniki mają odzwierciedlenie w sile zależności pomiędzy przynależnością do grup wyznaczonych na pod-

stawie wybranych czynników różnicujących a przynależnością do grup nierówności cyfrowych. Średnia wartość współczynnika kontyngencji wynosiła odpowiednio: dla klasy miejscowości – 0,200, dla grupy wieku – 0,190, dla regionu — 0,050 i dla płci – 0,047.

Badanie poziomu wykluczenia cyfrowego wykazało, że wraz ze wzrostem wieku zwiększał się udział w profilu wykluczenia grupy osób wykluczonych cyfrowo (GWC I), a zmniejszał się udział grupy osób w pełni korzystających z najnowszych rozwiązań ICT (GWC IV). We wszystkich grupach wieku największym udziałem w profilu wykluczenia cyfrowego charakteryzowały się osoby zakwalifikowane do grupy GWC I (przeciętna wartość poziomu wykluczenia wynosiła odpowiednio: 35,0%, 46,3%, 53,6% i 60,2%). Mniejszy udział cechował osoby z grupy GWC IV, czyli w pełni korzystające z najnowszych rozwiązań ICT (odpowiednio: 29,5%, 21,1%, 16,8% i 15,6%). Dwie pozostałe grupy wykluczenia cyfrowego (GWC II i III) cechował udział na podobnym, niskim poziomie wynoszącym ok. 10%. Spośród wszystkich grup wieku tylko w drugiej zaobserwowano profil wykluczenia cyfrowego zbliżony do ogólnopolskiego (profil ogólnopolski wynosił odpowiednio: 49,3%, 9,9%, 7,8% i 20,5%), (zob. tabl. 1).

W badanym okresie we wszystkich grupach wieku tylko grupa osób wykluczonych cyfrowo (GWC I) zmniejszyła swój udział w profilu wykluczenia. W pozostałych grupach nierówności cyfrowych we wszystkich przedziałach wieku występowała tendencja wzrostowa.

Tabl. 1. Profil wykluczenia cyfrowego w poszczególnych grupach wieku w latach 2003–2015 (przeciętna wartość poziomu wykluczenia)

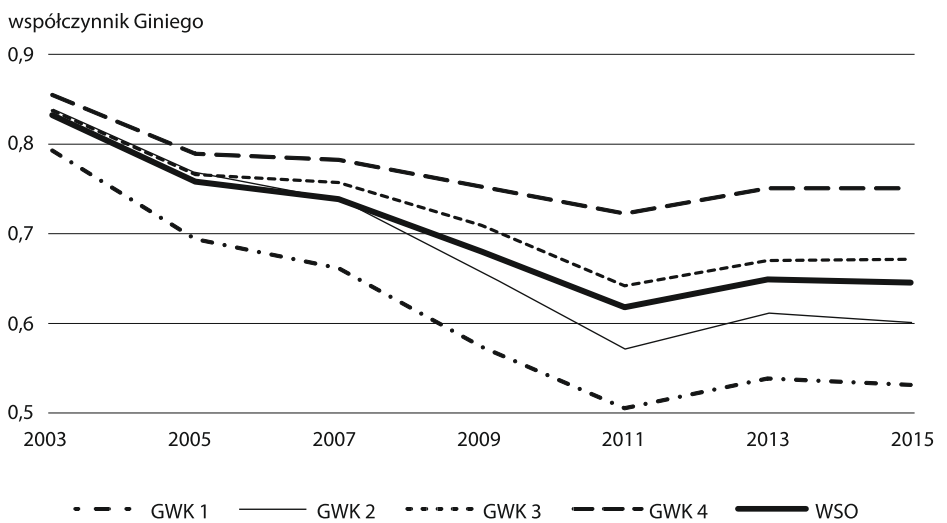
Grupy	GWK 1	GWK 2	GWK 3	GWK 4
	w %			
GWC I	35,0	46,3	53,6	60,2
GWC II	12,3	11,5	9,9	6,4
GWC III	10,7	8,5	7,2	5,3
GWC IV	29,5	21,1	16,8	15,6

Źródło: opracowanie własne na podstawie Czapiński i Panek (2003–2015) oraz danych GUS.

Znaczny udział grupy osób wykluczonych cyfrowo (GWC I) w profilu wykluczenia we wszystkich grupach wieku wpływał na poziom zróżnicowania (wykr. 1 i 2). Zdecydowanie najmniejszym zróżnicowaniem wykluczenia cyfrowego charaktery-

zowała się pierwsza grupa wieku (przeciętna wartość współczynnika Giniego – 0,537), a największym – czwarta (0,675). Zależności te mają odzwierciedlenie w tempie zmian zachodzących w czasie. Najszybszy spadek zróżnicowania wykluczenia cyfrowego został zaobserwowany w pierwszej grupie wiekowej, a najwolniejszy – w czwartej. Dekompozycja zróżnicowania wykluczenia cyfrowego w badanych grupach wieku wykazała, że udział w całkowitym zróżnicowaniu wykluczenia cyfrowego wszystkich grup wieku był na zbliżonym poziomie (od 19,9% do 22,5%). W przypadku zróżnicowania międzygrupowego udział w całkowitym zróżnicowaniu wykluczenia cyfrowego kształtował się na niskim poziomie i wynosił 3,5%.

Wykr. 1. Tendencja dyspersji wykluczenia cyfrowego w poszczególnych grupach wieku



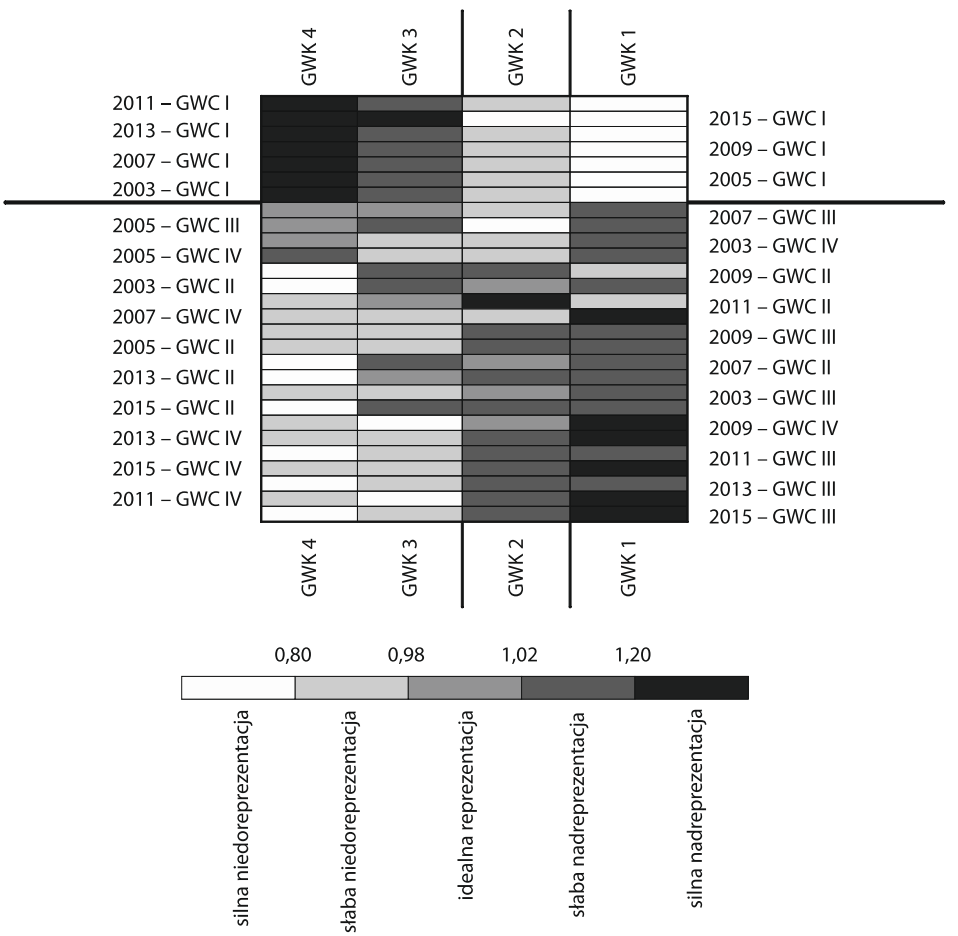
Źródło: opracowanie własne na podstawie Czapiński i Panek (2003–2015).

Analiza nadreprezentacji potwierdziła występowanie dużego zróżnicowania poziomu wykluczenia cyfrowego we wszystkich grupach wieku (wykr. 2). Grupa osób wykluczonych cyfrowo (GWC I) charakteryzowała się większym w stosunku do struktury wieku ogółem udziałem osób starszych (GWK 3 i 4) oraz mniejszym udziałem osób młodszych (GWK 1). W pozostałych grupach nierówności cyfrowych (GWC II, III i IV) wystąpiła odwrotna zależność.

Badanie poziomu wykluczenia cyfrowego w grupach wieku za pomocą gradacyjnej analizy danych wykazało występowanie podobieństw pomiędzy czwartą i trzecią grupą wieku. Między pozostałymi grupami (pierwszą i drugą) oraz grupami trzecią

i czwartą traktowanymi łącznie jako nowa kategoria nie wykazano podobieństw. W przypadku grup wykluczenia cyfrowego przeprowadzona analiza pozwoliła wyodrębnić dwie główne kategorie: złożoną z grupy osób wykluczonych cyfrowo (GWC I) i złożoną z pozostałych grup nierówności cyfrowych (GWC II, III i IV).

Wykr. 2. Nadreprezentacja struktury wieku w grupach nierówności cyfrowych



Uwaga. *Rho* = 0,1679; *tau* = 0,1120.
Źródło: jak przy wykr. 1.

Analiza wyników otrzymanych dla drugiego rozpatrywanego czynnika, jakim była klasa miejscowości, wykazała, że wraz ze spadkiem liczby mieszkańców zwiększał się

udział w profilu wykluczenia grupy osób wykluczonych cyfrowo (GWC I) i jednocześnie zmniejszał się udział grupy osób w pełni korzystających z najnowszych rozwiązań ICT (GWC IV). W miastach poniżej 500 tys. mieszkańców oraz na wsi (KLM 2, 3 i 4) największy udział w profilu wykluczenia cyfrowego miała grupa osób wykluczonych cyfrowo (GWC I – przeciętna wartość wynosiła od 38,5% do 58,4%). Z kolei w miastach powyżej 500 tys. mieszkańców największy udział miała grupa osób w pełni korzystających z najnowszych rozwiązań ICT (GWC IV – 37,0%). We wszystkich klasach miejscowości druga i trzecia grupa nierówności cyfrowych cechowały się zbliżonym niskim udziałem (od 6,7% do 10,6%) (tabl. 2).

Tabl. 2. Profil wykluczenia cyfrowego w poszczególnych klasach miejscowości w latach 2003–2015 (przeciętna wartość poziomu wykluczenia)

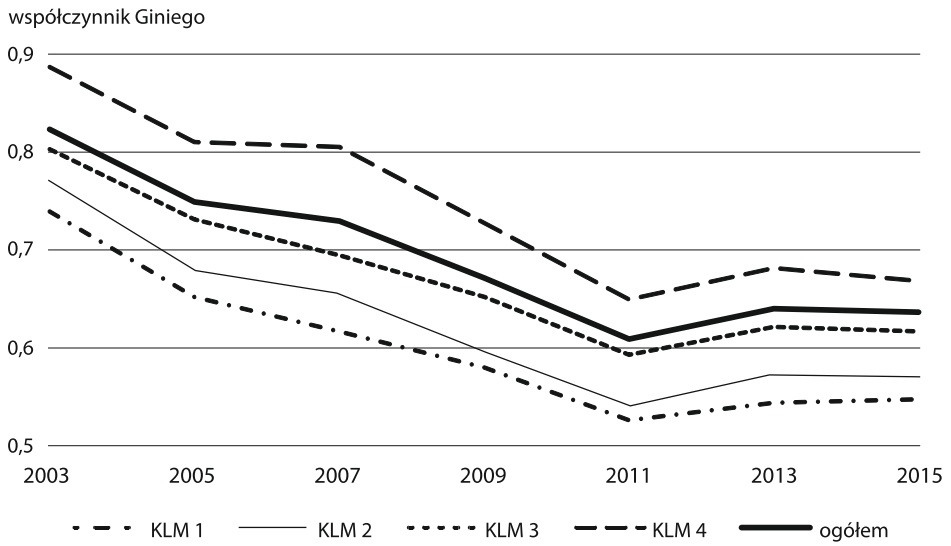
Grupy	KLM 1	KLM 2	KLM 3	KLM 4
	w %			
GWC I	33,0	38,5	46,5	58,4
GWC II	9,3	10,2	10,6	9,4
GWC III	8,2	9,0	8,6	6,7
GWC IV	37,0	29,9	21,8	13,0

Źródło: jak przy tabl. 1.

Przeprowadzone badanie wykazało również, że wraz ze wzrostem liczby mieszkańców zmniejszał się poziom zróżnicowania wykluczenia cyfrowego (wykr. 3 i 4). Tylko wśród osób mieszkających na wsi (przeciętna wartość współczynnika Giniego – 0,654) zaobserwowano większe zróżnicowanie niż dla wszystkich osób ogółem (0,615). W przypadku pozostałych klas miejscowości było ono niższe. Poziom zróżnicowania najbardziej zbliżony do ogólnopolskiego cechował grupę osób zamieszkujących miasta poniżej 100 tys. mieszkańców (0,597). Najniższym zróżnicowaniem wykluczenia cyfrowego charakteryzowały się osoby mieszkające w miastach od 100 tys. do 500 tys. mieszkańców (0,556) oraz osoby mieszkające w miastach powyżej 500 tys. mieszkańców (0,526). We wszystkich klasach miejscowości poziom zróżnicowania wykluczenia cyfrowego się zmniejszał. Zdecydowanie największy wpływ na całkowite zróżnicowanie miały wsie (29,0%) i miasta poniżej 100 tys. mieszkańców (27,2%). Miasta od 100 tys. do 500 tys. mieszkańców charakteryzowały się niższym wpływem (16,5%). Najmniejszy wpływ na całkowite zróżnicowanie wykluczenia

cyfrowego miały miasta powyżej 500 tys. mieszkańców (8,6%) oraz zróżnicowanie międzygrupowe (6,3%).

Wykr. 3. Tendencja dyspersji wykluczenia cyfrowego w poszczególnych klasach miejscowości



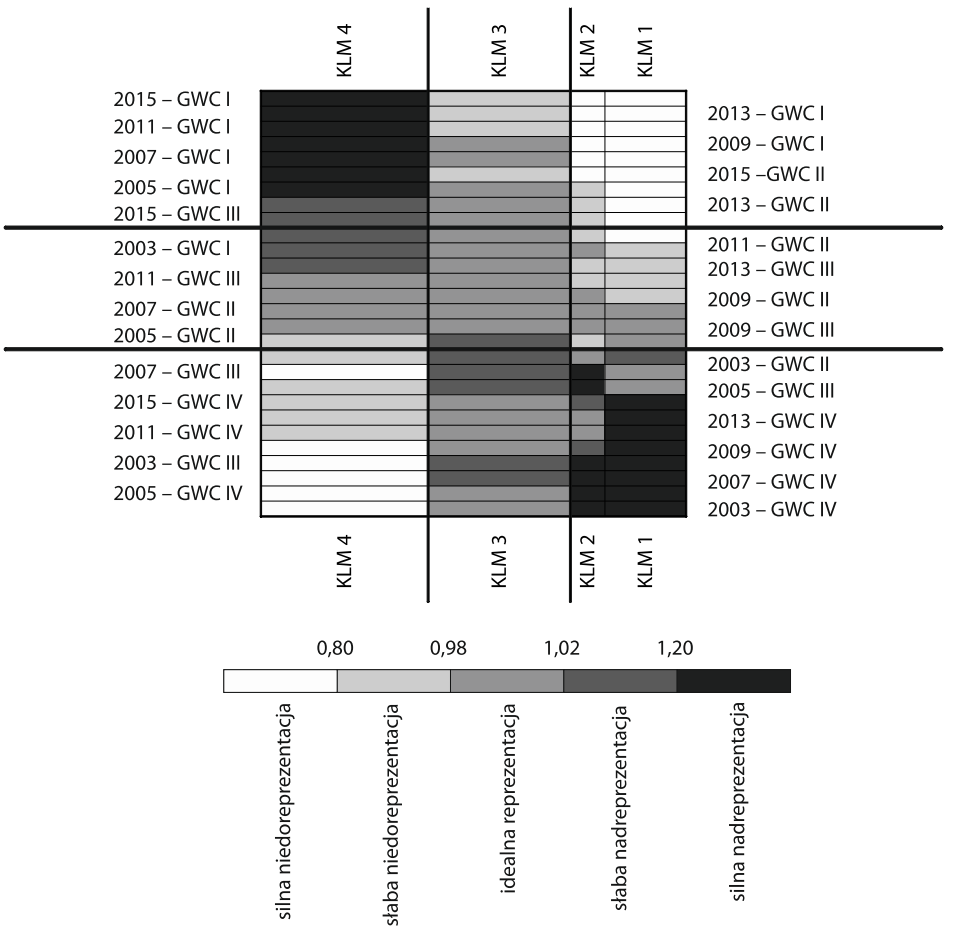
Źródło: jak przy wykr. 1.

Ocena nadreprezentacji potwierdziła, że we wszystkich klasach miejscowości występowało duże zróżnicowanie poziomu wykluczenia cyfrowego (wykr. 4). Grupa osób wykluczonych cyfrowo (GWC I) charakteryzowała się większym udziałem osób mieszkających na wsi i mniejszym udziałem osób mieszkających w miastach powyżej 100 tys. mieszkańców. Natomiast w grupie osób w pełni korzystających z najnowszych rozwiązań ICT (GWC IV) udział mieszkańców miast (KLM 1, 2 i 3), a udział mieszkańców wsi – większy. Struktura klas miejscowości w II i III grupie nierówności cyfrowych była zbliżona do struktury średniej.

Badanie poziomu wykluczenia cyfrowego w klasach miejscowości za pomocą gradacyjnej analizy danych wykazało także podobieństwa między miastami powyżej 500 tys. mieszkańców a miastami od 100 tys. do 500 tys. mieszkańców. Miasta poniżej 100 tys. mieszkańców, wsie oraz nowa kategoria (stworzona z pierwszej i drugiej klasy miejscowości) charakteryzowały się brakiem podobieństwa. W przypadku grup wykluczenia cyfrowego przeprowadzona analiza pozwoliła wyodrębnić następujące główne kategorie: pierwszą, w której dominowały osoby wykluczone; drugą, w której

dominowały osoby należące w poszczególnych latach badania do drugiej i trzeciej grupy nierówności cyfrowych; trzecią, w której dominowały osoby w pełni korzystające z najnowszych rozwiązań ICT.

Wykr. 4. Nadreprezentacja struktury miejscowości w grupach nierówności cyfrowych



Uwaga. *Rho* = 0,2187; *tau* = 0,1464.
Źródło: jak przy wykr. 1.

4. Podsumowanie

Coraz szybszy postęp technologiczny na świecie powoduje, że powszechniejszy staje się dostęp do technologii informacyjno-komunikacyjnych (ICT) będących fundamentalnym czynnikiem kształtującym obecny „wiek informacji". Rozwój ICT osiąg-

nął obecnie niespotykaną dotychczas skalę, ponieważ dotyczy zarówno sfery przedsiębiorstw, administracji, jak i gospodarstw domowych. Nie zachodzi jednak równomiernie, co prowadzi do zróżnicowania w dostępie do technologii ICT i ich wykorzystania, stanowiąc jedno z głównych zagrożeń dla zrównoważonego rozwoju i ładu społecznego.

W badaniu omawianym w niniejszym artykule wykorzystano autorską metodę estymacji opartą na podejściu całościowym. Zastosowanie takiej metody motywowane było brakiem miernika, który pozwoliłby na gruntowne zbadanie zjawiska nierówności cyfrowych w gospodarstwach domowych.

W badanym okresie 2003–2015 z czterech analizowanych czynników tylko wiek oraz klasa miejscowości wpływały na profil i zróżnicowanie wykluczenia cyfrowego. Zmienne płeć oraz region nie różnicowały skali wykluczenia cyfrowego, które było zbliżone do ogólnopolskiego. Z badania wynika, że osoby wykluczone cyfrowo to głównie osoby starsze, mieszkające na wsi, z kolei osoby w pełni korzystające z najnowszych rozwiązań ICT to przede wszystkim osoby młode, mieszkające w dużych miastach. Dominacja grupy osób wykluczonych cyfrowo oraz grupy osób w pełni korzystających z najnowszych rozwiązań ICT wpływała na skalę nierówności w posiadaniu i korzystaniu z ICT, która – mimo że stopniowo się zmniejszała – była bardzo wysoka. Zaobserwowano zależność pomiędzy wzrostem wieku oraz spadkiem liczby mieszkańców a rosnącym poziomem oraz zróżnicowaniem wykluczenia cyfrowego.

Omawiane badanie wpisuje się w nurt badań dotyczących nierówności, których celem jest zrozumienie mechanizmów wpływających na to zjawisko. Należy się spodziewać, że uzyskane wyniki pozwolą na lepsze poznanie wielowymiarowego problemu, jakim jest wykluczenie cyfrowe.

Bibliografia

Anand, S. (1983). *Inequality and Poverty in Malaysia: Measurement and Decomposition*. New York: Oxford University Press.

Brendesha, T. M., Kimberly, M. J. (2013). Black Youth Beyond the Digital Divide: Age and Gender Differences in Internet Use, Communication Patterns, and Victimization Experiences. *Journal of Black Psychology*, *40*(3), 291–307.

Bylok, F. (2012). Wpływ czynników demograficznych na przemiany konsumpcji w Polsce. *Studia Ekonomiczne / Uniwersytet Ekonomiczny w Katowicach*, (103), 17–32.

Czapiński, J., Panek, T. (red.). (2003–2015). Diagnoza społeczna. Pobrane z: www.diagnoza.com.

Dixon, L. J., Correa, T., Straubhaar, J., Covarrubias, L., Graber, D., Spence, J., Rojas, V. (2014). Gendered Space: The Digital Divide Between Male and Female Users in Internet Public Access Sites. *Journal of Computer-Mediated Communication*, *19*(4), 991–1009.

Koronacki, J., Mielniczuk, J. (2006). *Statystyka dla studentów kierunków technicznych i przyrodniczych*. Warszawa: Wydawnictwa Naukowo-Techniczne.

Kowalczyk, T., Pleszczyńska, E., Ruland, F. (2004). *Grade models and methods for data analysis. With applications for the analysis of data populations*. New York: Springer.

Podolec, B. (2008). Społeczno-ekonomiczne uwarunkowania sytuacji materialnej gospodarstw domowych. W: K. Jakóbik (red.), *Statystyka społeczna – dokonania, szanse, perspektywy* (s. 109–123). Warszawa: Główny Urząd Statystyczny.

Rószkiewicz, M. (2006). Tworzenie zabezpieczenia materialnego w świetle badań polskich gospodarstw domowych. *Gospodarka Narodowa*, (4), 69–86.

Sarama, M. (2011). Regionalne zróżnicowanie poziomu rozwoju społeczeństwa informacyjnego w Polsce. *Zeszyty Naukowe Uniwersytetu Szczecińskiego*, (67), 571–579.

Śmiałowski, T. (2019). Assessment of digital exclusion of Polish households. *Quantitative Methods in Economics*, *20*(1), 54–61.

Śmiałowski, T., Jałowiecki, P. (2012). Terytorialne zróżnicowanie obszarów wykluczeń technologicznych. *Roczniki Naukowe SERiA*, *14*(4), 125–131.

Śmiałowski, T., Ochnio, L. (2019). *Economic contexts of differences in digital exclusion*. *Acta Scientiarum Polonorum. Oeconomia*, *18*(2), 88–100.

Theil, H. (1967). *Economics and Information Theory*. Amsterdam: North Holland Publishing Company.

Turczak, A., Zwiech, P. (2013). Czynniki wpływające na strukturę wydatków konsumpcyjnych gospodarstw domowych w Polsce. *Zeszyty Naukowe Studia i Prace WNEiZ*, *2*(33), 189–207.

Van Dijk, J. A. G. M. (2012). The Evolution of the Digital Divide. The Digital Divide turns to Inequality of Skills and Usage. W: J. Bus, M. Crompton, M. Hildebrant, G. Metakides (red.), *Digital Enlightenment Yearbook* (s. 57–78). Amsterdam: IOS Press.

Witkowska, D., Matuszewska-Janica, A. (2013). Zróżnicowanie płac ze względu na płeć: zastosowanie drzew klasyfikacyjnych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (279), 58–66.

Witte, J., Kiss, M., Lynn, R. (2015). The internet and social inequalities in the U.S. W: M. Ragnedda, G. W. Muschert (red.), *The Digital Divide: The Internet and Social Inequality in International Perspective* (s. 67–85). London: Routledge.

Zillien, N., Marr, M. (2015). The digital divide in Europe. W: M. Ragnedda, G. W. Muschert (red.), *The Digital Divide: The Internet and Social Inequality in International Perspective* (s. 55–56). London: Routledge.

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 1, 46–53
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

Innowacje i badania innowacyjności[1]

Innovations and research on innovativeness

1. Wprowadzenie

Pojęcie innowacji może się odnosić zarówno do działania, jak i jego skutków. Kluczowe elementy koncepcji innowacji obejmują wiedzę (jako podstawę nowości i użyteczności) oraz tworzenie wartości lub wypracowanie określonego zachowania (jako celu). Nie należy jej utożsamiać z wynalazkiem czy wymogiem implementacji. Innowacja musi zostać wdrożona, tj. wprowadzona lub udostępniona do wykorzystania przez innych. Ogólna definicja innowacji w obu znaczeniach, na podstawie sformułowań zawartych w *Podręczniku Oslo* (*Oslo Manual 2018. Guidelines for Collecting, Reporting and Using Data on Innovation*), może brzmieć następująco: Innowacja to nowy lub ulepszony produkt lub proces (lub ich kombinacja), który różni się znacznie od wcześniejszych produktów lub procesów jednostki i który został udostępniony potencjalnym użytkownikom produktu lub wprowadzony do użycia przez jednostkę (OECD i Eurostat, 2018, s. 242, 246–248). W *Podręczniku Oslo* zdefiniowano również działania innowacyjne, innowację biznesową, innowację produktu oraz innowacje w zakresie procesów biznesowych (Kordos, 2019).

Pomiar i konceptualizacja innowacji są obecnie wystandaryzowane. Wiele krajów i organizacji międzynarodowych, uznając pomiar innowacji za potrzebny i ważny, stworzyło możliwości gromadzenia danych na ten temat. Ostatnie, czwarte wydanie *Podręcznika Oslo* (2018) wspiera te skoordynowane starania w celu uzyskania solidnych, porównywalnych w skali międzynarodowej danych, wskaźników i analiz. Opracowany w 1992 r. *Podręcznik* został trzykrotnie zaktualizowany ze względu na rozwój innowacji i zmieniające się potrzeby użytkowników. Najnowsza edycja uwzględnia postęp w pojmowaniu procesu innowacji i rozszerzenie zakresu badań na inne działy przemysłu i usług, a także zmiany międzynarodowych klasyfikacji standardowych.

Publikacja zawiera wytyczne do gromadzenia, prezentowania i interpretowania danych dotyczących innowacji, dzięki czemu łatwiejsze jest przeprowadzanie porównań na poziomie międzynarodowym. Tworzy również wspólną płaszczyznę badań i eksperymentalnych sposobów mierzenia innowacji. Wytyczne mają przede wszystkim wesprzeć krajowe urzędy statystyczne i innych producentów danych na temat innowacji w zakresie projektowania, gromadzenia, analizy i publikowania

[1] Artykuł stanowi drugą część opracowania, które powstało na podstawie referatu wygłoszonego na II Kongresie Statystyki Polskiej (Warszawa, 10–12 lipca 2018 r.). Pierwsza część została opublikowana w nr. 11/2019 „WS".

mierników innowacji. Celem jest zaspokajanie potrzeb badawczych i politycznych; ponadto wytyczne stanowią bezpośrednią wartość dla użytkowników informacji o innowacjach. Należy je traktować jako połączenie formalnych standardów statystycznych i porad dotyczących najlepszych praktyk, a także propozycje związane z rozszerzeniem pomiaru innowacji na nowe dziedziny.

Podręcznik Oslo odgrywa kluczową rolę w demonstrowaniu i komunikowaniu wielowymiarowego, a często ukrytego charakteru innowacji, chociaż kilka ważnych pytań badawczych i politycznych wymaga rozszerzonych i solidnych danych (Kordos, 2019). W czwartym wydaniu po raz pierwszy zapewniono wspólne podejście do pomiaru innowacji w całej gospodarce – w jednostkach rządowych, organizacjach pozarządowych i gospodarstwach domowych. Dokonano przeglądu zagadnień związanych z wykorzystaniem danych dotyczących innowacji do konstruowania wskaźników, a także prowadzenia analiz statystycznych i ekonometrycznych. Sformułowane na tej podstawie zalecenia są skierowane nie tylko do tych, którzy opracowują wskaźniki o oficjalnym charakterze, ale stanowią pomoc w ukierunkowaniu pracy osób zaangażowanych w projektowanie, produkcję i wykorzystanie wskaźników innowacji. W *Podręczniku* opisano metody analizy danych, ze szczególnym uwzględnieniem oceny wpływu innowacji i empirycznej oceny rządowych polityk innowacyjnych. Publikacja nadaje więc kierunek gromadzeniu i analizom danych dotyczących innowacji, a także zachęca do poszukiwań w celu poprawy jakości, widoczności oraz przydatności danych i wskaźników pochodzących z badań w zakresie innowacji.

2. Innowacje i innowacyjność

Pojęcie *innowacja* pochodzi od łacińskiego słowa *innovatio*, oznaczającego odnowienie. Tym terminem można objąć wszystko, co nowe, czyli zmiany techniczne, technologiczne, organizacyjne, zmiany w systemach zarządzania i komunikacji międzyludzkiej, w sferze mediów czy mody, a także w sposobie myślenia. Innowacje zmieniają istniejący stan rzeczy. Są przedmiotem zainteresowania wielu dyscyplin – nauk technicznych, ekonomicznych, społecznych czy nauki o zarządzaniu. Innowacyjne idee stają się innowacjami dopiero wtedy, gdy są wdrażane do praktyki – co wymaga ogromnej pracy i starannego przygotowania. Wywołują dalekosiężne skutki we wszystkich dziedzinach życia, stanowiąc decydujący czynnik rozwoju poszczególnych firm, a nawet całej gospodarki.

Produkty, procesy oraz metody organizacyjne i marketingowe nie muszą być nowością dla rynku, na którym działa przedsiębiorstwo, ale muszą być nowe przynajmniej dla samego przedsiębiorstwa. Mogą być opracowane przez dane przedsiębiorstwo albo powstać za sprawą innej firmy bądź jednostki o odmiennym charakte-

rze (np. instytutu naukowo-badawczego, ośrodka badawczo-rozwojowego, szkoły wyższej), przy czym warunkiem koniecznym jest ich wykorzystanie w praktyce.

Działalność innowacyjna polega na angażowaniu się przedsiębiorstw w różnego rodzaju działania naukowe, techniczne, organizacyjne, finansowe i komercyjne, które prowadzą lub mają w zamierzeniu doprowadzić do wdrażania innowacji. Niektóre z tych czynności mają charakter innowacyjny, inne – choć nie stanowią nowości – są konieczne do wprowadzenia innowacji w życie. Dotyczy to także działalności badawczej i rozwojowej (B+R), która nie jest bezpośrednio związana z tworzeniem konkretnej innowacji.

Innowacja to zatem ciąg działań zmierzających do wytworzenia nowych lub ulepszenia istniejących produktów, procesów technologicznych albo systemów organizacyjnych. Schumpeter już w 1912 r. próbował zdefiniować ten termin, wprowadzając go do ekonomii (Schumpeter, 1960). Określił pięć przypadków występowania innowacji:

- stworzenie nowego produktu;
- zastosowanie nowej technologii, metody produkcji;
- stworzenie nowego rynku zbytu;
- pozyskanie nieznanych dotąd surowców;
- reorganizacja określonej gałęzi gospodarki.

Za proces innowacyjny należy uznać stopniowe wprowadzanie zarządzania jakością, które jest traktowane jako integralny element składowy procesu zarządzania organizacją, obejmującego pięć funkcji zarządzania, czyli planowanie, organizowanie, kierowanie ludźmi, kontrolowanie i doskonalenie. Funkcje te są wzajemnie powiązane, przy czym odpowiednią jakość osiąga się – według aktualnych zasad zarządzania nią – przez planowanie, a nie przez inspekcję. Na etapie planowania określa się, co, jak i kiedy uzyskać, oraz sprawdza się możliwości realizacji. Dużego znaczenia nabierają więc badania i audyty, na podstawie których można uzyskać informacje dające podstawę porównań i ocen. Kontrola jakości (poczynając już od kontroli wstępnej) polega na rejestrowaniu wyników uzyskiwanych przez organizację w celu ustalenia, czy spełniają one normy jakościowe. Ważne jest także wskazanie sposobów eliminowania przyczyn niezadowalających wyników. Monitorowaniem objęte są wdrażane plany oraz stopień i jakość ich zrealizowania.

Zarządzanie jakością obejmuje całą organizację, ponieważ każde działanie w przedsiębiorstwie może mieć wpływ na jakość produktu. W zarządzaniu jakością każdy pracownik powinien dążyć do doskonalenia własnej pracy. Staje się to szczególnie widoczne w świetle koncepcji zarządzania przez jakość (Berdowski, 2014).

Zarządzanie innowacjami zawiera w sobie decyzje, działania i praktyki, za sprawą których realizacja danego pomysłu generuje nowe wartości. W przypadku zarządzania danymi statystycznymi innowacje mają kluczowe znaczenie dla długoterminowego sukcesu. W dzisiejszym szybko zmieniającym się środowisku biznesowym

efektywne zarządzanie innowacjami stało się niezbędnym warunkiem utrzymania konkurencyjności (Birkinshaw, Hamel i Mol, 2008).

Innowacyjność, dzięki rosnącemu rozumieniu innowacji, stała się ważnym punktem programu politycznego w większości krajów rozwiniętych. Polityka innowacyjna wyrosła przede wszystkim z polityki naukowej i technologicznej, ale wchłonęła również istotne elementy polityki przemysłowej. Początkowo zakładano, że postęp technologiczny zostanie osiągnięty poprzez prosty proces liniowy, zaczynający się od podstawowych badań naukowych, systematycznie kontynuowany w naukach stosowanych, a kończący się na zastosowaniach praktycznych i marketingowych. Naukę postrzegano jako siłę napędową, dlatego rząd potrzebował polityki naukowej. Nowatorskie myślenie o innowacjach uwydatniło funkcję systemów i doprowadziło do bardziej zintegrowanego podejścia do realizacji polityk związanych z innowacją.

3. Innowacje w statystyce

Innowacje można obserwować także w statystyce w postaci wprowadzania rozwiązań służących doskonaleniu prac instytucji lub urzędu, organizacji badań statystycznych, metod zbierania danych, sposobów ich opracowywania, analizy, dostępności, prezentacji, rozpowszechnienia oraz publikacji. Warto tu zwrócić uwagę na proces innowacyjny, który jest wyznaczany następującymi po sobie fazami – od powstania innowacyjnej idei do jej wdrożenia i komercjalizacji – a więc jest to zespół działań służących wprowadzeniu nowych rozwiązań w sferach technicznej, technologicznej, organizacyjnej i społecznej.

Innowacje w zakresie procesów statystycznych to nowy lub ulepszony proces statystyczny dla jednej lub wielu funkcji statystycznych, który różni się znacznie od wcześniejszych procesów. Tak więc proces innowacyjny w statystyce to następujące po sobie procedury dotyczące sformułowania celu badania, uzasadnienia jego podstaw teoretycznych, wyboru odpowiednich działań niezbędnych do jego realizacji oraz sposobów ich wykonania prowadzących do upowszechnienia uzyskanych rezultatów.

Każdy proces polega na przekształcaniu zasobów w produkty czy usługi. Problemem badawczym powinny być wyjaśnienie i opis prawidłowości rządzących interakcjami zachodzącymi w ramach tego procesu. Składają się na nie: koordynacja, komunikacja i podejmowanie decyzji, dzięki którym zasoby są przekształcane w produkty. Z tego powodu Organizacja Współpracy Gospodarczej i Rozwoju (OECD) i Eurostat podjęły na szerszą skalę badania innowacyjności w przedsiębiorstwach w różnych gałęziach przemysłu, a także w usługach.

Coraz częściej wyrażana jest opinia, że innowacja stanowi dla krajów członkowskich Unii Europejskiej (UE) podstawę trwałego wzrostu gospodarczego oraz popra-

wy warunków ekonomicznych i społecznych. Przyjmuje się, że polityka pomocy państwa w sferze badań i innowacji może przyczynić się do zwiększenia innowacyjności gospodarki nie tylko przez ochronę konkurencyjności rynkowej produktów jako stymulatora innowacyjności, lecz także przez ustanowienie ram ułatwiających państwom członkowskim opracowanie skutecznych form pomocy na rzecz innowacji.

4. Badania innowacyjności

Badania innowacyjności przeprowadzane są (co dwa lata) w całej UE, niektórych krajach European Free Trade Association (EFTA, Europejskie Stowarzyszenie Wolnego Handlu) i krajach kandydujących do UE. Koncentrują się na działalności innowacyjnej w przemyśle i usługach.

Unijne badania innowacyjności są narzędziem oceny wyników innowacyjnych w krajach członkowskich UE i podkreślają mocne i słabe strony ich systemów badań i innowacji[2]. Pomagają również w monitorowaniu wdrażania innowacji. W znacznym stopniu są oparte na *Podręczniku Oslo*. Zestaw wyników europejskiej innowacyjności opiera się na trzech typach wskaźników i ośmiu wymiarach innowacji. Aby sprostać potrzebom decydentów i społeczności naukowej, Eurostat zbiera dane statystyczne dotyczące innowacji z urzędów statystycznych krajów członkowskich. Zgromadzone dane statystyczne są ściśle powiązane z działaniami UE, ponieważ wskaźniki innowacji wykorzystuje się jako narzędzie do podejmowania decyzji; są one także pomocne w ocenie takich inicjatyw, jak Unia Innowacji czy Europejska Przestrzeń Badawcza (EPB) w kontekście strategii *Europa 2020*.

Warto podkreślić, że Unia Innowacji ma trzy cele:

- uczynienie z Europy ośrodka naukowego światowej rangi;
- usunięcie przeszkód w szybkim wprowadzaniu innowacji na rynek, takich jak kosztowne patentowanie, rozdrobnienie rynku, powolne ustanawianie standardów i brak kompetencji;
- zrewolucjonizowanie sposobu, w jaki współpracują ze sobą sektory publiczny i prywatny, m.in. poprzez partnerstwa innowacyjne między instytucjami europejskimi, władzami krajowymi i regionalnymi oraz biznesem.

Unia Innowacji obejmuje ponad 30 punktów działania. Przykładowo jednym z nich jest stymulowanie innowacji w Europie poprzez zwiększenie dostępu do finansowania dla innowacyjnych firm.

Zgodnie z zaleceniami Eurostatu GUS prowadzi badania innowacyjności w przemyśle i usługach. Badania te są realizowane i analizowane przez Urząd Statystyczny w Szczecinie (GUS i US Szczecin, 2010, 2011, 2012a, 2012b, 2013, 2014, 2015, 2016).

[2] http://europa.eu/rapid/press-release_IP-16-2486_pl.htm.

Publikacja *Działalność innowacyjna przedsiębiorstw w latach 2014–2016* (GUS i US Szczecin, 2017) prezentuje wyniki badań, w których wykorzystano metodykę opracowaną przez Eurostat i OECD, a omówioną w *Podręczniku Oslo.* Badania zostały przeprowadzone przez GUS w ramach Wspólnotowego Badania Innowacji (Community Innovation Survey, CIS) na podstawie formularza modelowego opracowanego przez centralny urząd statystyczny Norwegii i centralne urzędy statystyczne innych krajów UE. Zakres przedmiotowy badań został rozszerzony w porównaniu z poprzednią edycją (za lata 2013–2015), która służyła zaspokojeniu jedynie krajowych potrzeb informacyjnych.

5. Uwagi końcowe

Innowacje i innowacyjność należą do bardzo złożonych zagadnień i pomimo bogatej literatury ekonomicznej na ich temat, zarówno w wymiarze teoretycznym, jak i w podejściu praktycznym, nadal wymagają pogłębionych analiz i badań.

Dotychczas nie ustalono standardowej metodyki badań innowacyjności w sektorze publicznym nierynkowym, do którego zalicza się GUS. Ostatnie wydanie *Podręcznika Oslo* ułatwi to zadanie. Metodologia tam przedstawiona jest bez wątpienia udaną propozycją bezpośredniego pomiaru działalności innowacyjnej, wykraczającą poza tzw. linearny model innowacji. Opiera się ona na założeniu, że działalność innowacyjna jest dynamicznym procesem ciągłym, czyli fenomenem znacznie trudniejszym do pomiaru niż zjawiska, które zwykle bada statystyka.

W kontekście współczesnych globalnych uwarunkowań innowacje coraz częściej stają się podstawowym narzędziem rozwoju nie tylko podmiotów gospodarczych, lecz także regionów i krajów, ponieważ umożliwiają tworzenie przewagi konkurencyjnej na poziomie przedsiębiorstw, sektorów i całych systemów gospodarczych. W okresie silnego wzrostu gospodarczego w efekcie postępu technicznego i informatycznego wdrażanie, rozpowszechnianie i wykorzystywanie rozwiązań innowacyjnych należy traktować nie tylko jako konieczność mającą zagwarantować utrzymanie czy poprawę własnej konkurencyjności, ale przede wszystkim jako przejaw nowoczesności organizacji podejmujących i realizujących takie przedsięwzięcia.

Organizacje statystyczne coraz częściej dążą do rozwijania nowego rodzaju partnerstwa z dostawcami danych, środowiskiem akademickim, krajowymi i lokalnymi organami rządowymi, sektorem prywatnym i innymi jednostkami, aby badać rozwój innowacji. Jednak wiele organizacji przyznaje, że ich doświadczenie w skutecznym rozwijaniu takiego partnerstwa i zarządzaniu nim jest ograniczone.

Warto dodać, że w sytuacji niepewności korzystanie ze wszelkich użytecznych i wiarygodnych informacji należy uznać za racjonalne. Jak zauważa Szreder (2017), wkraczania big data w dziedziny, w których wcześniej dominowały badania statystycz-

ne (pełne lub próbkowe), nie trzeba rozpatrywać w perspektywie rywalizacji tych dwu źródeł danych, ponieważ właściwa jest im relacja komplementarności. Big data, podobnie jak rejestry administracyjne, mogą stanowić – i w praktyce już stanowią – wartościowe dopełnienie badań opartych na próbie. W szczególności mogą one dostarczyć ważnych informacji w sytuacji zagrożenia badania reprezentacyjnego dużymi błędami nielosowymi, np. błędami pokrycia lub braków odpowiedzi. Innymi słowy, dodatkowe informacje o populacji, potrzebne do efektywnego zastosowania mechanizmów ważenia danych z próby lub kalibracji danych, mogą mieć swoje źródło w big data.

Unia Europejska koncentruje się na kilku kluczowych cechach dotyczących rozwoju innowacji wdrażanych przez przedsiębiorstwa. Szczególną rolę w zakresie jakości danych, a także badań innowacyjnych, odgrywają OECD (2009, 2010, 2015, 2016) i Eurostat (1996, 2007, 2009, 2013). Badania innowacyjne stanowią część strategii *Europa 2020* ze względu na rolę UE w tworzeniu możliwości zatrudnienia, zwiększania konkurencyjności przedsiębiorstw na rynkach globalnych, poprawie jakości życia i przyczynianiu się do bardziej zrównoważonego wzrostu gospodarczego. Polityka UE często koncentruje się na zachęcaniu do innowacji i stymulowaniu ich. Wspólnotowe Badanie Innowacji (CIS) dostarcza danych statystycznych, które analizowane są według rodzaju innowatorów, działalności gospodarczej i klasy wielkości przedsiębiorstwa.

W ostatnich latach podjęto intensywne prace badawcze nad innowacjami i innowacyjnością w statystyce oficjalnej. Potrzebę ich prowadzenia zwiększyły globalizacja i przyspieszony rozwój technologiczny.

Bibliografia

Berdowski, J. B. (2014). *Zintegrowany system zarządzania jakością*. Warszawa: Europejska Uczelnia Informatyczno-Ekonomiczna.

Birkinshaw, J., Hamel, G., Mol, M. J. (2008). Management Innovation. *Academy of Management Review, 33*(4), 825–845. Pobrane z: http://faculty.london.edu/jbirkinshaw/assets/documents/5034421969.pdf.

Eurostat. (1996). *The Future of European Social Statistics: Use of Administrative Registers and Dissemination Strategies*. Luxembourg: Eurostat.

Eurostat. (2007). *Handbook on Data Quality Assessment: Methods and Tools*. Luxembourg: Eurostat.

Eurostat. (2009*). Handbook on Quality Report.* Luxembourg: Eurostat.

Eurostat. (2013). *Handbook on improving quality by analysis of process variables*. Luxembourg: Eurostat.

GUS. (2011). *Metodologia badania budżetów gospodarstw domowych*. Warszawa: Główny Urząd Statystyczny.

GUS, US Szczecin. (2010). *Działalność innowacyjna przedsiębiorstw w latach 2006–2009*. Warszawa: Główny Urząd Statystyczny.

GUS, US Szczecin. (2012a). *Działalność innowacyjna przedsiębiorstw w latach 2008–2010*. Warszawa: Główny Urząd Statystyczny.

GUS, US Szczecin. (2012b). *Działalność innowacyjna przedsiębiorstw w latach 2009–2011*. Warszawa: Główny Urząd Statystyczny.

GUS, US Szczecin. (2013). *Działalność innowacyjna przedsiębiorstw w latach 2010–2012*. Warszawa: Główny Urząd Statystyczny.

GUS, US Szczecin. (2014). *Działalność innowacyjna przedsiębiorstw w latach 2011–2013*. Warszawa: Główny Urząd Statystyczny.

GUS, US Szczecin. (2015). *Działalność innowacyjna przedsiębiorstw w latach 2012–2014*. Warszawa: Główny Urząd Statystyczny.

GUS, US Szczecin. (2016). *Działalność innowacyjna przedsiębiorstw w latach 2013–2015*. Warszawa: Główny Urząd Statystyczny.

GUS, US Szczecin. (2017). *Działalność innowacyjna przedsiębiorstw w latach 2014–2016*. Warszawa: Główny Urząd Statystyczny.

Kordos, J. (2019). Pomiar i wykorzystanie innowacji. Czwarte wydanie Podręcznika Oslo. *Wiadomości Statystyczne, 64*(4), 85–88.

OECD. (2009). *Innovation in Firms: A Microeconomic Perspective*. Paris: OECD Publishing. DOI: https://doi.org/10.1787/9789264056213-en.

OECD. (2010). *Measuring Innovation: A New Perspective*. Paris: OECD Publishing. DOI: https://doi.org/10.1787/9789264059474-en.

OECD. (2015). *The Innovation Imperative: Contributing to Productivity, Growth and Well-Being*. Paris: OECD Publishing. DOI: https:/doi.org/10.1787/9789264239814-en.

OECD. (2016). System innovation. *OECD Science, Technology and Innovation Outlook 2016*. DOI: https://doi.org/10.1787/sti_in_outlook-2016-9-en.

OECD, Eurostat (2018). *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation*. Paris, Luxembourg: OECD Publishing – Eurostat. DOI: https://doi.org/10.1787/9789264304604-en.

Schumpeter, J. A. (1960). *Teoria rozwoju gospodarczego*. Warszawa: PWN.

Szreder, M. (2017). Nowe źródła informacji i ich wykorzystywanie w podejmowaniu decyzji. *Wiadomości Statystyczne*, *62*(7), 5–17.

Jan Kordos (dawniej Szkoła Główna Handlowa w Warszawie)

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 1, 54–61
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

XXXVIII Międzynarodowa Konferencja Naukowa Wielowymiarowa Analiza Statystyczna WAS 2019

The 38th International Scientific Conference on Multivariate Statistical Analysis MSA 2019

W dniach 4–6 listopada 2019 r. odbyła się w Centrum Szkoleniowo-Konferencyjnym Uniwersytetu Łódzkiego (UŁ) XXXVIII Międzynarodowa Konferencja Naukowa Wielowymiarowa Analiza Statystyczna WAS 2019 (The 38th International Scientific Conference on Multivariate Statistical Analysis MSA 2019). Zorganizowały ją: Katedra Metod Statystycznych UŁ, Instytut Statystyki i Demografii UŁ, łódzki oddział Polskiego Towarzystwa Statystycznego oraz Komitet Statystyki i Ekonometrii Polskiej Akademii Nauk (PAN). Komitetowi Naukowemu przewodniczył prof. dr hab. Czesław Domański; funkcję sekretarzy naukowych konferencji pełniły dr hab. Aleksandra Baszczyńska i dr hab. Katarzyna Bolonek-Lasoń. Finansowego wsparcia przy organizacji wydarzenia udzieliły Ministerstwo Nauki i Szkolnictwa Wyższego (MNiSW)[1], PAN oraz StatSoft Polska.

Głównymi celami konferencji były prezentacja najnowszych osiągnięć z zakresu wielowymiarowej analizy statystycznej oraz wymiana doświadczeń będących wynikiem jej stosowania. Poruszone zostały następujące zagadnienia: rozkłady wielowymiarowe, testy statystyczne, metody nieparametryczne, analiza czynnikowa, analiza skupień, analiza dyskryminacyjna, analiza wariancji i regresji, metody bayesowskie, analizy Monte Carlo, data mining, procedury odporne, analiza danych cenzurowanych, rozpoznawanie obrazów, analizy stochastyczne oraz zastosowanie metod statystycznych w naukach ekonomicznych, marketingu, finansach, ubezpieczeniach, rynku kapitałowym i zarządzaniu ryzykiem.

W konferencji wzięły udział 72 osoby z ośrodków akademickich w Gdańsku, Katowicach, Krakowie, Łodzi, Poznaniu, Rzeszowie, Szczecinie, Warszawie i we Wrocławiu, przedstawiciele Głównego Urzędu Statystycznego (GUS) i urzędów statystycznych w Łodzi, Poznaniu i Rzeszowie, a także goście z Włoch i Niemiec. Łącznie odbyło się 15 sesji plenarnych i panelowych, na których wygłoszono 42 referaty. Konferencję zainaugurowały wystąpienia Czesława Domańskiego, rektora UŁ prof. dr. hab. Antoniego Różalskiego oraz prodziekana Wydziału Ekonomiczno-Socjologicznego UŁ dr. hab. Michała Przybylińskiego, prof. UŁ.

[1] Organizacja konferencji Multivariate Statistical Analysis 2019 została sfinansowana w ramach umowy 712/P-DUN/202019 ze środków MNiSW przeznaczonych na działalność upowszechniającą naukę.

W pierwszej sesji plenarnej, której przewodniczył Czesław Domański, przedstawiono referaty dotyczące aktualnych problemów statystyki wielowymiarowej. Prof. dr hab. inż. Jacek Wesołowski (Politechnika Warszawska, GUS) wykazał w wystąpieniu *Optymalna alokacja próbki w schematach warstwowych – metody algebry liniowej i algorytmy*, że powszechnie stosowany rekurencyjny algorytm Neymana zapewnia optymalność alokacji w schemacie warstwowym z ograniczeniami na liczebność próbki w warstwach. Następnie prof. dr hab. Mirosław Krzyśko (Uniwersytet im. Adama Mickiewicza w Poznaniu) przedstawił w referacie *Jądrowe współrzędne dyskryminacyjne w przypadku geograficznie ważonych danych czasowo-przestrzennych z wyborem zmiennych* rozszerzenie analizy jądrowych współrzędnych dyskryminacyjnych na dane czasowo-przestrzenne oraz ważone dane czasowo-przestrzenne wraz z propozycją procedury wyboru zmiennych.

Drugą sesję, pod przewodnictwem Mirosława Krzyśki, poświęcono historii polskiej statystyki. Czesław Domański przybliżył sylwetkę Jakuba Kazimierza Haura, a dr hab. Jerzy T. Kowaleski, prof. UŁ – Marcina Kromera.

Trzecia sesja plenarna miała charakter wspomnieniowy. Przywołano postaci wybitnych polskich statystyków, którzy odeszli w 2017 i 2018 r. Mirosław Krzyśko wspominał Krystynę Katulską, dr Ewa Wycinka (Uniwersytet Gdański, UG) – Mirosława Krzysztofiaka, prof. dr hab. Janusz Wywiał (Uniwersytet Ekonomiczny w Katowicach, UE w Katowicach) – Józefa Kolonkę, zaś dr hab. Stanisław Wanat, prof. UEK (Uniwersytet Ekonomiczny w Krakowie) – Stanisława Wydymusa i Michała Majora.

Następnie rozpoczęły się sesje panelowe. Sesji IV A, prowadzonej w języku angielskim, przewodniczył prof. dr hab. Marek Walesiak (Uniwersytet Ekonomiczny we Wrocławiu, UEW). Dr hab. inż. Jacek Białek, prof. UŁ (GUS), omówił w referacie *Chain drift problem in the CPI measurement based on scanner data* wyniki badań symulacyjnych, które wskazują na sytuacje na rynku generujące największe obciążenie. Występuje ono wtedy, gdy indeks cen różni się od jedności w momencie powrotu cen produktów do wartości bazowej. Dr hab. Tomasz Żądło, prof. UE w Katowicach, w referacie *On generalization of Quatember's bootstrap* zaprezentował uogólnienie algorytmu Quatembera, pozwalającego na bezpośrednie ponowne próbkowanie z oryginalnej próby, oraz przedstawił wyniki badania jego własności wraz z porównaniem z aktualnymi konkurencyjnymi metodami. Referat *Asymptotic properties of duration-based VaR backtests* wygłoszony przez dr Martę Małecką (UŁ) dotyczył własności granicznych testu geometrycznego i geometrycznego testu VaR, zbudowanych na zasadzie ilorazu wiarygodności, jak również własności testu Giniego. Autorka wyprowadziła odpowiednie rozkłady asymptotyczne i określiła użyteczność otrzymanych wyników dla prób o realistycznych rozmiarach.

Podczas sesji IV B, której przewodniczyła dr hab. Iwona Markowicz, prof. US (Uniwersytet Szczeciński), wygłoszono cztery referaty. W prezentacji *Alokacja próby w problemie estymacji wskaźnika struktury w populacjach skończonych* Dominik Sieradzki i prof. dr hab. Wojciech Zieliński (Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, SGGW) przeprowadzili porównanie dokładności szacowania w zależności od wyboru alokacji próby. Przedstawili także wyniki badań, w których zestawili swoją propozycję ze znaną alokacją Neymana oraz alokacją proporcjonalną. Ewa Wycinka i dr hab. Beata Jackowska, prof. UG, w wystąpieniu *Estymacja rozkładu czasu istnienia przedsiębiorstw z uwzględnieniem rodzaju zdarzenia kończącego działalność* omówiły rezultaty badania zastosowania estymatorów uwzględniających rodzaj zdarzenia kończącego działalność w modelowaniu rozkładów czasu istnienia przedsiębiorstw. W badaniu porównano własności estymatorów dla zdarzeń konkurujących: naiwnego estymatora Kaplana-Meiera, estymatora Aalena-Johansena oraz estymatora uwzględniającego informacyjność cenzurowania IPCW (Inverse Probability Censoring Weighted). Następnie dr Stanisław Jaworski (SGGW) wygłosił *Kilka uwag o estymacji poziomu bezrobocia w Polsce*, opisując estymację poziomu bezrobocia za pomocą modelu strukturalnych szeregów czasowych. Ostatnim referatem w tej sesji było wystąpienie Janusza L. Wywiała i Grzegorza Sitka (UE w Katowicach) *Wariancja wartości własnych macierzy*. Autorzy rozważali problem dokładności estymacji wariancji maksymalnych wartości własnych macierzy kowariancji, omówili konstrukcję tego estymatora i zaproponowali metodę aproksymacji wariancji estymatora korelacji kanonicznej.

Sesję V A (w języku angielskim), pod przewodnictwem dr. hab. Wojciecha Gamrota, prof. UE w Katowicach, otworzył referat *On EBLUP under some linear mixed model with correlated random effects*, którego autorka Małgorzata Krzciuk (UE w Katowicach) przedstawiła wnioski z badań symulacyjnych wpływu występowania zależności między efektami losowymi na własności najlepszego liniowego nieobciążonego predyktora empirycznego dla liniowego modelu mieszanego z dwoma skorelowanymi efektami losowymi. Następnie dr Maciej Beręsewicz (Uniwersytet Ekonomiczny w Poznaniu, UEP, Urząd Statystyczny w Poznaniu) wygłosił przygotowany z Katarzyną Zadrogą (UEP) referat *Estimation of the number of illegally residing foreigners in Poland in 2017 i 2018 using Bayesian non-linear mixed count regression models*. Jego przedmiotem było oszacowanie liczby cudzoziemców nielegalnie przebywających w Polsce w latach 2017–2018, w którym zastosowano bayesowski nieliniowy model mieszany dla danych dyskretnych, wykorzystujący wyłącznie dane zagregowane raportowane przez straż graniczną i policję oraz rejestr PESEL.

Prowadzonej równolegle sesji V B przewodniczył Wojciech Zieliński. W sesji tej dr Wioletta Grzenda (Szkoła Główna Handlowa w Warszawie, SGH) wygłosiła re-

ferat *Bayesowskie wielomianowe modele logitowe dla kategorii nieuporządkowanych w badaniu sytuacji osób młodych na rynku pracy w Polsce*, w którym przedstawiła wyniki badań sytuacji respondentów na rynku pracy z zastosowaniem dwumianowego modelu logitowego. Prelegentka wykazała, że kontynuacja nauki często wynika z problemów ze znalezieniem pracy, a łączenie pracy zawodowej z nauką nie należy do preferowanych form aktywności zawodowej osób młodych. Referat *Triady czy tetrady? Porównanie niepełnych metod pomiaru podobieństwa preferencji,* przedstawiony przez dr. hab. Artura Zaborskiego (UEW) dotyczył zestawienia dwóch niepełnych metod pomiaru podobieństwa preferencji, tj. metody triad oraz metody tetrad. Zostały one porównane ze względu na ich pracochłonność oraz zdolność odwzorowania znanej struktury obiektów, także w przypadku kiedy nie jest spełniony warunek co do równej liczby par obiektów w podgrupach prezentowanych respondentom. Prezentacja *Uogólniona ekonometria entropii krzyżowej a sprzeczne transgraniczne (duże) źródła danych. Aktualizacja rachunków narodowych* autorstwa dr. Seconda Bwanakare (Wyższa Szkoła Informatyki i Zarządzania w Rzeszowie) oraz dr. Marka Cierpiał-Wolana (Uniwersytet Rzeszowski, Urząd Statystyczny w Rzeszowie) dotyczyła skutecznej metody łączenia danych z różnych źródeł oraz porównania wyników z techniką stosowaną dotychczas w oficjalnych statystykach. Autorzy podzielili się wynikami badań, w których zastosowali związaną z prawem potęgowym metodę dywergencji informacji Kullbacka-Leiblera – znaną z uogólnienia entropii Shannona – do rozwiązania nieliniowych, źle uwarunkowanych problemów odwrotnych za pomocą teorii bayesowskiej.

Drugiego dnia konferencji odbyły się dwie sesje plenarne i dwie równoległe sesje panelowe. Podczas pierwszej sesji plenarnej (w języku angielskim), poprowadzonej przez dr hab. Alinę Jędrzejczak, prof. UŁ, wygłoszono wykłady gościnne. Dr hab. Francesca Greselin (University of Milan-Bicocca) przedstawiła prezentację *Advances in learning from contaminated datasets,* przygotowaną wspólnie z Andreą Cappozzo (University of Milan-Bicocca) i prof. dr. hab. Thomasem Brendanem Murphym (University College Dublin). Dotyczyła ona modyfikacji procedury analizy dyskryminacyjnej, stosowanej w przypadku gdy dane rzeczywiste niektórych jednostkek w zestawie uczenia się mogą być zawodne (zjawisko definiowane jako szum etykiety) lub część obserwacji może odbiegać od głównej struktury danych (co jest określane jako wartości odstające). Wykład prof. dr. hab. Hansa-Joachima Mittaga (University of Hagen) *A new virtual library containing interactive learning objects for statistics education* zawierał opis projektu służącego opracowaniu interaktywnych obiektów edukacyjnych do celów edukacji statystycznej. Autor zapoznał publiczność z wynikami projektu, opisał dwie biblioteki wirtualne oraz przedstawił dotychczasowe doświadczenia z różnymi scenariuszami uczenia się mieszanego, a także perspektywy

współpracy międzynarodowej. Prof. dr hab. Józef Dziechciarz i dr Marta Dziechciarz-Duda (UEW) w wystąpieniu *Selected aspects of households' well-being measurement* dokonali dogłębnej analizy problemów metodologicznych pomiaru dobrostanu gospodarstw domowych i przedstawili propozycje dotyczące tego zagadnienia.

Drugą sesję, której przewodniczył prof. dr hab. Grzegorz Kończak (UE w Katowicach), otworzyło wystąpienie Iwony Markowicz *Rozbieżności w handlu wewnątrzwspólnotowym: przypadek Polski*, opracowane wspólnie z dr. Pawłem Baranem (US). Prelegentka przedstawiła wyniki analizy rozbieżności danych w handlu Polski w relacjach: Polska – kraj UE (relacje dwustronne) i Polska – kraje UE (relacja zagregowana kraj – kraje). Dr hab. Grażyna Dehnel, prof. UEP, i Marek Walesiak (UEW) w wystąpieniu *Ocena spójności społecznej województw Polski na podstawie danych klasycznych oraz symbolicznych interwałowych* przedstawili ocenę spójności społecznej województw w 2018 r., opracowaną z wykorzystaniem podejścia hybrydowego, łączącego zastosowanie skalowania wielowymiarowego z porządkowaniem liniowym. Dr hab. Beata Bieszk-Stolorz, prof. US, wygłosiła referat *Wybrane modele zdarzeń wielokrotnych w ocenie ryzyka powtórnej rejestracji w urzędzie pracy*, który zawierał wyniki analizy ryzyka kolejnej rejestracji w urzędzie pracy w zależności od wybranych cech osób bezrobotnych: płci, wieku, wykształcenia oraz stażu pracy. W badaniu porównano wyniki otrzymane dla modelu danych zliczanych Andersona-Gila, modeli warunkowych Prentince'a-Williamsa-Petersona: czasu całkowitego i luki czasowej oraz modelu Wei, Lin i Weissfelda. Sesję zamknął referat Wojciecha Gamrota (UE w Katowicach) *Skala Likerta i współczynnik regresji*. Autor omówił skutki postępowania polegającego na stosowaniu metod wnioskowania statystycznego przeznaczonych dla zmiennych ciągłych, a w szczególności metod opartych na założeniu wielowymiarowej normalności w przypadku modelu regresji liniowej, gdy skala Likerta stosowana jest w statystycznych badaniach kwestionariuszowych.

Sesji III A przewodniczył Tomasz Żądło; w jej trakcie wygłoszono trzy referaty. Dr Piotr Sulewski (Akademia Pomorska w Słupsku) w wystąpieniu *Rozpoznawanie rozkładów zamiast testowania zgodności* przedstawił koncepcję rozpoznawania rozkładów, zgodnie z którą stosowana jest reguła *k*-najbliższych sąsiadów – jako metoda konkurencyjna wobec przeprowadzania klasycznych testów zgodności wykorzystujących miarę rozbieżności. Referat Dominiki Polko-Zając (UE w Katowicach) *O permutacyjnych testach porównywania populacji wielowymiarowych* dotyczył permutacyjnej, jednoczesnej procedury identyfikacji różnic występujących między wektorami wartości przeciętnych oraz macierzami wariancji – kowariancji w dwóch badanych populacjach. Autorka przedstawiła również wyniki badania symulacyjnego w celu określenia rozmiaru i mocy tych testów. Następnie Krzysztof Szymoniak-Książek (UE w Katowicach) w prezentacji *Własności nieparametrycznych testów izotropii* omówił wyniki

badania symulacyjnego polegającego na wygenerowaniu ciągów realizacji pola losowego o zadanym rozkładzie teoretycznym, dla których testowano hipotezę zerową stanowiącą o izotropii. Autor wyznaczył empiryczne prawdopodobieństwa odrzuceń i porównał je z zakładanym z góry poziomem istotności.

Sesję III B, której przewodniczyła Grażyna Dehnel, otworzył referat dr Katarzyny Budny (UEK) *Wielowymiarowa nierówność Czebyszewa – oszacowania dla prawdopodobieństwa przyjmowania przez wektor losowy wartości z kuli euklidesowej*. Autorka zaprezentowała wybrane wielowymiarowe uogólnienia klasycznej nierówności Czebyszewa oraz oszacowania dla prawdopodobieństwa przyjmowania przez wektor losowy wartości z kuli euklidesowej, wyrażone za pomocą momentów wektora losowego opartych na definicji potęgi wektora. Wystąpienie dr. hab. Michała Bernardelliego (SGH) *Identyfikacja punktów zwrotnych szeregów czasowych z rynku kryptowalut* dotyczyło wykorzystania ścieżek Viterbiego do analizy jednowymiarowych szeregów cen z rynku kryptowalut za pomocą ukrytych modeli Markowa. Adam Juszczak (UŁ, Narodowy Bank Polski) w referacie *Zastosowanie danych scrapowanych w pomiarze inflacji* rozpatrzył pozytywne i negatywne aspekty wykorzystania web scrapingu przy obliczaniu wskaźnika dóbr i usług konsumpcyjnych (CPI).

Sesji IV A, prowadzonej w języku angielskim, przewodniczyła prof. dr hab. Grażyna Trzpiot (UE w Katowicach). Autorka pierwszego referatu dr hab. Aneta Ptak-Chmielewska (SGH) w prezentacji *Application of multidimensional classification to prediction of SME* dokonała porównania efektywności analizy dyskryminacyjnej z dyskryminacją wielowymiarową, taką jak metoda wektorów wspierających (*support vector machines*), z wykorzystaniem próby przedsiębiorstw MŚP. Dr hab. Jerzy Korzeniewski, prof. UŁ, w wystąpieniu *Determining semantic relatedness of concepts – modifications proposals* przedstawił metody badania podobieństwa semantycznego pojęć wraz z własną propozycją modyfikacji metody Leacocka i Chodorowa stosowanej dla pojęć, których wzajemne powiązania są bardzo słabe. Następnie dr Elżbieta Roszko-Wójtowicz (UŁ) i dr hab. Maria M. Grzelak, prof. UŁ, wygłosiły referat *Innovation activities and competitiveness of manufacturing divisions in Poland in the years 2009–2017*. Autorki przeprowadziły analizę pomiaru i oceny wpływu działalności innowacyjnej na konkurencyjność działów przetwórstwa przemysłowego, wykorzystując w tym celu modele panelowe, zarówno statystyczne z opóźnieniami, jak i dynamiczne, co pozwoliło na wnikliwe prześledzenie zmian w czasie w odniesieniu do rozpatrywanych zmiennych. Ostatnim głosem w sesji było wystąpienie dr. Łukasza Wawrowskiego (UEP) *Impact of dependent variable transformation on poverty rate estimates in poviats* dotyczące badania ubóstwa z wykorzystaniem danych o niskim stopniu agregacji przestrzennej. Prelegent zapoznał uczestników z wynikami estymacji stopy ubóstwa na poziomie powiatów z wykorzystaniem

danych z Europejskiego Badania Dochodów i Warunków Życia i Narodowego Spisu Powszechnego Ludności i Mieszkań oraz metod estymacji pośredniej bazujących na liniowych modelach mieszanych i symulacjach Monte Carlo.

W sesji IV B, której przewodniczył dr hab. Andrzej Dudek, prof. UEW, wygłoszono cztery referaty. Dr Agnieszka Stanimir (UEW) w wystąpieniu *Metody wielowymiarowej analizy statystycznej w analizie pytań wielokrotnego wyboru* przedstawiła wyniki zastosowania metod wielowymiarowej analizy statystycznej w sytuacjach, gdy wybór kategorii nie ma charakteru następstw oraz gdy z punktu widzenia badacza istotne jest poznanie sekwencji dokonywanych wyborów kategorii w ramach jednej zmiennej. Referat Łukasza Ziarki (UŁ) *O możliwości zastosowania analizy asocjacji do określenia wzorca zachowania się wykonawców w przetargach publicznych* dotyczył możliwości zastosowania analizy asocjacji (analizy koszykowej) do identyfikacji nielegalnych porozumień zawieranych pomiędzy wykonawcami ubiegającymi się o udzielenie zamówienia publicznego. Prelegent pokusił się również o próbę oceny zaproponowanego podejścia. Następnie dr Anna Denkowska i Stanisław Wanat (UEK) w prezentacji *Wzajemne powiązania a ryzyko systemowe w europejskim sektorze ubezpieczeniowym. Nowe wyniki bazujące na wykorzystaniu dynamicznych minimalnych drzew rozpinających* przedstawili wyniki badań wzajemnych powiązań europejskich instytucji ubezpieczeniowych oraz ich wkład w ryzyko systemowe związane z wykorzystaniem sieci korelacyjnych. Autorki ostatniego referatu *Zastosowanie modelu Zengi do opisu rozkładu dochodów ludności Polski dla grup społeczno-ekonomicznych* Alina Jędrzejczak i dr Kamila Trzcińska (UŁ) wykazały, że model Zengi bardzo dobrze opisuje rozkład dochodów ludności Polski dla grup społeczno-ekonomicznych.

Trzeciego dnia konferencji odbyły się dwie sesje plenarne. Pierwszej przewodniczył prof. dr hab. Andrzej Bąk (UEW). Dwa referaty wygłosili dr hab. Małgorzata Graczyk i prof. dr hab. Bronisław Ceranka (Uniwersytet Przyrodniczy w Poznaniu). Wystąpienie *Uwagi o wysoce D-efektywnych sprężynowych układach wagowych* autorzy poświęcili związanym z nowymi metodami konstrukcji sprężynowym układom wagowym o wysokiej D-efektywności w klasach, w których nie istnieje układ D-optymalny. W drugim wystąpieniu *Nowe wyniki dotyczące metod konstrukcji D-optymalnych chemicznych układów wagowych* omówili doświadczenie, w którym wyznaczane są nieznane miary p obiektów przy użyciu n operacji pomiarowych, zgodnie z modelem chemicznego układu wagowego, oraz podali warunki optymalności i serie parametrów układów D-optymalnych. Grażyna Trzpiot poświęciła swój referat *Seniorzy w miastach a miasta przyjazne seniorom – analiza dla wybranych miast Polski* zmianom i tempu procesu starzenia się wybranych miast w Polsce z wykorzystaniem modelu opisowego oraz odpornego podejścia taksonomicznego.

Grzegorz Kończak w wystąpieniu *Wielowymiarowe permutacyjne rozszerzenie testu McNemara* zreferował propozycję k-wymiarowego ($k > 2$) testu dla różnic pomiędzy prawdopodobieństwami sparowanych wektorów losowych o rozkładach dwupunktowych; propozycja ta stanowi wielowymiarowe rozszerzenie testu McNemara.

W finalnej sesji, pod przewodnictwem Czesława Domańskiego, Andrzej Bąk wystąpił z referatem *Metody imputacji brakujących danych z wykorzystaniem programu R na przykładzie Banku Danych Lokalnych*. Przedstawiał w nim wyniki badań związanych z uzupełnieniem brakujących danych, wykorzystujących metody zaproponowane w literaturze przedmiotu, jak również zestawy programu R. Wystąpienie Czesława Domańskiego *Uwagi o testach normalności opartych na charakterystykach procesów stochastycznych* zawierało rozważania o testach normalności, w szczególności tych, w których wykorzystywane są różnorodne charakterystyki procesów stochastycznych.

Czesław Domański, po zakończeniu i podsumowaniu wydarzenia, poinformował że kolejna konferencja Wielowymiarowa Analiza Statystyczna WAS odbędzie się w dniach 16–18 listopada 2020 r.

Aleksandra Baszczyńska (Uniwersytet Łódzki)

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 1, 62–65
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

WYDAWNICTWA GUS. GRUDZIEŃ 2019
PUBLICATIONS OF STATISTICS POLAND. DECEMBER 2019

W grudniowej ofercie wydawniczej warto zwrócić uwagę na publikację ***Co warto wiedzieć o inflacji?*** oraz opracowania cykliczne ***Kapitał ludzki w Polsce w latach 2014–2018***, ***Gospodarka morska w Polsce w latach 2017 i 2018***, ***Beneficjenci środowiskowej pomocy społecznej w 2018 r.*** i ***Wskaźniki zielonej gospodarki w Polsce 2019***.

Co warto wiedzieć o inflacji? to nowa pozycja wydawnicza, przystępnie i poglądowo przybliżająca tematykę pomiaru zmian cen. Skierowana jest do szerokiego grona odbiorców niezwiązanych zawodowo ze statystyką, lecz będących potencjalnymi obserwatorami zjawisk ekonomicznych i użytkownikami danych statystycznych.

Najistotniejsze informacje na temat wskaźników cen towarów i usług konsumpcyjnych przedstawiono w formie pytań i odpowiedzi, unikając opisu skomplikowanej metodologii oraz naukowej terminologii. Autorzy publikacji postawili sobie za cel wyjaśnienie, jak przebiega badanie i jak należy odczytywać jego wyniki. Z publikacji można dowiedzieć się m.in., w jaki sposób zbierane są dane do badania cen, co jest uwzględniane w tzw. koszyku inflacyjnym oraz gdzie można znaleźć wyniki badań i jak je poprawnie interpretować. Zamieszczono również wyjaśnienia dotyczące sposobów obliczania inflacji.

Opracowanie ukazało się w języku polskim.

Kapitał ludzki w Polsce w latach 2014–2018 jest najnowszym wydaniem ukazującej się co dwa lata publikacji, która informuje o stanie kapitału ludzkiego w naszym kraju. Dane w niej przedstawione pochodzą ze zbiorów statystyki publicznej oraz ze źródeł pozastatystycznych. Opracowanie zawiera również opis koncepcji badania oraz uwagi metodyczne.

Podstawowym celem autorów jest dostarczenie zestawu wskaźników umożliwiających prowadzenie samodzielnych badań i analiz w ujęciu lokalnym, regionalnym i krajowym w następujących dziedzinach: demografia, zdrowie, edukacja, rynek pracy, kultura, na-

uka, technologia i innowacje oraz ekonomiczne i społeczne uwarunkowania rozwoju kapitału ludzkiego. Informacje te pozwalają także na formułowanie programów rozwoju społeczno-gospodarczego kraju i jego regionów.

Opracowanie wydano w wersji polsko-angielskiej. Elektroniczną postać wydania uzupełniono o tablice w formacie Excel, zawierające dane szerszym zakresie przedmiotowym i szczegółowe przekroje terytorialne.

Gospodarka morska w Polsce w latach 2017 i 2018 zawiera podstawowe informacje z zakresu działalności podmiotów z tego sektora na tle wyników osiąganych na świecie. Dane dotyczą lat 2017 i 2018, ponieważ opracowanie poświęcone gospodarce morskiej ukazuje się z dwuletnią częstotliwością.

W kolejnych rozdziałach omówiono strukturę przestrzenno-funkcjonalną gospodarki, podmioty, pracujących i wynagrodzenia oraz inwestycje i środki trwałe. Dokonano również analizy gospodarki morskiej, uwzględniając m.in.: porty morskie, żeglugę morską i przybrzeżną, przemysł stoczniowy, gospodarkę rybną, szkolnictwo morskie i naukę oraz turystykę morską i przybrzeżną. Opracowanie wzbogacono o przegląd międzynarodowy.

Publikację wydano w wersji polsko-angielskiej.

Beneficjenci środowiskowej pomocy społecznej w 2018 r. to najnowsze wydanie publikacji ukazującej się co trzy lata. Stanowi podsumowanie pierwszej dekady prowadzenia stałej obserwacji beneficjentów środowiskowej pomocy społecznej i ich gospodarstw domowych. Odpowiada na potrzeby przedstawicieli administracji publicznej, środowiska naukowego i organizacji społecznych.

W kolejnych rozdziałach przedstawiono m.in. dane o zmianach wielkości populacji beneficjentów środowiskowej pomocy społecznej i skali korzystania z tej pomocy wraz z regionalnym zróżnicowaniem jej zasięgu, o uwarunkowaniu pomocy społecznej oraz o ubogich beneficjentach, czyli osobach ubogich korzystających z pomocy społecznej. Dokonano również analizy gospodarstw domowych beneficjentów środowiskowej pomocy społecznej, uwzględniając zróżnicowanie regionalne oraz zbiorowość ubogich beneficjentów żyjących samot-

nie. Ponadto podjęto problem trwałości korzystania z pomocy społecznej oraz po raz pierwszy scharakteryzowano beneficjentów środowiskowej pomocy społecznej przejmujących ubóstwo.

Publikację wydano w wersji polsko-angielskiej. Tablice w postaci elektronicznej przygotowano w formacie Excel.

Uzupełnieniem omawianego opracowania jest zeszyt metodologiczny *Beneficjenci środowiskowej pomocy społecznej*, zawierający kompleksowy opis metodologii badań.

Wskaźniki zielonej gospodarki w Polsce 2019 to pierwsza edycja publikacji analitycznej opracowanej na podstawie wyników badania z zakresu zielonej gospodarki (działalności ekonomicznej, której efektem jest poprawa jakości życia człowieka i jednocześnie zmniejszenie zagrożeń dla środowiska naturalnego), po włączeniu tego badania do Programu badań statystycznych statystyki publicznej. Przyjęte w nim rozwiązania metodologiczne bazują przede wszystkim na propozycjach Organizacji Współpracy Gospodarczej i Rozwoju (OECD).

Informacje statystyczne przedstawiono w czterech obszarach, w których monitorowany jest stan zielonej gospodarki: kapitału naturalnego, środowiskowej efektywności produkcji, środowiskowej jakości życia ludności oraz polityk gospodarczych i ich następstw. W celu przystępniejszego zobrazowania poruszanych zagadnień zaprezentowano wskaźniki kontekstowe, stanowiące tło i źródło podstawowych informacji o sytuacji społeczno-gospodarczej kraju. Proponowany zestaw miar do monitorowania stanu zielonej gospodarki, oprócz informacji pochodzących ze statystyki publicznej, obejmuje również dane różnych instytucji krajowych. Ponadto zaprezentowano porównania międzynarodowe na podstawie danych Eurostatu, OECD, Banku Światowego, Organizacji Narodów Zjednoczonych i Europejskiej Agencji Środowiska.

Publikację wydano w języku polskim.

W grudniu 2019 r. ukazały się ponadto:

- *Bezrobocie rejestrowane. I–III kwartał 2019 r.*;
- „Biuletyn statystyczny" nr 11/2019;
- *Ceny robót budowlano-montażowych i obiektów budowlanych (październik 2019 r.)*;
- *Ceny w gospodarce narodowej w latach 2014–2018*;
- *Działalność badawcza i rozwojowa w Polsce w 2018 r.*;

- *Ekonomiczne aspekty ochrony środowiska 2019*;
- *Gospodarka finansowa jednostek samorządu terytorialnego 2018*;
- *Koniunktura w przetwórstwie przemysłowym, budownictwie, handlu i usługach 2000–2019 (grudzień 2019)*;
- *Nakłady i wyniki przemysłu. I–III kwartał 2019 r.*;
- *Produkcja i handel zagraniczny produktami rolnymi w 2018 r.*;
- *Produkcja ważniejszych wyrobów przemysłowych w listopadzie 2019 r.*;
- *Produkt krajowy brutto – rachunki regionalne w latach 2015–2017*;
- *Rocznik Statystyczny Rzeczypospolitej Polskiej 2019*;
- *Rocznik Statystyki Międzynarodowej 2019*;
- *Społeczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2015–2019*;
- *Sytuacja społeczno-gospodarcza kraju w listopadzie 2019 r.*;
- *Sytuacja społeczno-gospodarcza województw nr 3/2019*;
- *Środki trwałe w gospodarce narodowej w 2018 r.*;
- „Wiadomości Statystyczne" nr 12/2019;
- *Zatrudnienie i wynagrodzenia w gospodarce narodowej w I–III kwartale 2019 r.*;
- *Zeszyt metodologiczny. Badanie produkcji przemysłowej*;
- *Zeszyt metodologiczny. Organizacje non-profit: stowarzyszenia, fundacje, samorząd gospodarczy i zawodowy oraz społeczne jednostki wyznaniowe*;
- *Zeszyt metodologiczny. Rewitalizacja w gminie*;
- *Zeszyt metodologiczny. Statystyka sportu*;
- *Zużycie paliw i nośników energii w 2018 roku.*

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

Justyna Gustyn (Główny Urząd Statystyczny)

Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 1, 66–68
Dyskusje. Recenzje. Informacje / Discussions. Reviews. Information

ZAPOWIEDZI WYDARZEŃ W STATYSTYCE
UPCOMING EVENTS IN OFFICIAL STATISTICS

Informacje o wybranych wydarzeniach międzynarodowych, w których wezmą udział przedstawiciele polskiej statystyki publicznej:
Selected international events which will be attended by the representatives of Polish official statistics:

51. Sesja Komisji Statystycznej ONZ, 3–6 marca 2020 r.
51st Session of the United Nations Statistical Commission, 3–6 March 2020

51. Sesja Komisji Statystycznej Organizacji Narodów Zjednoczonych odbędzie się w siedzibie ONZ w Nowym Jorku. Komisja Statystyczna składa się z 24 krajów – członków ONZ, którzy wybierani są przez Radę Gospodarczą i Społeczną ONZ na zasadach sprawiedliwego podziału geograficznego. Kadencja trwa cztery lata.

Robocza agenda sesji wskazuje następujące zagadnienia do omówienia oraz uzgodnienia: dane i wskaźniki do Agendy Zrównoważonego Rozwoju 2030, koordynacja programów statystycznych, przyszłość statystyki gospodarczej, rachunki narodowe, handel międzynarodowy i statystyka przedsiębiorstw, wskaźniki cen, Międzynarodowy Program Porównawczy (ICP), rachunki ekonomiczne środowiska, statystyka wsi i rolnictwa, statystyka demograficzna, rejestry cywilne i ruch naturalny ludności, statystyka zdrowia, statystyka płci, statystyka dotycząca uchodźców, statystyka rządów, pokoju i bezpieczeństwa, regionalny rozwój statystyki, podstawowe zasady statystyki publicznej, zarządzanie i modernizacja systemów statystycznych, statystyka informacji i technologii komunikacyjnych, big data, integracja informacji statystycznych i geoprzestrzennych, otwarte dane, metody robocze Komisji Statystycznej oraz Światowy Dzień Statystyki. Podczas sesji przekazane zostaną również informacje dotyczące: kontynuacji i rozwinięcia decyzji Rady Gospodarczej i Społecznej oraz Zgromadzenia Ogólnego, krótkoterminowej statystyki gospodarczej, statystyki cen, statystyki usług, statystyki środowiska, statystyki niepełnosprawności oraz statystyki rozwoju społecznego.

51. Sesja Komisji Statystycznej będzie się toczyć w językach: angielskim, arabskim, chińskim, francuskim, hiszpańskim i rosyjskim również dokumentacja zostanie zapisana w sześciu oficjalnych językach ONZ. Pozostałe spotkania zostaną poprowadzone w języku angielskim.

Wydarzenie będzie transmitowane na żywo pod adresem http://webtv.un.org/ (tu będzie również dostępne zarchiwizowane nagranie).

Więcej informacji: https://unstats.un.org/unsd/statcom/51st-session/.

Regionalne Forum Zrównoważonego Rozwoju dla Regionu EKG ONZ, 19–20 marca 2020 r.

Regional Forum on Sustainable Development for the UNECE Region, 19–20 March 2020

Czwarta sesja Regionalnego Forum Zrównoważonego Rozwoju odbędzie się w Genewie. Forum jest organizowane przez Europejską Komisję Gospodarczą (EKG) przy wsparciu regionalnego systemu Organizacji Narodów Zjednoczonych.

Regionalne Forum Zrównoważonego Rozwoju to dostępna dla wielu interesariuszy platforma rozwiązań prowadzących do osiągnięcia Celów Zrównoważonego Rozwoju. W ramach Forum obserwuje się i ocenia realizację zadań, które zostały określone w Agendzie na rzecz zrównoważonego rozwoju 2030, w regionie EKG. Forum stwarza przestrzeń dla wymiany rozwiązań politycznych i najlepszych praktyk oraz dyskusji nad wyzwaniami we wdrażaniu Celów Zrównoważonego Rozwoju, a także pomaga rozpoznać główne regionalne trendy. Wydarzenie jest otwarte dla wszystkich zainteresowanych realizacją Celów Zrównoważonego Rozwoju, włączając w to organizacje o zasięgu regionalnym i międzynarodowym oraz przedstawicieli społeczeństwa obywatelskiego, środowisk akademickich i sektora prywatnego.

W roboczym programie czwartej sesji Forum znalazły się: panel z zakresu polityki wysokiego szczebla poświęcony przyspieszeniu działania i osiągania zrównoważonego rozwoju w regionie EKG, cztery wydarzenia tematyczne, a także obrady przy okrągłym stole skupione wokół siedmiu zagadnień. Wydarzenia tematyczne będą dotyczyły kwestii: przyspieszenia postępu w realizacji Celów Zrównoważonego Rozwoju za pomocą dobrowolnych przeglądów narodowych, nowych podmiotów działających na rzecz zielonej zmiany i innowacji, zmiany przyzwyczajeń na rzecz zrównoważonego rozwoju i transformacji gospodarczej, jak również roli finansów i technologii w przyspieszeniu osiągania Celów Zrównoważonego Rozwoju. Z kolei do dyskusji podczas spotkań przy okrągłym stole zostały wybrane takie zagadnienia, jak: podnoszenie jakości życia poprzez objęcie ochroną zdrowia i edukacją wszystkich ludzi, przyspieszenie przekształcania systemów gospodarczych w zrównoważone gospodarki obiegu zamkniętego, zapewnienie zrównoważonego wyżywienia i zrównoważonej żywności, osiągnięcie neutralności węglowej w regionie EKG, drogi do zrównoważonego rozwoju miast w regionie EKG, rozszerzanie działań na rzecz zrównoważonego gospodarowania zasobami naturalnymi oraz wykorzystanie danych i statystyki do osiągnięcia Celów Zrównoważonego Rozwoju.

Więcej informacji: https://www.unece.org/rfsd.html.

Geospatial World Forum 2020, 7–9 kwietnia 2020 r.
Geospatial World Forum 2020, 7–9 April 2020

Geospatial World Forum to coroczne spotkanie liderów i specjalistów z obszaru informacji geoprzestrzennej – kreatorów polityk publicznych, reprezentantów krajowych instytucji zapewniających dostęp do informacji kartograficznej, przedsiębiorstw sektora prywatnego, organizacji multilateralnych i rozwojowych, instytucji naukowych i akademickich oraz użytkowników końcowych: rządu, przedsiębiorstw i instytucji świadczących usługi dla obywateli. Jak przewidują organizatorzy, w wydarzeniu weźmie udział około 1500 delegatów reprezentujących ponad 70 krajów i 500 organizacji, w tym 250 prelegentów. Organizatorem wydarzenia jest firma Geospatial Media and Communications.

Tematem 12. edycji Forum, która odbędzie się w Amsterdamie, jest rola informacji geoprzestrzennej w transformacji gospodarek w erze 5G – technologii mobilnej piątej generacji. Technologia, na której opiera się złożony system zintensyfikowanych powiązań pomiędzy ludźmi, organizacjami i maszynami, zaczyna zmieniać sposób organizacji przedsiębiorstw, ich współpracę oraz sposób pozyskiwania i rodzaj informacji, towarów i usług dostarczanych konsumentom. W tym kontekście lokalizacja geoprzestrzenna jest uznawana za jeden z filarów – obok sztucznej inteligencji, internetu rzeczy i big data – procesu przemian określanych mianem czwartej rewolucji przemysłowej. Nowe trendy w rozwoju technologii geoprzestrzennych opartym na technologii 5G wpłyną na wiele sektorów gospodarki i będą sprzyjać przełomowym innowacjom.

Bogaty program konferencji obejmuje panele plenarne, równoległe panele dyskusyjne skupione w dwóch blokach tematycznych: „przemysł" oraz „biznes, technologie i innowacje", a także seminaria. Ponadto podczas Forum zostaną przyznane nagrody Geospatial World Leadership Awards oraz Geospatial Excellence Awards. Wydarzeniu będzie towarzyszyć wystawa najnowszych technologii, produktów i rozwiązań z branży geoprzestrzennej. Przed konferencją, 5 kwietnia, rozpocznie się pięciodniowe szkolenie poświęcone wzrastającej roli infrastruktury dla informacji geoprzestrzenej w światowej gospodarce i społeczeństwach, organizowane wspólnie z Wydziałem Statystyki Organizacji Narodów Zjednoczonych. Z kolei 6 kwietnia odbędą się warsztaty dla dziennikarzy, podczas których uczestnicy będą mieli okazję zapoznać się z najnowszymi trendami i praktykami dziennikarskimi, w tym z możliwościami wykorzystania w pracy zdjęć satelitarnych, map i innych narzędzi cyfrowych.

Więcej informacji: https://geospatialworldforum.org.

Dodatkowe informacje znajdą Państwo na Portalu Naukowym GUS nauka.stat.gov.pl.
For more information go to Statistics Poland Research Portal research.stat.gov.pl.

DLA AUTORÓW
FOR THE AUTHORS
(for information go to 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: POL-index, CEJSH, BazEkon oraz AGRO.

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

1. Zgłaszanie artykułów

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

Artykuł powinien być utrzymany w formie bezosobowej i zawierać streszczenie, słowa kluczowe oraz kod/kody JEL. Tytuł, streszczenie i słowa kluczowe powinny być podane w języku polskim i angielskim. Jeżeli w pracy występują tablice, wykresy lub mapy, powinny być umieszczone w treści artykułu. W osobnym pliku należy podać dane do wykresów. **Prosimy o niestosowanie stylów i ograniczenie formatowania do wymogów redakcyjnych.** Więcej informacji – w podrozdziale *Wymogi redakcyjne* i następnych podrozdziałach.

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 oraz dane kontaktowe autora i afiliację zgłaszanego artykułu 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

Redakcja rozpoczyna postępowanie kwalifikujące artykuł do opublikowania po przedłożeniu przez autora oświadczenia o przeniesieniu praw majątkowych do artykułu.

Zgłoszony artykuł jest oceniany i opracowywany zgodnie ze schematem:

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 artykułu, która jest kierowana 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.
2. **Ocena recenzentów**, dokonywana przez specjalistów w danej dziedzinie. Artykuł oceniają dwaj recenzenci spoza jednostki naukowej, do której afiliowana jest zgłoszona praca; w przypadku artykułu 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 dostarczają redakcji zmodyfikowaną wersję pracy. Jeśli pojawi się różnica zdań dotycząca zasadności proponowanych zmian, autorzy są zobligowani do uzasadnienia swojego stanowiska.
3. **Ocena dopuszczająca do publikacji**, dokonywana przez Kolegium Redakcyjne (KR) 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. Kolegium Redakcyjne ocenia artykuł pod względem poprawności i spójności merytorycznej oraz zaleca autorowi wprowadzenie poprawek, jeśli są one konieczne, aby praca spełniała wymogi czasopisma. W przypadku podjęcia decyzji o niepublikowaniu artykułu autorowi przysługuje prawo do odwołania. W tym celu 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. W przypadku nieuwzględnienia danej uwagi autor jest zobligowany do uzasadnienia swojego stanowiska.
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, redakcję, 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 opracowaniu 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 publikacji.

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 poświadczeniem na piśmie 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 o tym niezwłocznie poinformować redakcję w celu dokonania korekty, wycofania tekstu lub zamieszczenia odpowiedniego sprostowania. W przypadku korekty artykułu już opublikowanego jego nowa wersja jest zamieszczana na stronie internetowej „WS" wraz ze stosownym wyjaśnieniem.

3.2. Odpowiedzialność redakcji

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 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 wszelkie 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 że jest ich oryginalnym dziełem, oraz określają swój wkład w opracowanie artykułu.
5. Podczas oceny wstępnej zgłoszony maszynopis jest weryfikowany przez redaktorów pod względem zgodności z celem i zakresem tematycznym czasopisma oraz spełniania wymogów redakcyjnych „WS", a także ewentualnych przejawów nierzetelności naukowej i możliwości wystąpienia konfliktu interesów.
6. Po ocenie wstępnej opracowania mające charakter naukowy przekazywane są do recenzji specjalistom z poszczególnych dziedzin. Redakcja jest odpowiedzialna za ustalenie spójnych kryteriów oceny artykułu oraz wymaga od recenzentów podpisania oświadczenia o przestrzeganiu zasad etyki recenzowania COPE (https://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 innym 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 tematycznych i merytorycznych. 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. W uzasadnionych przypadkach 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 „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 badań, 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 https://ws.stat.gov.pl/ForAuthors.

4.1. Struktura i zawartość artykułu

Wymagane elementy artykułu:

1. Tytuł, autor.
2. Streszczenie (objętość do 1200 znaków ze spacjami, forma bezosobowa). W przypadku artykułu opisującego badanie empiryczne powinno zawierać: cel badania, przedmiot, okres i metodę badania, źródła danych, najważniejsze wnioski z badania. W przypadku artykułów o innym charakterze należy podać co najmniej cel artykułu, 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.
3. Słowa kluczowe – najistotniejsze użyte w pracy pojęcia lub wyrażenia (nie mniej niż trzy). Słowa kluczowe powinny być zawarte w streszczeniu i/lub tytule.
4. 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).
5. Kod/kody z klasyfikacji Journal of Economic Literature (JEL).
6. W przypadku artykułu opisującego badanie empiryczne wymagane są następujące części:
   - wprowadzenie, zawierające: cel badania, uzasadnienie podjętego problemu badawczego i odniesienie do literatury przedmiotu, chyba że przegląd literatury stanowi odrębną część artykułu;
   - metoda badania, zawierająca: przedmiot i okres badania, źródła danych i zastosowane metody badawcze;
   - wyniki badania;
   - podsumowanie, które powinno być zwięzłe i odzwierciedlać istotę problemu badawczego przedstawionego w artykule, bez podawania danych liczbowych; wnioski powinny odnosić się do treści artykułu, a w szczególności do celu badań.

   Wszystkie części powinny być opatrzone numerami.
7. Bibliografia, zawierająca pełny wykaz prac i materiałów przywołanych w artykule, przygotowana zgodnie z wymogami czasopisma.

4.2. Przygotowanie artykułu

1. Tekst należy zapisać alfabetem łacińskim. Nazwy własne, tytuły itp. oryginalnie zapisane innym alfabetem powinny być poddane transliteracji.
2. Nie należy stosować stylów; formatowanie należy ograniczyć do wymogów redakcyjnych.
3. Objętość artykułu łącznie ze streszczeniem, słowami kluczowymi, bibliografią, tablicami, wykresami i innymi materiałami graficznymi nie powinna być mniejsza niż 10 stron ani przekraczać 20 stron maszynopisu.

4. Edytor tekstu: Microsoft Word, format *.doc lub *.docx.
5. Krój czcionki:
   - Arial – tytuł, autor, streszczenia, słowa kluczowe, kody JEL, tablice, zestawienia, wykresy, przypisy, śródtytuły;
   - Times New Roman – tekst główny, bibliografia.
6. Wielkość czcionki:
   - 14 pkt – tytuł, autor, tytuły rozdziałów;
   - 12 pkt – tekst główny, tytuły podrozdziałów;
   - 10 pkt – pozostałe elementy.
7. Marginesy – 2,5 cm z każdej strony.
8. Interlinia – 1,5 wiersza; tablice i przypisy – 1 wiersz; przed tytułami rozdziałów i podrozdziałów oraz po nich – pusty wiersz.
9. Wcięcie akapitowe – 0,4 cm; bibliografia – bez wcięcia, wysunięcie 0,4 cm.
10. Przy wyliczeniach należy posłużyć się listą punktowaną z punktorami w postaci kropek (wysunięcie 0,4 cm, wcięcie 0 cm); wiersze (oprócz ostatniego) zakończone średnikiem.
11. Strony ponumerowane automatycznie.
12. Tablice i elementy graficzne (wykresy, mapy, schematy) muszą być przywołane w tekście.
13. Wykresy, mapy i schematy powinny być zamieszczone w tekście głównym. Dane, na podstawie których opracowano wykresy, powinny być przekazane osobno w pliku programu Excel (lub innym edytowalnym w pakiecie Microsoft Office), ewentualnie wykresy powinny dawać możliwość odczytania z nich danych.
14. Tablice muszą być edytowalne. Nie należy stosować rastrów, cieniowania, pogrubiania czy też podwójnych linii itp.
15. Wskazówki dotyczące opracowywania map znajdują się w publikacji *Mapy statystyczne. Opracowanie i prezentacja danych*, dostępnej na stronie internetowej GUS https://stat.gov.pl/statystyka-regionalna/publikacje-regionalne/podreczniki-atlasy/podreczniki/mapy-statystyczne-opracowanie-i-prezentacja-danych,1,1.html.
16. Pod tablicami, wykresami, schematami i innymi elementami graficznymi należy podać źródło opracowania.
17. 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. $\mathbf{A}$, $\mathbf{M}$, $\mathbf{Y}(x)$, $\mathbf{M}_i$).
18. 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.
19. Stosowane są skróty: tablica – tabl., wykres – wykr.
20. Przypisy rzeczowe, słownikowe lub informacyjne należy umieszczać na dole strony. Przypisy bibliograficzne, zgodnie ze standardem APA (American Psychological Association), należy podawać w tekście głównym.
21. 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 wskazaniach tej samej pracy należy zastosować określenie „i współpracownicy" (w przypadku umieszczenia przywołania nazwisk w strukturze zdania) lub „i in." (gdy nazwiska autorów nie stanowią części struktury zdania). 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, natomiast pozostałych autorów należy zastąpić określeniem „i współpracownicy" (w przypadku umieszczenia przywołania nazwisk w strukturze zdania) lub „i in." (gdy nazwiska autorów nie stanowią części struktury zdania). W literaturze 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 jedynie pracę czytaną. Przykład zapisu: Jak wykazał Nowakowski (1990; za: Zieniecka, 2007)... Badania sugerują, że... (Nowakowski, 1990; za: Zieniecka, 2007).
7. Bibliografia powinna być zamieszczona na końcu opracowania. Prace należy zapisać alfabetycznie według nazwiska pierwszego autora. W przypadku dwóch lub więcej prac tego samego autora / tych samych autorów należy je uporządkować według roku publikacji. Jeśli kilka prac tego samego autora / tych samych autorów zostało opublikowanych w tym samym roku, należy uporządkować prace alfabetycznie według tytułu i wstawić litery a, b, c itd. po roku publikacji.

4.4. Przykłady opisu bibliograficznego

1. 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.
2. 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.
3. Książka: Nazwisko, X., Nazwisko 2, X. Y. (rok). *Tytuł książki*. Miejsce wydania: Wydawnictwo.
4. Książka napisana pod redakcją: Nazwisko, X. (red.). (rok). *Tytuł książki*. Miejsce wydania: Wydawnictwo.
5. 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.
6. 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 informacje 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: 21.03.2019).

Artykuł przygotowany w sposób niezgodny z powyższymi wskazówkami będzie odesłany z prośbą o dostosowanie jego formy do wymagań redakcji.

ZAKRES TEMATYCZNY DZIAŁÓW
THEMATIC SCOPE OF SECTIONS

(for information go to 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.