

Cena 12,00 zł  
(VAT 5%)

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

# WIADOMOŚCI STATYSTYCZNE

## THE POLISH STATISTICIAN

GŁÓWNY URZĄD  
STATYSTYCZNY  
STATISTICS  
POLAND

POLSKIE TOWARZYSTWO  
STATYSTYCZNE  
POLISH STATISTICAL  
ASSOCIATION

MIESIĘCZNIK  
MONTHLY JOURNAL  
LIPIEC 2019  
JULY

Numer **7** (698)  
Issue



Cena 12,00 zł  
(VAT 5%)

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

---

# WIADOMOŚCI STATYSTYCZNE

## THE POLISH STATISTICIAN

---

GŁÓWNY URZĄD  
STATYSTYCZNY  
STATISTICS  
POLAND

POLSKIE TOWARZYSTWO  
STATYSTYCZNE  
POLISH STATISTICAL  
ASSOCIATION

---

MIESIĘCZNIK  
MONTHLY JOURNAL  
ROK LXIV  
VOLUME 64  
LPIEC 2019  
JULY

Numer **7** (698)  
Issue

---

---

## RADA NAUKOWA / SCIENTIFIC COUNCIL

dr Dominik Rozkrut (przewodniczący/chairman) – Uniwersytet Szczeciński, prof. Anthony Arundel – University of Tasmania in Hobart, dr hab. Bożena Balcerzak-Paradowska, prof. IPiSS – Instytut Pracy i Spraw Socjalnych, prof. Eric Bartelsman – Vrije Universiteit Amsterdam, prof. dr hab. Czesław Domański – Uniwersytet Łódzki, dr hab. Elżbieta Gołata, prof. UEP – Uniwersytet Ekonomiczny w Poznaniu, prof. Semen Matkovskiy, PhD – Ivan Franko National University of Lviv, prof. dr hab. Włodzimierz Okrasa – Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie, prof. dr hab. Józef Oleński – Uczelnia Łazarskiego, prof. dr hab. Tomasz Panek – Szkoła Główna Handlowa w Warszawie, prof. Juan Manuel Rodríguez Poo – University of Cantabria, assoc. prof. ing. Iveta Stankovičová, PhD – Comenius University in Bratislava, prof. dr hab. Marek Walesiak – Uniwersytet Ekonomiczny we Wrocławiu, prof. dr hab. Józef Zegar – Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowy Instytut Badawczy

sekretarz/secretary: Paulina Kucharska-Singh

---

## KOLEGIUM REDAKCYJNE / EDITORIAL BOARD

prof. Tudorel Andrei – Bucharest Academy of Economic Studies, mgr Renata Bielak – Główny Urząd Statystyczny, dr Marek Cierpień-Wolan – Uniwersytet Rzeszowski, dr hab. Grażyna Dehnel, prof. UEP – Uniwersytet Ekonomiczny w Poznaniu, dr Jacek Kowalewski – Uniwersytet Ekonomiczny w Poznaniu, dr Jan Kubacki – Urząd Statystyczny w Łodzi, mgr Władysław Wiesław Łagodziński – Polskie Towarzystwo Statystyczne, dr Grażyna Marciniak – Główny Urząd Statystyczny, dr hab. Andrzej Młodak – Państwowa Wyższa Szkoła Zawodowa im. Prezydenta Stanisława Wojciechowskiego w Kaliszu, dr Stanisław Paradysz, dr hab. Mateusz Pipień – Uniwersytet Ekonomiczny w Krakowie, ing. Marek Rojiček, PhD – University of Economics Prague, assoc. prof. Anna Shostya, PhD – Pace University in New York, dr hab. Małgorzata Tarczyńska-Łuniewska, prof. US – Uniwersytet Szczeciński, dr Wioletta Wrzaszcz – Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – Państwowy Instytut Badawczy, dr inż. Agnieszka Zgierska – Główny Urząd Statystyczny

## ZESPÓŁ REDAKCYJNY / EDITORIAL STAFF

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

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

redaktorzy tematyczni / thematic editors: Jan Kubacki, Małgorzata Tarczyńska-Łuniewska, Agnieszka Zgierska

redaktor merytoryczny / substantive editor: Wioletta Wrzaszcz

redaktorzy językowi / language editors: Ewa Antoniak, Xawery Stańczyk, Małgorzata Zygmunt

sekretarz/secretary: Małgorzata Zygmunt

---

## ADRES REDAKCJI / EDITORIAL OFFICE ADDRESS

GUS, al. Niepodległości 208, 00-925 Warszawa, tel./phone +48 22 608 32 25, stat.gov.pl

e-mail: redakcja.ws@stat.gov.pl

---

**Wersja elektroniczna, stanowiąca wersję pierwotną czasopisma, jest dostępna na stat.gov.pl**  
An electronic edition of the journal is an original one. It is available at stat.gov.pl

© Copyright by Główny Urząd Statystyczny / Statistics Poland



Zakład Wydawnictw  
Statystycznych

Zakład Wydawnictw Statystycznych / Statistical Publishing Establishment

al. Niepodległości 208, 00-925 Warszawa, tel./phone +48 22 608 31 45

Informacje w sprawie nabywania czasopism / Information on purchasing of the journal

tel./phone +48 22 608 32 10, +48 22 608 38 10

Zbigniew Karpiński (redaktor techniczny / technical editor)

Ewa Krawczyńska (skład i tkanie / typesetting)

Wydział Korekty pod kierunkiem Bożeny Górczy / Proof-Reading Section supervised by Bożena Górczyca

Andrzej Kajkowski (wykresy/figures)

## Indeks 381306

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

---

## **SPIS TREŚCI**

### **CONTENTS**

#### **STUDIA METODOLOGICZNE**

#### **METHODOLOGICAL STUDIES**

Piotr Sulewski	
Porównanie generatorów liczb pseudolosowych .....	<b>5</b>
Comparison of normal random number generators	

#### **STATYSTYKA W PRAKTYCE**

#### **STATISTICS IN PRACTICE**

Wirginia Doryń	
Zróżnicowanie krajów pod względem udziału krajowej wartości dodanej w eksporcie	<b>32</b>
Differentiation of countries in terms of the share of domestic value added in exports	

#### **EDUKACJA STATYSTYCZNA**

#### **STATISTICAL EDUCATION**

Bogdan Stefanowicz	
Metafory informacyjne .....	<b>48</b>
Information metaphors	

#### **Z DZIEJÓW STATYSTYKI**

#### **FROM THE HISTORY OF STATISTICS**

Józef Pocięcha	
Demograficzny obraz ziem polskich w przededniu I wojny światowej .....	<b>56</b>
Demographic picture of the Polish lands before the outbreak of the First World War	

#### **IN MEMORIAM**

Zofia Mielecka-Kubień, Grażyna Trzpiot	
Profesor dr hab. Andrzej Stanisław Barczak, dr h.c. (1939–2019) .....	<b>70</b>
Andrzej Stanisław Barczak, PhD, DSc, Prof Tit, Dr HC (1939–2019)	

#### **INFORMACJE. RECENZJE. DYSKUSJE**

#### **INFORMATION. REVIEWS. DISCUSSIONS**

Artur Mikulec	
Konferencja naukowa <i>230 lat Statystyki Publicznej w Polsce</i> .....	<b>75</b>
Scientific conference <i>230 Years of Public Statistics in Poland</i>	
Justyna Gustyn	
Wydawnictwa GUS. Czerwiec 2019 .....	<b>84</b>
Publications of Statistics Poland. June 2019	
Do Autorów .....	<b>87</b>
To the Authors	



## Porównanie generatorów liczb pseudolosowych

Piotr Sulewski<sup>a</sup> 

**Streszczenie.** Losowanie prób w badaniach statystycznych i w obliczeniach numerycznych, jak również symulacyjne badanie modeli probabilistycznych właściwie we wszystkich dziedzinach wiedzy wymagają wyposażenia komputera w generatory liczb pseudolosowych. Głównym celem pracy jest porównanie generatorów liczb pseudolosowych normalnych na podstawie ich analizy dokonanej za pomocą różnego rodzaju kryteriów. Zbadano właściwości 12 generatorów liczb pseudolosowych o rozkładzie normalnym. Zaproponowano rozszerzenie rodziny generatorów o dwa tzw. generatory aplikacyjne oraz przyjęcie nowego podejścia do sprawdzania jakości generatorów. Przedstawiono narzędzie przygotowane w języku C++ oraz w języku Visual Basic for Application (VBA) do prowadzenia samodzielnych badań z użyciem generatorów. Symulacje Monte Carlo przeprowadzono w języku C++, a obliczenia wykonano w edytorze VBA przy użyciu arkusza kalkulacyjnego Microsoft Excel 2016. Analiza uzyskanych wyników wskazuje, że najlepsze właściwości mają generatory: MP Monty Pythona, R, Biegun oraz Ziggurat. Najmniej użyteczne okazują się generatory: BM Boxa-Mullera, Wallace'a, Iloraz oraz Excel.

**Słowa kluczowe:** rozkład normalny, generator liczb pseudolosowych, symulacje Monte Carlo

## Comparison of normal random number generators

**Summary.** The sampling in statistical surveys and numerical calculations as well as simulation testing of probabilistic models in virtually all fields of knowledge require a computer endowed with pseudorandom numbers generators. The main goal of the study is to compare the normal random number generators using various criteria. The properties of 12 random number generators for a normal distribution were investigated. Then, the family of generators was extended by two so-called application generators and a new approach for checking the quality of generators was adopted. A ready-made tool prepared in C++ and in Visual Basic for Application (VBA) for conducting self-contained research using generators was presented. All Monte Carlo simulations were carried out in C++, while the calculations were performed in the VBA editor using the Microsoft Excel 2016 spreadsheet. The analysis of the obtained results shows that the generators with best properties are: MP Monty Python, R, Biegun and Ziggurat. The worst generators, are: BM Box-Muller, Wallace, Iloraz and Excel.

**Keywords:** normal distribution, pseudo-random number generator, Monte Carlo simulation

**JEL:** C12, C15

---

<sup>a</sup> Akademia Pomorska w Słupsku, Wydział Matematyczno-Przyrodniczy.

Losowanie prób w kontekście badań statystycznych (badania reprezentacyjne, symulacyjne badanie estymatorów, testów i statystycznych reguł decyzyjnych) oraz w kontekście obliczeń numerycznych (metody Monte Carlo dla całek, równań z operatorami liniowymi oraz dla zadań optymalizacji), jak również symulacyjne badanie modeli probabilistycznych, szczególnie podczas analizowania zjawisk ekonomicznych, lecz także w technice, naukach przyrodniczych i właściwie we wszystkich pozostałych dziedzinach wiedzy, wymagają wyposażenia komputera w odpowiednie narzędzia (Wieczorkowski i Zieliński, 1997). Takimi narzędziami są generatory liczb pseudolosowych.

W ekonomii generatory liczb pseudolosowych znajdują zastosowanie w symulacjach i modelowaniu matematycznym (Ohirko, Zaniewski i Ogirko, 2016). Dotyczą one takich zagadnień, jak: ekonomia gospodarstwa domowego, obliczanie kosztów i budżetów, obliczanie podatków dochodowych, wyliczanie stóp procentowych w kredytach udzielanych przez banki, obliczenia procentowe z uwzględnieniem operacji bankowych, inwestycje i planowanie remontów, obliczanie wynagrodzeń pracowniczych, koszty finansowe zakładu pracy, księgi podatkowe, bilanse księgowy, ustalanie wyniku finansowego metodą statystyczną, wycena aktywów i kapitałów z uwzględnieniem amortyzacji danego środka oraz obliczanie inwestycji długoterminowych. Modelowanie i symulacje tych zjawisk mają pozytywny wpływ na podejmowane decyzje.

Liczby pseudolosowe o rozkładzie normalnym są powszechnie stosowane m.in. w symulacjach Monte Carlo. W ciągu ostatnich 60 lat opublikowano kilkadziesiąt algorytmów generujących takie liczby. O ich użyteczności przesądza jakość. Pierwszą publikacją na ten temat była praca Mullera (1959). Później tematykę tę podejmowali m.in.: Kronmal (1964), Ahrens i Dieter (1972), Chen i Burford (1981), Schollmeyer i Tranter (1991) oraz Molle, Hinich i Morrice (1992). Porównania pseudolosowych liczb o rozkładzie normalnym dokonali Kabal (2000) oraz Thomas, Luk, Leong i Villasenor (2007). W pracy Thomasa i współpracowników (2007) zbadano zarówno jakość generatorów, jak i ich szybkość. W badaniach tych skorzystano z testu zgodności *chi-kwadrat*, z testów high-sigma multiple, konwersji liczb do rozkładu jednostajnego oraz testowania korelacji. Ze względu na szybkość działania wyróżniono generator Wallace, a ze względu na jakość – generator Ziggurat. Algorytmy zapisane w pseudokodzie rozpatrywano oddzielnie dla głównej części rozkładu Gaussa ( $|x| \leq 3$ ) i dla tzw. ogonów tego rozkładu ( $|x| > 3$ ).

W dobie dynamicznie rozwijającego się rynku komputerowego, w tym błyskawicznie rosnącej wydajności procesorów, badanie szybkości działania generatorów jest rzeczą drugorzędną. W związku z powyższym w badaniu omawianym w niniejszej pracy ograniczono się do porównania jakości generatorów.

Analizie statystycznej poddano 12 generatorów liczb pseudolosowych o rozkładzie normalnym oraz dwa tzw. generatory aplikacyjne. Wszystkie, z wyjątkiem

generatorów aplikacyjnych, zostały zaimplementowane w języku C++, a większość z nich także w języku VBA (Visual Basic for Application) wbudowanym w Microsoft Excel 2016. Dzięki temu czytelnik może samodzielnie przeanalizować kody źródłowe algorytmów i prowadzić badania, wykorzystując dostarczone generatory liczb pseudolosowych o rozkładzie normalnym<sup>1</sup>.

W testowaniu jakości generatorów posłużono się narzędziami statystycznymi niewykorzystanymi w pracy Thomasa i współpracowników (2007), a mianowicie: testem Lilleforsa, testem Andersona-Darlinga, testami trendu i sezonowości, a także dwoma nowymi testami, z których jeden wykorzystuje moduł z różnicy dystrybuant teoretycznej i empirycznej, a drugi – statystyki pozycyjne.

Głównym celem artykułu jest porównanie generatorów liczb pseudolosowych normalnych na podstawie ich analizy przeprowadzonej przy użyciu różnego rodzaju kryteriów. Zbadano właściwości 12 generatorów.

## ROZKŁAD NORMALNY

Rozkład Gaussa z wartością oczekiwaną  $m = 0$  i odchyleniem standardowym  $s = 1$ , zwany także rozkładem normalnym standaryzowanym  $N(0,1)$ , ma funkcję gęstości prawdopodobieństwa w postaci:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (1)$$

Odwroćenie funkcji gęstości (1) w wyniku nieskomplikowanych przekształceń daje równość:

$$\varphi^{-1}(y) = \sqrt{-\ln(2\pi y^2)} \quad (2)$$

Dystrybuanta teoretyczna rozkładu  $N(0,1)$  dana jest wzorem:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt = \frac{1}{2} \left[ 1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right] \quad (3)$$

gdzie  $\operatorname{erf}$  jest funkcją błędu Gaussa. Wartość dystrybuanty (3) można wyznaczyć za pomocą wbudowanej funkcji Excela `ROZKŁAD.NORMALNY(x;0;1;1)`. Funkcja ta jest akceptowalna do stosowania w rozbudowanej analizie statystycznej (Mélard, 2014). Wartość dystrybuanty (3) można obliczyć także za pomocą kodu języka VBA w postaci `APPLICATION.NORMDIST(x,0,1,1)`.

<sup>1</sup> Stosowne pliki są umieszczone pod adresem <https://sulewski.apsl.edu.pl/index.php/publikacje>.



Niech zmienna losowa  $X$  ma rozkład  $N(0,1)$ , a  $\varphi(x)$  będzie funkcją gęstości daną wzorem (1). Wartość oczekiwana zmiennej losowej  $X$  wynosi 0, ponieważż:

$$m = \int_{-\infty}^{\infty} x \varphi(x) dx = 0 \quad (4)$$

Momenty centralne zmiennej losowej  $X$  mają w tej sytuacji postać:

$$\mu_k = \int_{-\infty}^{\infty} (x-m)^k \varphi(x) dx = \int_{-\infty}^{\infty} x^k \varphi(x) dx \quad (5)$$

Jeżeli zmienna losowa  $X$  ma rozkład  $N(0,1)$ , to zmienna losowa  $Y = Xs + m$  ma rozkład  $N(m, s)$ .

Charakterystyki rozkładu  $N(0,1)$  przyjmują następujące wartości teoretyczne (Cramér, 1958):

- wartość oczekiwana  $m = 0$ ,
- odchylenie standardowe  $s = \sqrt{\mu_2} = 1$ ,
- momenty centralne:  $\mu_3 = 0, \mu_4 = 3$ .

Niech  $x_i (i = 1, 2, \dots, n)$  będzie  $n$ -elementową próbą o rozkładzie normalnym. Wartość średnią z próby definiuje się jako:

$$\alpha_1 = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

natomiast momenty centralne z próby to:

$$M_k = \frac{1}{n} \sum_{i=1}^n (x_i - \alpha_1)^k \quad (7)$$

Estymatory nieobciążone momentów centralnych mają postać (Cramér, 1958):

$$M_2^n = \frac{n}{n-1} M_2 \quad M_3^n = \frac{n^2}{(n-1)(n-2)} M_3 \quad (8)$$

$$M_4^n = \frac{n(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)} M_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} M_2$$

## GENERATORY LICZB PSEUDOLOSOWYCH O ROZKŁADZIE JEDNOSTAJNYM

W generacji liczb pseudolosowych o rozkładzie normalnym ważne są nie tylko algorytmy opisujące poszczególne jej metody, lecz przede wszystkim algorytmy, które charakteryzują liczby pseudolosowe o rozkładzie jednostajnym (inaczej równomiernym)  $U(0,1)$  wykorzystywane przez te generatory.

Testowanie liczb losowych o rozkładzie jednostajnym omówił L'Ecuyer (1992, 2001). Zestawy testów i procedur można znaleźć także w pracy Rukhina i współpracowników (2001). Testy Dieharda zostały opracowane przez Marsaglię (1997). Innym kompleksowym zestawem jest TestU01 (L'Ecuyer, Simard, 2005), który zapewnia szeroki wybór testów, takich jak np. test Crusha (L'Ecuyer, 2001).

W niniejszej pracy skorzystano z bardzo popularnego algorytmu Mersenne Twister autorstwa japońskich naukowców Matsumoto i Nishimura (1998), generującego liczby pseudolosowe o rozkładzie jednostajnym o bardzo dużym okresie  $2^{19937} - 1$ . Algorytm ten jest wbudowany w takie środowiska obliczeniowe, jak Excel, w tym VBA (od wersji 2010), Calc (od wersji Apache OpenOffice 4.0), Scilab i język R. Na potrzeby tego artykułu algorytm Mersenne Twister został zapisany także w języku C++, ponieważ z procedury Mersenne Twister korzysta każdy z omawianych tu generatorów. Zapewniło to możliwość porównywania ze sobą generatorów liczb o rozkładzie normalnym niezależnie od tego, czy zostały one zaprogramowane w języku C++, czy w języku VBA (Excel 2016), czy też pochodzą z programu Scilab. Opierając się na sugestiach zawartych w pracy Mélarida (2014), do generacji liczb losowych o rozkładzie jednostajnym w języku VBA wykorzystano formułę EVALUATE ("RAND()") zamiast formuły RND.

## WŁAŚCIWOŚCI GENERATORÓW LICZB PSEUDOLOSOWYCH O ROZKŁADZIE NORMALNYM

Dwanaście najważniejszych zdaniem autora generatorów liczb pseudolosowych o rozkładzie normalnym zostało zdefiniowanych i zapisanych w formie algorytmów. Algorytmy te dotyczą gęstości rozkładu Gaussa w całej swojej dziedzinie wraz z tzw. ogonami rozkładu.

### **Generator CdfInv odwracania dystrybuanty**

Niech  $U_1$  będzie liczbą pseudolosową  $U(0,1)$ . Wówczas liczba pseudolosowa  $X$  z  $N(0,1)$  ma postać  $X = \Phi^{-1}(U_1)$ , gdzie  $\Phi^{-1}$  jest funkcją odwrotną do dystrybuanty  $N(0,1)$  daną wzorem (2). Jedną z pierwszych metod aproksymacji funkcji  $\Phi^{-1}$  zaproponował Muller (1958). Zagadnieniem tym zajmowali się także:

Gebhardt (1964), Wetherill (1965), Wichura (1988), McCollum, Lancaster, Boul-din i Peterson (2003) oraz Chen, Moon i Bazargan (2004). W artykule skorzy-stano z wbudowanej funkcji użytkownika APPLICATION.NORMINV ( $U_1, 0, 1$ ) z edytora VBA. Analiza pracy Méléarda (2014) wskazuje, że z funkcji tej można korzystać w badaniach statystycznych.

### Generator Boxa-Mullera (BM)

Niech  $U_1, U_2$  będą liczbami pseudolosowymi  $U(0,1)$ . Metoda Boxa-Mullera (Box i Muller, 1958) generująca parę niezależnych liczb pseudolosowych ( $Y_1, Y_2$ ) z  $N(m, s)$  jest opisana za pomocą algorytmu:

1.  $a = \sqrt{-2\ln(U_1)}, b = 2\pi U_2$
2.  $X_1 = a\sin(b), X_2 = a\cos(b)$
3.  $Y_1 = sX_1 + m, Y_2 = sX_2 + m$

W testach można np. skorzystać tylko z jednej wartości  $Y_1$  lub  $Y_2$ . Innym roz-wiązaniem jest zwracanie przez algorytm wartości  $Y_1$ , a wartość  $Y_2$  jest buforo-wana i zwracana podczas ponownego uruchomienia algorytmu.

### Generator Tuzin

Niech  $U_1, U_2, \dots, U_{12}$  będą liczbami pseudolosowymi  $U(0,1)$ . Metoda tuzina generująca liczbę pseudolosową  $Y$  z  $N(m, s)$  jest opisana za pomocą algorytmu:

1.  $S = \sum_{i=1}^{12} U_i$
2.  $X = S - 6$
3.  $Y = sX + m$

### Generator Biegun

Niech  $U_1, U_2$  będą liczbami pseudolosowymi  $U(0,1)$ . Metoda biegunowa (Bell, 1968; Knop, 1969) generująca parę niezależnych liczb pseudolosowych ( $Y_1, Y_2$ ) z  $N(m, s)$  jest opisana za pomocą algorytmu:

1.  $a = -1 + 2U_1, b = -1 + 2U_2$
2.  $d = a^2 + b^2$
3. Jeżeli  $d \geq 1$ , to przejdź do pkt 1
4.  $e = \sqrt{-2\ln(d)/d}$
5.  $X_1 = ae, X_2 = be$
6.  $Y_1 = sX_1 + m, Y_2 = sX_2 + m$

W testach można skorzystać np. tylko z jednej wartości  $Y_1$  lub  $Y_2$ . Innym rozwiązaniem jest zwracanie przez algorytm wartości  $Y_1$ , podczas gdy wartość  $Y_2$  jest buforowana i zwracana przy ponownym uruchomieniu algorytmu.

### Generator Iloraz

Niech  $U_1, U_2$  będą liczbami pseudolosowymi  $U(0,1)$ . Metoda ilorazowa (Kinderman i Monahan, 1977) generująca liczbę pseudolosową  $Y$  z  $N(m, s)$  jest opisana za pomocą algorytmu (Wieczorkowski i Zieliński, 1997):

1.  $u = U_1, e = \exp(1), v = -\sqrt{2/e} + 2\sqrt{2/e}U_2$
2.  $X = v/u$
3. Jeżeli  $X^2 \leq 2(3 - u(4 + u))$ , to przejdź do pkt 6
4. Jeżeli  $X^2 \leq 2/u - 2u$  i  $X^2 \leq -4\ln(u)$ , to przejdź do pkt 6
5. Przejdź do pkt 1
6.  $Y = sX + m$

### Generator Levy

Niech  $U_1, U_2$  będą liczbami pseudolosowymi  $U(0,1)$ . Metoda Levy (Leva, 1992) generująca liczbę pseudolosową  $Y$  z  $N(m, s)$  jest opisana za pomocą poniższego algorytmu. W pkt 7 przedstawiono drobną korektę autorską, bez uwzględnienia której zbyt liczne wartości  $X = 0$  obniżają jakość generatora.

1.  $e = \exp(1), a = 0,196, b = 0,25472, ss = 0,449871, t = -0,386595$
2.  $u = U_1, v = -\sqrt{2/e} + 2U_2\sqrt{2/e}$
3.  $x = u - ss, y = |v| - t, Q = x^2 + y(ay - bx)$
4. Jeżeli  $Q < 0,27597$ , to przejdź do pkt 7
5. Jeżeli  $Q > 0,27846$ , to przejdź do pkt 2
6. Jeżeli  $v^2 > -4u^2$ , to przejdź do pkt 2
7. Jeżeli  $v/u = 0$ , to przejdź do pkt 2
8.  $X = v/u$
9.  $Y = sX + m$

### Generator Ahrensa-Dietera (AD)

Niech  $U_1, U_2, U_3$  będą liczbami pseudolosowymi  $U(0,1)$ . Metoda Ahrensa-Dietera (Ahrens i Dieter, 1988) generująca parę niezależnych liczb pseudolosowych  $(Y_1, Y_2)$  z  $N(m, s)$  jest opisana za pomocą algorytmu:

1. Jeżeli  $U_1 < 0,5$ , to  $a = 1$ . Przejdź do pkt 3
2.  $a = -1$

3.  $b = -\ln(U_2)$
4.  $c = \text{tg}[\pi(U_3 - 0,5)]$
5.  $d = \sqrt{2b/(1 + c^2)}$
6.  $X_1 = ad, X_2 = cd$
7.  $Y_1 = sX_1 + m, Y_2 = sX_2 + m$

W testach można skorzystać np. tylko z jednej wartości  $Y_1$  lub  $Y_2$ . Inne rozwiązanie polega na zwracaniu przez algorytm wartości  $Y_1$  i buforowaniu wartości  $Y_2$ , która jest zwracana podczas ponownego uruchomienia algorytmu.

### Generator Marsaglii-Braya (MB)

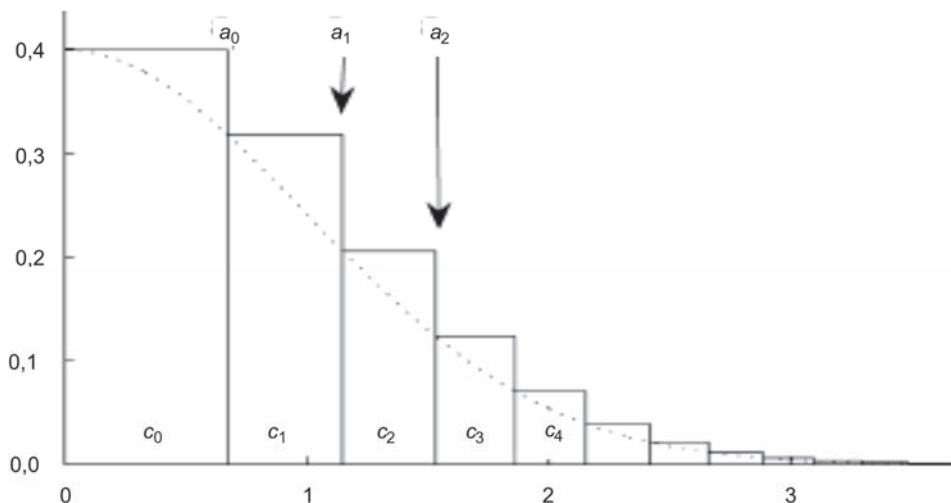
Niech  $U_1, U_2, \dots, U_{10}$  będą liczbami pseudolosowymi  $U(0,1)$ . Metoda Marsaglii-Braya (Marsaglia i Bray, 1964) generująca liczbę pseudolosową  $Y$  z  $N(m, s)$  jest opisana za pomocą algorytmu, który wywołuje funkcję pomocniczą w postaci:

$$g_3(x, a, b, c, d) = \begin{cases} a \exp(-0,5x^2) - b(3 - x^2) - c(1,5 - |x|) & \text{dla } |x| \leq 1 \\ a \exp(-0,5x^2) - d(3 - |x|)^2 - c(1,5 - |x|) & \text{dla } 1 < |x| \leq 1,5 \\ a \exp(-0,5x^2) - d(3 - |x|)^2 & \text{dla } 1,5 < |x| \leq 3 \\ 0 & \text{dla } 3 < |x| \end{cases}$$

1.  $a = 17,49731196, b = 4,73570326, c = 2,15787533, d = 2,36785163$
2.  $a_1 = 0,8638, a_2 = 0,1107, a_3 = 0,0228002039$
3.  $a_4 = 1 - a_1 - a_2 - a_3$
4.  $u = U_1$
5. Jeżeli  $u < a_1$ , to  $X = 2(U_2 + U_3 + U_4 - 1,5)$ . Przejdź do pkt 14
6. Jeżeli  $u < a_1 + a_2$ , to  $X = 1,5(U_5 + U_6 - 1)$ . Przejdź do pkt 14
7. Jeżeli  $u < a_1 + a_2 + a_3$ , to  $x = 6U_7 - 3, y = 0,358U_8$
8. Jeżeli  $y \geq g_3(x, a, b, c, d)$ , to przejdź do pkt 7
9.  $X = x$ , przejdź do pkt 14
10.  $f = -1 + 2U_9, g = U_{10}$
11.  $x = -\ln(|f|)/3, y = -\ln(g)$
12. Jeżeli  $2y \leq x^2$ , to przejdź do pkt 10
13. Jeżeli  $f > 0$ , to  $X = x + 3$ . W sytuacji przeciwnej  $X = -x - 3$
14.  $Y = sX + m$

### Generator Grand

Niech  $U_1, U_2, \dots, U_6$  będą liczbami pseudolosowymi  $U(0,1)$ . Wykres 1 przedstawia ideę generatora Grand, tzn. podział gęstości normalnej na oddzielne zakresy. Zakres  $c_i$  jest wybierany z prawdopodobieństwem  $2^{-i-1}$ .

WYKR. 1. PODZIAŁ GĘSTOŚCI NORMALNEJ NA ODDZIELNE ZAKRESY  $c_i$ 

U w a g a.  $i = 0, 1, 2 \dots$

Ź r ó d ł o: Thomas i współpracownicy (2007).

Metoda Grand (Brent, 1974) generująca liczbę pseudolosową  $Y$  z  $N(m, s)$  jest opisana za pomocą poniższego algorytmu, w którym  $t_i = \Phi^{-1}(1 - 2^{-i-1})$ ,  $i = 0, 1, 2, \dots$

1.  $a = U_1$ ,  $i = 0$
2. Dopóki  $a < 0,5$  wykonuj kroki 2.1 i 2.2
  - 2.1.  $a = 2a$
  - 2.2.  $i = i + 1$
3.  $u = (t_{i+1} - t_i)U_2$ ,  $v = u(0,5u + t_i)$
4. Dopóki  $v \geq U_6$  wykonuj kroki 4.1–4.3
  - 4.1. Jeżeli  $v \geq U_3$ , to przejdź do pkt 4.3
  - 4.2. Jeżeli  $U_4 < 0,5$ , to  $X = t_i + u$ , w przeciwnym razie  $X = -t_i - u$ . Przejdź do pkt 5
  - 4.3.  $v = U_5$
5.  $Y = sX + m$

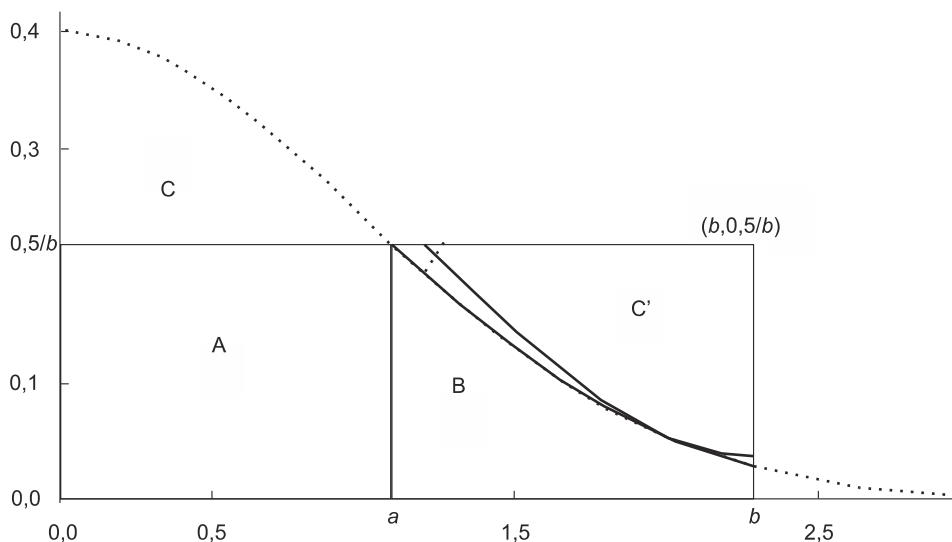
W środowisku matematycznym SciLab generator Grand został zaimplementowany dla różnych rozkładów, w tym dla rozkładu normalnego. Wygenerowanie na potrzeby symulacji Monte Carlo tablicy  $n \times u$  liczb pseudolosowych o rozkładzie normalnym  $N(m, s)$  i wyeksportowanie tej tablicy do pliku data.csv zostało

wykonane w środowisku SciLab za pomocą polecenia: `csvWrite(grand(n, u, "nor", m, s), fullfile("data.csv"), ",", ",")`.

### Generator Monty Pythona (MP)

Niech  $U_1, U_2, \dots, U_5$  będą liczbami pseudolosowymi  $U(0,1)$ . Wykres 2 przedstawia ideę generatora Monty Pythona, tzn. spakowanie rozkładu Gaussa do obszaru prostokątnego. Jeżeli punkt losowy  $(x, y)$  znajduje się w obszarze  $C'$ , to należy go przesunąć do obszaru  $C$  (zob. pkt 6 poniższego algorytmu). Funkcja *int* oznacza część całkowitą z liczby, funkcja gęstości  $\varphi(x)$  dana jest wzorem (1), a wyrażenie  $\varphi^{-1}(x)$  dane jest wzorem (2).

WYKR. 2. SPAKOWANIE ROZKŁADU GAUSSA DO OBSZARU PROSTOKĄTNEGO



Źródło: jak przy wyk. 1.

Metoda Monty Pythona (Marsaglia i Tsang, 1998) generująca liczbę pseudolosową  $Y$  z  $N(m, s)$  została opisana za pomocą algorytmu:

1.  $b = \sqrt{2\pi} = 2,5066, a = \varphi^{-1}(1/2b) = 1,1774$
2.  $c = 2\text{int}(2U_1) - 1, x = bU_2$
3. Jeżeli  $x < a$ , to  $X = cx$ . Przejdź do pkt 12
4.  $y = U_3/(2b)$
5. Jeżeli  $y < \varphi(x)$ , to  $X = cx$ . Przejdź do pkt 12
6.  $(x, y)^{C'} \rightarrow (x, y)^C$
7. Jeżeli  $y < \varphi(x)$ , to  $X = cx$ . Przejdź do pkt 12

8.  $d = -1 + 2U_4$ ,  $f = U_5$
9.  $x = -\ln(|d|)/3$ ,  $y = -\ln(f)$
10. Jeżeli  $2y \leq x^2$ , to przejdź do pkt 8
11. Jeżeli  $d > 0$ , to  $X = x + 3$ . W sytuacji przeciwnej  $X = -x - 3$
12.  $Y = sX + m$

### Generator Wallace'a

Generator Wallace'a (2005) jest najszybszym generatorem liczb pseudolosowych o rozkładzie normalnym  $N(0,1)$ . Ze względu na programistyczny i bardzo skomplikowany charakter algorytmu go opisującego w niniejszej pracy ograniczono się do implementacji komputerowej generatora. Szczegółowy algorytm został przedstawiony w pracy Thomasa i współpracowników (2007).

### Generator Ziggurat

Niech  $U_1, U_2, \dots, U_5$  będą liczbami pseudolosowymi  $U(0,1)$ . Metoda Ziggurat (Marsaglia i Tsang, 2000) generująca liczbę pseudolosową  $Y$  z  $N(m, s)$  jest opisana za pomocą poniższego algorytmu. Funkcja *int* oznacza część całkowitą z liczby, natomiast wyrażenia  $\varphi(x)$  oraz  $\varphi^{-1}(x)$  dane są odpowiednio wzorami (1) i (2).

1.  $n = 255$ ,  $v = 0,00492867323399$ ,  $r = 3,6541528853610088$
2.  $x_{255} = r$
3. Dla  $i = 254, 253, \dots, 1$ 
  - 3.1.  $x_i = \varphi^{-1}\left(\frac{v}{x_{i+1}} + \varphi(x_{i+1})\right)$
4.  $i = 1 + \text{int}(nU_1)$
5.  $x = x_i U_2$
6. Jeżeli  $|x| < x_{i-1}$ , to  $X = x$ . Przejdź do pkt 10
7. Jeżeli  $i \neq n$ , to  $y = (\varphi(x_{i-1}) - \varphi(x_i))U_3$
8. Jeżeli  $y < (\varphi(x) - \varphi(x_i))$ , to  $X = x$ . Przejdź do pkt 10
9. Jeżeli  $|x| \geq x_{i-1}$ , to:
  - 9.1.  $d = -1 + 2U_4$ ,  $f = U_5$
  - 9.2.  $x = -\ln(|d|)/3$ ,  $y = -\ln(f)$
  - 9.3. Jeżeli  $2y \leq x^2$ , to przejdź do pkt 9.1
  - 9.4. Jeżeli  $d > 0$ , to  $X = x + 3$ . W sytuacji przeciwnej  $X = -x - 3$
10.  $Y = sX + m$

### GENERATORY APLIKACYJNE

Pierwszy generator aplikacyjny to Excel 2016. Liczby o rozkładzie jednostajnym powstają w tym programie za pomocą algorytmu Mersenne Twister (Matsumoto i Nishimura, 1998). Jednakże algorytm generowania liczb pseudoso-



wych o rozkładzie normalnym za pomocą narzędzia „analiza danych” nie jest znany. Drugim generatorem aplikacyjnym jest język R. Także w nim liczby o rozkładzie jednostajnym powstają za pomocą algorytmu Mersenne Twister. Z kolei liczby o rozkładzie normalnym są tworzone metodą odwracania dystrybuanty zaproponowaną przez Wichurę (1988). Do generatorów aplikacyjnych nie można zaliczyć Mathcada, ponieważ w jego przypadku brakuje informacji na temat sposobu generowania liczb pseudolosowych o rozkładzie jednostajnym. Do tej grupy nie można zaliczyć również programu Calc z zestawu Apache OpenOffice 4.1.5. Wprawdzie liczby losowe o rozkładzie jednostajnym generowane są w tym programie za pomocą algorytmu Mersenne Twister, ale brakuje narzędzia do generowania liczb pseudolosowych o rozkładzie normalnym. Ze względu na specyfikę tych generatorów poniżej opisano sposób ich otrzymywania w dwóch środowiskach aplikacyjnych.

### Generator Excel

W programie Microsoft Excel 2016 do generowania liczb pseudolosowych o rozkładzie normalnym wykorzystuje się narzędzie „analiza danych” oraz opcję „generowanie liczb pseudolosowych”. Należy wybrać polecenie „rozkład normalny” oraz określić wartości  $m$  i  $s$ , liczbę liczb pseudolosowych  $n$  i liczbę zmiennych  $u$ .

### Generator R

W środowisku R za pomocą formuły  $mnorm(n, mean = 0, sd = 1)$  otrzymano  $n$  liczb  $X$  z  $N(0,1)$ . Wartość  $Y$  z  $N(m, s)$  otrzymano za pomocą formuły  $Y = Xs + m$ . Ciąg takich  $n$  liczb można wyeksportować do pliku csv za pomocą polecenia: `write.csv2(mnorm(n), 'plik.csv', row.names = F, na = "", fileEncoding = "Windows-1250")`.

## TESTOWANIE GENERATORÓW

Testowanie generatorów liczb pseudolosowych polega na sprawdzaniu hipotezy, że otrzymany za pomocą danego generatora ciąg liczb jest próbka z populacji o danej dystrybucji, w tym przypadku o dystrybucji rozkładu normalnego wyrażonej wzorem (3).

Celem tej części pracy jest uzupełnienie metod testowania generatorów o sposoby nieopisane w pracy Thomasa i współpracowników (2007). Przedstawione wyniki dotyczą rozkładu normalnego standaryzowanego  $N(0,1)$ . Testowanie generatorów przeprowadzono za pomocą sześciu kryteriów: czterech klasycznych i dwóch nowych. W przypadku kryteriów klasycznych wyznaczono przy użyciu omawianych generatorów podstawowe charakterystyki rozkładu

i porównano je z wartościami teoretycznymi, zastosowano testy Lillieforsa i Andersona-Darlinga oraz test trendu i test sezonowości. W przypadku nowych kryteriów zbadano zgodność dystrybuant teoretycznej i empirycznej, a także przetestowano generatory za pomocą statystyk pozycyjnych. Symulacje Monte Carlo na potrzeby analizowanych kryteriów przeprowadzono dla  $u = 1000$  zestawów (liczby powtórzeń) prób o liczebności  $n = 1000$ . Wszystkie obliczenia przeprowadzono na takich samych zbiorach danych. Wartości  $n$  mogłyby być większe, gdyby nie kryterium opisane przez statystyki pozycyjne (zob. *Kryterium 4: statystyki pozycyjne*). Ograniczeniami dla wartości  $u$  są tylko czas trwania symulacji i liczba 16384 kolumn w arkuszu. Przy bardzo wydajnych procesorach z drugim ograniczeniem na pewno można sobie poradzić.

### Kryterium 1: charakterystyki rozkładu normalnego standaryzowanego

Algorytm wyznaczania wartości średnich charakterystyk dla generatora jest następujący:

1. Ustal liczebność próby  $n$
2. Ustal wartość zmiennych  $S_j = 0$  ( $j = 1, 2, \dots, 4$ )
3. Kroki 3.1–3.3 powtórz  $u$  razy:
  - 3.1. Wygeneruj  $n$ -elementową próbę za pomocą generatora
  - 3.2. Oblicz wartości charakterystyk  $\alpha_1, \sqrt{M_2^n}, M_3^n, M_4^n$  zgodnie ze wzorami (6)–(8)
  - 3.3. Oblicz  $S_1 = S_1 + |\alpha_1 - m|$ ,  $S_2 = S_2 + |\sqrt{M_2^n} - s|$ ,  $S_3 = S_3 + |M_3^n - \mu_3|$  i  $S_4 = S_4 + |M_4^n - \mu_4|$ , gdzie  $m, s, \mu_3$  i  $\mu_4$  przyjmują odpowiednio wartości: 0, 1, 0 i 3
4. Oblicz  $\bar{S}_j = \frac{S_j}{u}$  ( $j = 1, 2, \dots, 4$ )

Tablica 1 przedstawia wyniki uzyskane za pomocą powyższego algorytmu dla liczebności próby  $n$  i liczby powtórzeń  $u$ . Wartości najmniejsze zostały pogrubione.

TABL. 1. WARTOŚCI MIAR  $\bar{S}_j$  DLA  $n = u = 10^3$  WRAZ Z RANKINGIEM

Generatory	$\bar{S}_1$		$\bar{S}_2$		$\bar{S}_3$		$\bar{S}_4$	
		R		R		R		R
CdfInV .....	0,02520	7	0,01829	10	0,06207	10	<b>0,11811</b>	1
BM .....	0,02527	8	0,02762	14	<b>0,05057</b>	1	0,22614	13
Tuzin .....	0,02570	10	<b>0,01649</b>	1	<b>0,05836</b>	2	0,13594	12
Biegun .....	<b>0,02447</b>	3	0,01800	6	0,05901	4	0,12149	4
Iloraz .....	0,02616	13	0,01846	11	0,06432	12	<b>0,12016</b>	3
Levy .....	0,02619	14	0,01886	13	<b>0,05900</b>	3	0,12341	8
AD .....	0,02497	6	0,01805	8	0,06203	9	<b>0,12005</b>	2
MB .....	0,02567	9	<b>0,01736</b>	2	0,06154	6	0,12579	10

TABL. 1. WARTOŚCI MIAR  $\bar{S}_j$ , DLA  $n = u = 10^3$  WRAZ Z RANKINGIEM (dok.)

Generatory	$\bar{S}_1$		$\bar{S}_2$		$\bar{S}_3$		$\bar{S}_4$	
		R		R		R		R
Grand .....	<b>0,02388</b>	2	0,01828	9	0,06257	11	0,12555	9
MP .....	<b>0,02383</b>	1	0,01778	4	0,06167	7	0,12309	7
Wallace'a .....	0,02583	11	0,01788	5	0,06473	13	0,12231	6
Ziggurat .....	0,02469	4	0,01805	7	0,06177	8	0,13350	11
Excel .....	0,02591	12	0,01868	12	0,08069	14	0,29809	14
R .....	0,02489	5	<b>0,01776</b>	3	0,06057	5	0,12180	5

U w a g a.  $j = 1, 2, 3, 4$ . R – ranking.

Ź r ó d ł o: opracowanie własne.

Wyniki zaprezentowane w tabl. 1 wskazują, że ze względu na charakterystyki rozkładu  $N(0,1)$  wyróżniają się generatory Biegun, R i MP. Najgorzej z analizowanych generatorów wypadają Excel, Iloraz i Levy.

### Kryterium 2: testy Lillieforsa i Andersona-Darlinga

Generatory pseudolosowych liczb normalnych sprawdzono za pomocą testu Andersona-Darlinga (A-D) (Anderson i Darling, 1952) oraz testu Lillieforsa (LF) (Lilliefors, 1967). Pominięto test Shapiro-Wilka (Shapiro i Wilk, 1965), stosowany w swojej pierwotnej postaci dla  $n \leq 5000$ , którego skuteczność jest podobna do skuteczności testu A-D (Razali i Wah, 2011). Natomiast test Kołmogorowa-Smirnowa (K-S), również pominięty, charakteryzuje się najmniejszą mocą spośród wyżej wymienionych (Razali i Wah, 2011).

Test LF, sprawdzający, czy próba  $x_1, x_2, \dots, x_n$  ma rozkład  $N(0,1)$ , korzysta ze statystyki:

$$L = \max_{1 \leq i \leq n} \left\{ \max \left\{ \left| \Phi \left( \frac{x - \alpha_1}{\sqrt{M_2^n}} \right) - \frac{i-1}{n} \right|, \left| \Phi \left( \frac{x - \alpha_1}{\sqrt{M_2^n}} \right) - \frac{i}{n} \right| \right\} \right\} \quad (9)$$

gdzie wartość  $\alpha_1$  wyrażona jest wzorem (6), zaś wartość  $\Phi(\cdot)$  – wzorem (3) oraz

$$\sqrt{M_2^n} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \alpha_1)^2} \quad (10)$$

Wartość  $p$ -value  $p_V$  dla testu LF wyznaczono za pomocą algorytmu (Abdi i Molin, 2007):

1. Niech  $b_0 = 0,37872256037043$ ,  $b_1 = 1,3074818507879 + n$
2. Niech  $b_2 = 0,08861783849346$

3. Oblicz  $a = \left(-b_1 + \sqrt{b_1^2 - 4b_2(b_0 - L^{-2})}\right) / (2b_2)$
4. Oblicz  $p_1 = -0,37782822932809 + 1,67819837908004a - 3,02959249450445a^2$
5. Oblicz  $p_2 = 2,8001579814210a^3 - 1,39874347510845a^4 - 0,40466213484419a^5$
6. Oblicz  $p_3 = -0,06353440854207a^6 + 0,0028746208762a^7 + 0,000696500131a^8$
7. Oblicz  $p_4 = -0,00011872227037a^9 + 0,00000575586834a^{10}$
8. Oblicz  $p_V = p_1 + p_2 + p_3 + p_4$

Test A-D stanowi modyfikację testu Craméra-von Misesa (Cramér, 1928) dokonaną w celu poprawy jego czułości w „ogonach” rozkładu normalnego. Wykorzystuje on statystykę w postaci (Anderson i Darling, 1952; D’Agostino, 1986):

$$AD = A^2 \left(1 + \frac{0,75}{n} + \frac{2,25}{n^2}\right) \quad (11)$$

gdzie:

$$A^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} \left[ \ln \Phi(X_{(i)}) + \ln \left(1 - \Phi(X_{(n+1-i)})\right) \right],$$

$X_{(i)}$  –  $i$ -ta zaobserwowana wartość w uporządkowanej rosnąco  $n$ -elementowej próbie.

Formuła na wyznaczenie  $p_V$  dla testu A-D dana jest wyrażeniem (D’Agostino, 1986):

$$p_V = \begin{cases} \exp(1,2937 - 5,709AD + 0,0186AD^2) & \text{dla } AD \geq 0,6 \\ \exp(0,9177 - 4,279AD - 1,38AD^2) & \text{dla } 0,34 \leq AD < 0,6 \\ 1 - \exp(-8,318 + 42,796AD - 59,938AD^2) & \text{dla } 0,2 < AD < 0,34 \\ 1 - \exp(-13,436 + 101,14AD - 223,73AD^2) & \text{dla } AD \leq 0,2 \end{cases} \quad (12)$$

Wyznaczanie wartości  $p$ -value dla testów LF i A-D zaimplementowano komputerowo (szczegóły znajdują się w rozdziale *Jak skorzystać z plików internetowych*).

Tablica 2 przedstawia średnie wartości  $p$ -value dla analizowanych testów, liczebności próby  $n = 10^3$  i liczby powtórzeń  $u = 10^3$ . Wartości największe zostały pogrubione. Dla  $n = 10^6$  otrzymano tak wysokie wartości statystyk (11), że formuła (12) sygnalizowała błąd. Liczebność próby  $n = 10^6$  wpływała negatywnie także na test LF, czego skutkiem były ujemne wartości  $p$ -value dla generatorów Tuzin, Levy i MB.

TABL. 2. ŚREDNIE WARTOŚCI  $p$ -value DLA  $n = u = 10^3$  WRAZ Z RANKINGIEM

Generatory	LF		A-D	
		R		R
Cdflnv .....	0,48018	5	0,13879	12
BM .....	0,44516	14	0,12578	14
Tuzin .....	0,46894	13	0,14794	6
Biegun .....	0,47086	12	<b>0,15380</b>	2
lloraz .....	0,48130	4	0,14165	9
Levy .....	0,47912	6	0,13917	10
AD .....	0,47580	9	0,14902	5
MB .....	<b>0,48298</b>	3	0,15017	4
Grand .....	<b>0,47315</b>	2	<b>0,13563</b>	3
MP .....	0,47835	7	0,14185	8
Wallace'a .....	0,47485	10	0,13908	11
Ziggurat .....	0,47832	8	0,13820	13
Excel .....	0,47232	11	0,14396	7
R .....	<b>0,49191</b>	1	<b>0,16000</b>	1

Źródło: jak przy tabl. 1.

Wyniki zaprezentowane w tabl. 2 wskazują, że w testach LF i A-D pozytywnie wyróżniają się generatory R, Grand i MB. Najgorsze okazują się generatory BM, Wallace i Ziggurat.

### Kryterium 3: zgodność dystrybuant teoretycznej i empirycznej

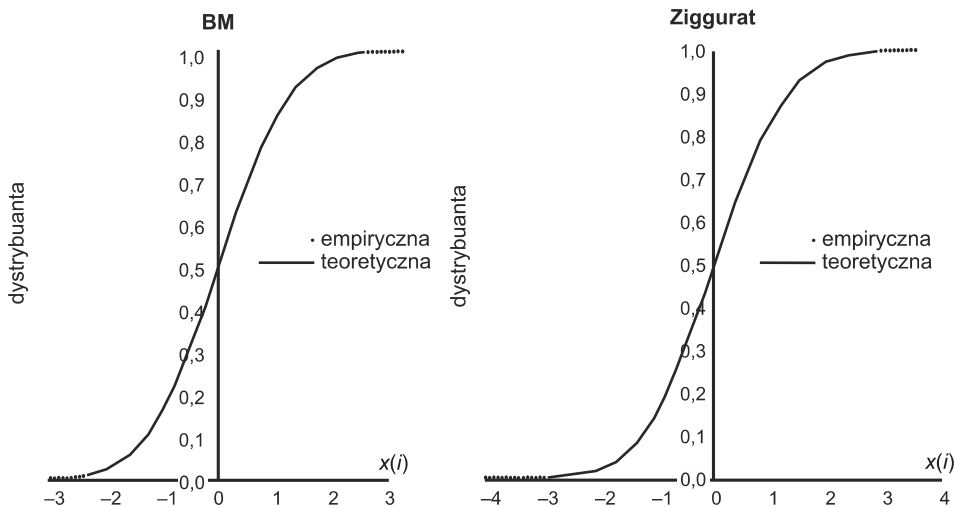
Propozycja autora dotycząca testowania jakości generatorów polega na zbadaniu zachowania się dystrybuanty empirycznej i teoretycznej. W literaturze statystycznej można znaleźć różne definicje dystrybuanty empirycznej. Dystrybuanta empiryczna, przeznaczona wyłącznie dla rozkładu normalnego, ma postać (Bloom, 1958):

$$F_i = \frac{i - 0,375}{n + 0,25} \quad (13)$$

Chociaż przebiegi dystrybuant empirycznej i teoretycznej można oczywiście przedstawić na wykresie, trudno jest w tym przypadku sporządzić ranking generatorów. Wykres 3 przedstawia przebiegi dystrybuant dla generatorów BM oraz Ziggurat, gdy  $n = 10^4$ .

Omówione wcześniej testy LF i A-D, podobnie jak testy K-S (Kołmogorow, 1933; Smirnov, 1948), Craméra-von Misesa (Cramér, 1928), Watsona (1961), R Rényiego (Rényi, 1953) i Kuipera (1960) zaliczane są do rodziny testów do badania normalności danej próby; grupa ta jest zdefiniowana na podstawie dystrybuanty empirycznej.

WYKR. 3. DYSTRYBUANTY EMPIRYCZNA I TEORETYCZNA  
DLA GENERATORÓW BM I ZIGGURAT



Źródło: jak przy tabl. 1.

Do badania jakości generatorów można także wykorzystać miarę:

$$M = \sum_{i=1}^n |\Phi(x_i) - F_i| \quad (14)$$

gdzie:

$F_i$  – dystrybuanta empiryczna (13),

$\Phi(x_i)$  – dystrybuanta teoretyczna rozkładu normalnego (3),

$M \geq 0$ .

Przy idealnej zgodności próby z rozkładem normalnym standaryzowanym  $M = 0$ . Im mniejsza wartość  $M$ , tym lepszą jakością cechuje się dany generator.

Algorytm wyznaczania wartości średniej miary  $M$  dla generatora jest następujący:

1. Ustal liczebność próby  $n$
2. Kroki 2.1–2.3 powtórz  $u$  razy:
  - 2.1. Wygeneruj  $n$  liczb pseudolosowych  $x_1, x_2, \dots, x_n$  z  $N(0,1)$
  - 2.2. Liczby uzyskane w kroku 2.1 posortuj w porządku rosnącym
  - 2.3. Oblicz wartość miary  $M = \sum_{i=1}^n |F_i - \Phi(x_i)|$
3. Oblicz  $\bar{M} = \frac{1}{u} \sum_{j=1}^u M_j$

Tablica 3 przedstawia średnie wartości miary  $M$  otrzymane przy użyciu tego algorytmu dla różnych wartości  $n$  i  $u$  wraz z rankingiem. Wartości najmniejsze zostały pogrubione.

TABL. 3. WARTOŚCI MIARY  $\bar{M}$  DLA  $n = u = 10^3$  WRAZ Z RANKINGIEM

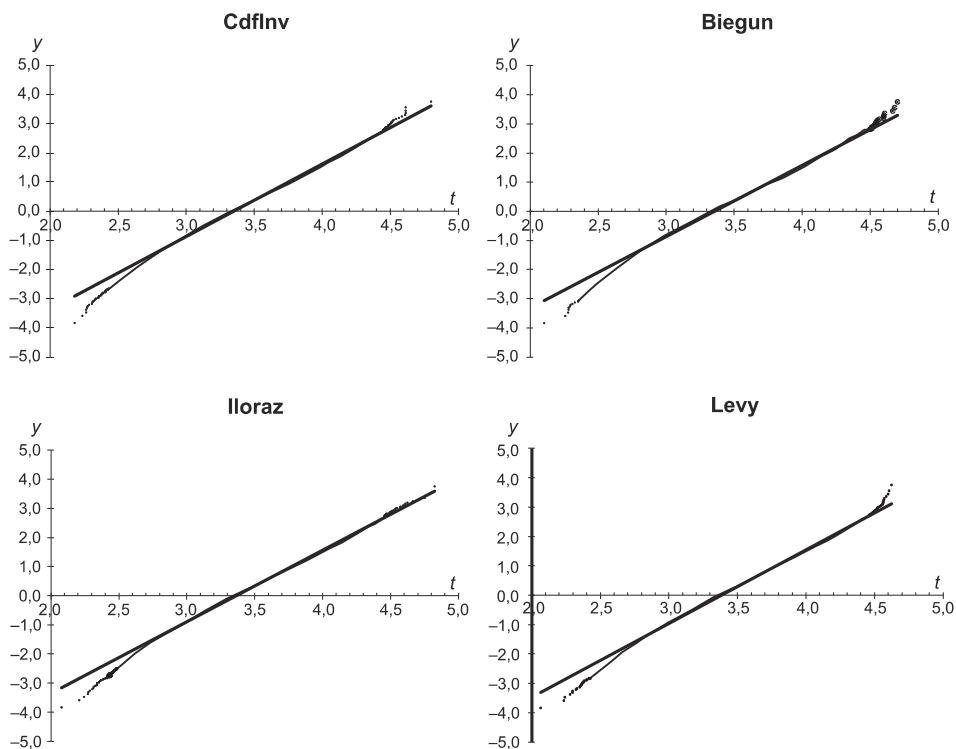
Generatory	$\bar{M}$	R	Generatory	$\bar{M}$	R
Cdflnv .....	9,9766	8	MB .....	9,9069	5
BM .....	10,1210	11	Grand .....	10,1767	14
Tuzin .....	10,0909	10	MP .....	<b>9,6993</b>	1
Biegun .....	<b>9,7563</b>	3	Wallace'a .....	10,0787	9
Iloraz .....	10,1480	13	Ziggurat .....	<b>9,7460</b>	2
Levy .....	10,1355	12	Excel .....	9,9222	7
AD .....	9,8337	4	R .....	9,9105	6

Źródło: jak przy tabl. 1.

Wyniki zaprezentowane w tabl. 3 wskazują, że na podstawie kryterium (19) na docenienie zasługują generatory MB, Ziggurat i Biegun. Najgorzej wypadają generatory Grand, Iloraz i Levy.

Rozkład wartości miary  $M$  dla generatorów ma charakter logarytmiczno-normalny. Sytuację tę potwierdza siatka prawdopodobieństwa dla wybranych generatorów, gdzie  $t = \ln(M)$  oraz funkcja kwantylowa  $y = \Phi^{-1}(F_i) = \Phi^{-1}[(i - 0,375)/(n + 0,25)]$  (wykr. 4).

WYKR. 4. SIATKA PRAWDOPODOBIEŃSTWA



Uwaga. Jak przy wyk. 1.

Źródło: jak przy tabl. 1.

### Kryterium 4: statystyki pozycyjne

Drugą propozycją autora w kwestii testowania jakości generatorów jest zastosowanie statystyk pozycyjnych. Wcześniej już wyznaczono wartość miary (14), która bada zachowanie się dystrybuant empirycznej i teoretycznej. Ciekawym zagadnieniem jest także wyznaczenie wartości miary oznaczonej symbolem MSP, opisującej różnicę pomiędzy tymi dystrybuantami dla różnych statystyk pozycyjnych  $x_{(i)}$ , gdy  $i$  przyjmuje wartości początkowe, środkowe i końcowe. Dokonano oceny zarówno numerycznej (tabl. 4), jak i wizualnej (wykr. 5).

Algorytm wyznaczania wartości średniej miary MSP (17) przedstawia się następująco:

1. Wygeneruj tablicę  $n \times u$  liczb pseudolosowych  $x_{ij}$  ( $i = 1, \dots, n$ ;  $j = 1, \dots, u$ ) z  $N(0,1)$
2. Tablicę  $n \times u$  posortuj kolumnami w kolejności rosnącej
3. Z tablicy  $n \times u$  wybierz wiersze (statystyki pozycyjne)  $i = 1, 10i$  ( $i = 1, \dots, n/10$ ) i utwórz tablicę  $(n/10 + 1) \times u$
4. Transponuj tablicę z pkt 3
5. Tablicę  $u \times (n/10 + 1)$  posortuj kolumnami w kolejności rosnącej. Kolumny odpowiadają statystynom pozycyjnym  $i = 1, 10i$  ( $i = 1, \dots, n/10$ ). W kolumnach znajdują się  $u$ -elementowe próby statystyk pozycyjnych
6. Oblicz dystrybuanty empiryczne rozkładów statystyk pozycyjnych  $F_j = j/(u + 1)$
7. Ustal  $S = 0$  oraz  $licznik = 0$ . Dla  $i = 1, 10i$  ( $i = 1, \dots, n/10$ ) wykonaj kroki 7.1–7.4:
- 7.1. Oblicz dystrybuanty teoretyczne rozkładów statystyk pozycyjnych dla  $j = 1, \dots, u$  za pomocą formuły:

$$\Phi_{j(i)}(x_{ji}) = \sum_{k=i}^n \binom{n}{k} \Phi(x_{ji})^k [1 - \Phi(x_{ji})]^{n-k} \quad (15)$$

gdzie  $\Phi(\cdot)$  – dystrybuanta teoretyczna (3).

W celu skorzystania z odpowiednio dużych wartości  $n$  symbol Newtona z (15) obliczono za pomocą samodzielnie wyprowadzonej formuły powstałej w wyniku przekształceń podstawowej definicji symbolu Newtona. Formuła ta ma postać:

$$\binom{n}{k} = \exp \left[ \sum_{i=1}^n \ln(i) - \sum_{i=1}^k \ln(i) - \sum_{i=1}^{n-k} \ln(i) \right] \quad (16)$$



7.2. Oblicz wartość miary dla  $j = 1, \dots, u$

$$MSP_{(i)} = \sum_{j=1}^u |F_j - \Phi_{j(i)}| \quad (17)$$

7.3. Oblicz  $S = S + MSP_{(i)}$

7.4. Zwiększ iterację *licznik* = *licznik* + 1

8. Oblicz wartość średnią miary (17) za pomocą formuły  $\overline{MSP} = S/\text{licznik}$

Tablica 4 przedstawia wyniki uzyskane za pomocą tego algorytmu dla  $n = 10^3$  oraz liczby powtórzeń  $u = 10^3$ . Pogrubiono najmniejsze wartości. Ze względu na specyfikę wzoru (15) zbyt duże wartości  $n$  przysparzają problemów numerycznych. Analiza numeryczna wykazała, że maksymalna wielkość próby, z jakiej można skorzystać w (15) dla dowolnej wartości  $k \leq n$ , to tylko  $n = 1029$ . Ten fakt nie pozostaje bez wpływu na jakość omawianej metody.

**TABL. 4. WARTOŚCI MIARY  $\overline{MSP}$  DLA  $n = u = 10^3$  WRAZ Z RANKINGIEM**

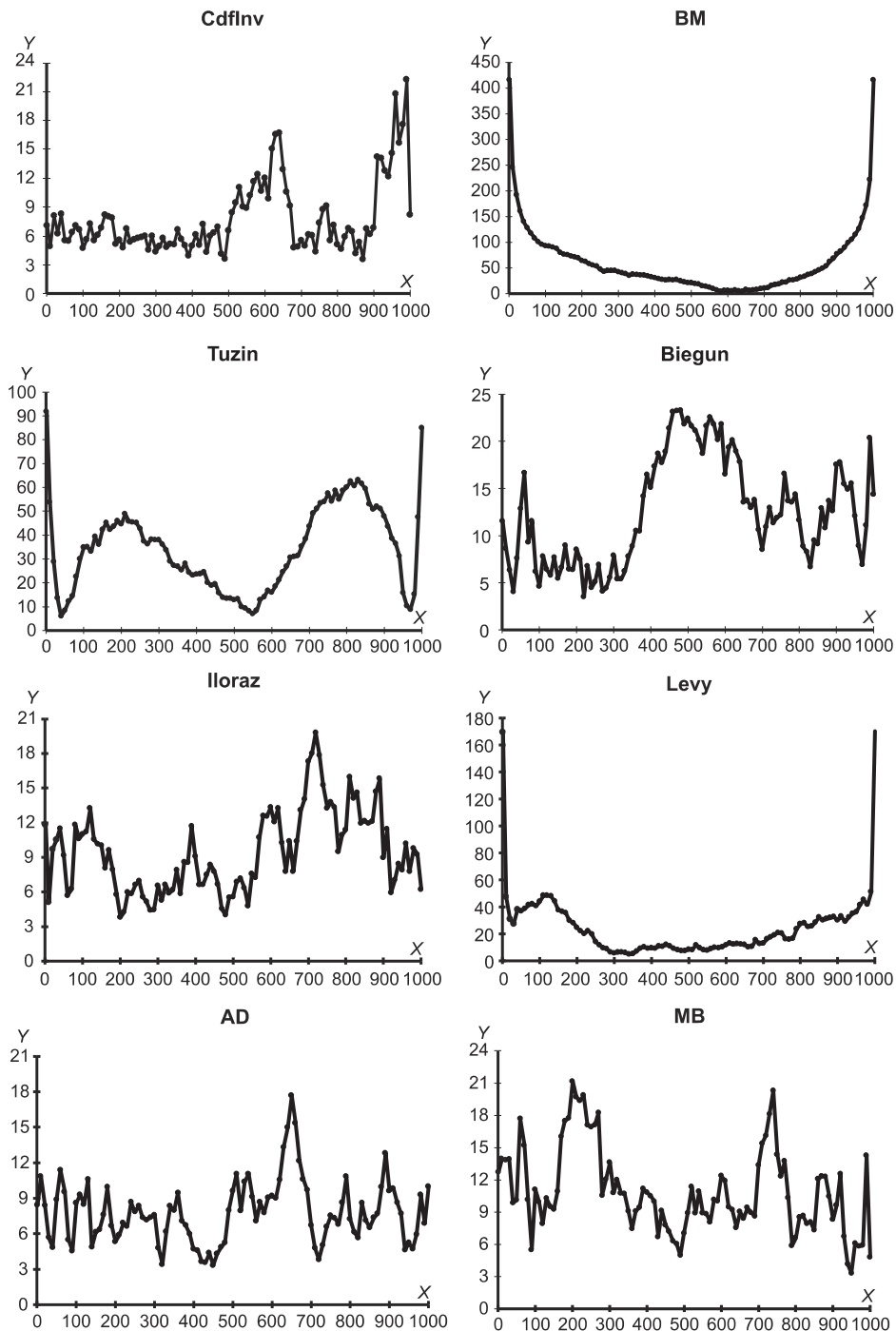
Generatory	<i>MSP</i>	R
Cdflnv .....	<b>7,7493</b>	2
BM .....	61,1743	14
Tuzin .....	34,2330	13
Biegun .....	12,5343	9
Iloraz .....	9,4371	6
Levy .....	24,4643	12
AD .....	<b>7,7820</b>	3
MB .....	10,9223	7
Grand .....	12,0608	8
MP .....	9,4006	5
Wallace'a .....	15,5867	11
Ziggurat .....	<b>7,3861</b>	1
Excel .....	13,9730	10
R .....	8,9455	4

Źródło: jak przy tabl. 1.

Z tabl. 4 wynika, że dla statystyk pozycyjnych  $i = 1, 10, 20, \dots, 1000$  średnia wartość miary rozbieżności między dystrybuantami teoretyczną i empiryczną (13) przyjmuje najmniejsze wartości dla generatorów Ziggurat, Cdflnv i AD. Według tego kryterium najniższą jakością cechują się generatory BM, Tuzin i Levy.

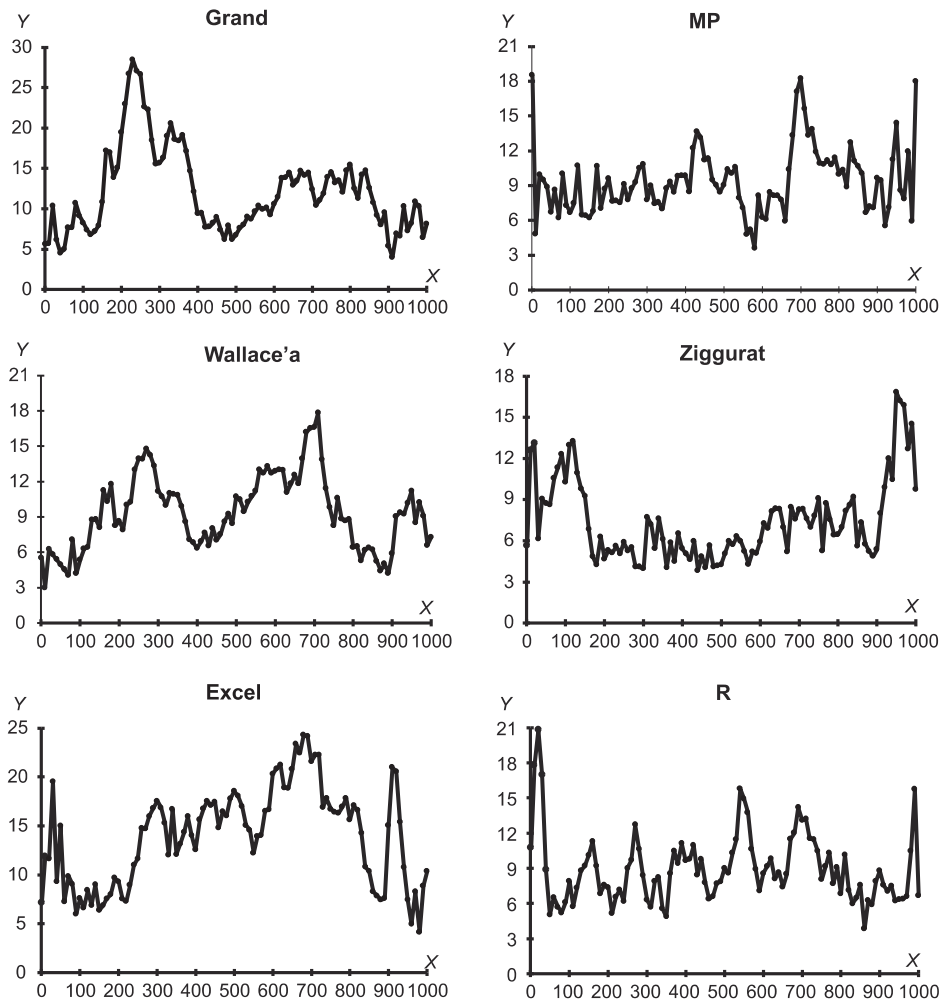
W celu precyzyjniejszej analizy wartości zaprezentowanych w tabl. 4 przedstawiono na wyk. 5 przykładowe zależności między wartościami  $i$ -tej ( $i = 1, 10, 20, \dots, 1000$ ) statystyki pozycyjnej a wartościami miary rozbieżności (17), której wartości średnie przedstawia tabl. 4.

**WYKR. 5. MIARA ROZBIEŻNOŚCI DYSTRYBUANT DLA STATYSTYK POZYCYJNYCH**



oś Y – miara rozbieżności dystrybuant, oś X – numer statystyki pozycyjnej

WYKR. 5. MIARA ROZBIEŻNOŚCI DYSTRYBUANT DLA STATYSTYK POZYCYJNYCH (dok.)



oś Y – miara rozbieżności dystrybuant, oś X – numer statystyki pozycyjnej

Źródło: jak przy tabl. 1.

Przebieg miary rozbieżności dystrybuant dla poszczególnych statystyk pozycyjnych okazał się szczególnie nietypowy w przypadku generatorów BM, Tuzin i Levy.

### Kryteria 5 i 6: testy trendu i sezonowości

W celu rzetelnej interpretacji wyników zaprezentowanych na wyk. 5 przeprowadzono testy trendu i sezonowości dla każdej  $n$ -elementowej próby wygenerowanej  $u$ -krotnie przy poziomie istotności  $\alpha = 0,05$ , zakładając hipotezę  $H_0$ : w ciągu

liczb pseudolosowych trend i sezonowość nie występują. Testy trendu i sezonowości zostały omówione np. w: Sulewski (2000). Implementacja komputerowa tych testów dla  $n$ -elementowej próby znajduje się w pliku trend\_sezon.xlsm.

Algorytm wyznaczania średnich wartości  $p$ -value testów trendu i sezonowości dla generatora jest następujący:

1. Ustal liczebność próby  $n$
2. Ustal wartość zmiennych  $H_T = 0, H_S = 0$
3. Poniższe trzy kroki powtórz  $u$  razy:
  - 3.1. Wygeneruj  $n$  liczb pseudolosowych  $x_1, x_2, \dots, x_n$  z  $N(0,1)$
  - 3.2. Oblicz wartości  $p$ -value testów trendu i sezonowości odpowiednio  $p_{vT}$  i  $p_{vS}$
  - 3.3. Jeżeli nie ma podstaw do odrzucenia  $H_0$ , to  $H_T = H_T + 1$  (dla trendu) i  $H_S = H_S + 1$  (dla sezonowości)
4. Oblicz średnie wartości  $p$ -value  $\overline{p_{vT}}$  (dla trendu) i  $\overline{p_{vS}}$  (dla sezonowości)
5. Prawdopodobieństwo braku podstaw do odrzucenia  $H_0$  wynosi  $p_T = H_T/u$  (dla trendu) i  $p_S = H_S/u$  (dla sezonowości).

Tablica 5 przedstawia średnie wartości  $p$ -value testu trendu, średnie wartości  $p$ -value testu sezonowości oraz prawdopodobieństwo braku podstaw do odrzucenia  $H_0$ . Wartości największe zostały pogrubione.

**TABL. 5. TESTY TRENDU I SEZONOWOŚCI DLA  $n = u = 10^3$  WRAZ Z RANKINGIEM**

Generatory	Test trendu				Test sezonowości			
	$p_{vT}$		$p_T$		$p_{vS}$		$p_S$	
		R		R		R		R
CdfInv .....	0,2466	13	0,932	13	0,5265	8	0,839	14
BM .....	0,2486	10	0,945	9	0,5261	9	0,870	4
Tuzin .....	<b>0,2580</b>	1	0,946	7	<b>0,5426</b>	1	<b>0,872</b>	2
Biegun .....	0,2510	5	0,946	7	0,5349	4	0,869	5
Iloraz .....	0,2481	11	0,942	10	0,5252	10	0,851	13
Levy .....	0,2543	4	<b>0,954</b>	2	<b>0,5366</b>	2	0,867	6
AD .....	0,2509	6	<b>0,951</b>	3	0,5116	14	0,854	12
MB.....	0,2497	9	<b>0,955</b>	1	0,5224	12	0,864	8
Grand .....	0,2474	12	0,929	14	0,5349	5	<b>0,871</b>	3
MP .....	<b>0,2543</b>	3	0,951	3	0,5304	7	0,861	9
Wallace'a .....	0,2509	7	0,940	11	0,5172	13	0,857	11
Ziggurat .....	0,2464	14	0,949	5	0,5324	6	<b>0,874</b>	1
Excel .....	<b>0,2545</b>	2	0,949	5	0,5244	11	0,860	10
R .....	0,2501	8	0,940	11	<b>0,5365</b>	3	0,866	7

Źródło: jak przy tabl. 1.

Z tabl. 5 wynika, że najlepiej z punktu widzenia testów trendu i sezonowości prezentują się generatory Levy, Tuzin i MP. Najniższymi wartościami  $p$ -value cechują się generatory CdfInv, Iloraz i Wallace'a.

## JAK SKORZYSTAĆ Z PLIKÓW INTERNETOWYCH

Pliki niezbędne do przeprowadzenia badań zostały umieszczone pod adresem <https://sulewski.apsl.edu.pl/index.php/publikacje>. Plik `generatory.xlsm` zawiera kody źródłowe większości generatorów zapisane w języku VBA wraz z przykładami ich zastosowania. Plik `testy.xlsm` umożliwia obliczanie wartości  $p$ -value dla testów LF i A-D. Plik `trend_sezon.xlsm` pozwala na wyznaczanie wartości  $p$ -value dla testów trendu i sezonowości. Plik `c++.zip` zawiera foldery dotyczące omawianych metod generacji liczb pseudolosowych. W każdym z tych folderów znajdują się pliki `*.cpp` i `*.exe`, gdzie symbol `*` oznacza nazwę danego generatora.

Wyniki są zapisywane w formie tablicy dwuwymiarowej w pliku `results.txt` zamieszczonym w każdym folderze. W związku z tym przy uruchomieniu danego generatora należy podać liczbę kolumn  $kol$ , liczbę wartości w każdej kolumnie  $n$  oraz parametry rozkładu normalnego  $m$  i  $s$ . Dzięki separatorowi typu tabulator można bardzo łatwo wczytać otrzymane wyniki do arkusza kalkulacyjnego w celu dalszej obróbki statystycznej.

## PODSUMOWANIE

Głównym celem pracy jest analiza generatorów liczb pseudolosowych normalnych za pomocą różnego rodzaju kryteriów. Do badań wybrano najważniejsze zdaniem autora `generatory`, z których każdy zdefiniowano algorytmicznie i oprogramowano w języku C++, a większość także w języku VBA. Oprócz powszechnie stosowanych kryteriów, takich jak charakterystyki rozkładu (kryterium 1), testy zgodności (kryterium 2) czy testy trendu i sezonowości (kryteria 5 i 6), zastosowano także dwa nowe podejścia w postaci miary wykorzystującej sumy modułów różnic dystrybuant empirycznej i teoretycznej (kryterium 3) oraz miary korzystającej ze statystyk pozycyjnych (kryterium 4). Na podstawie uśrednienia rankingów analizowanych generatorów można dokonać ich końcowej klasyfikacji. Wyniki pozwalają stwierdzić, że najwyższą jakością cechują się generatory Monty Pythona (MP), R, Biegun i Ziggurat, podczas gdy najgorsze okazały się generatory Boxa-Mullera (BM), Wallace'a, Iloraz i Excel.

## BIBLIOGRAFIA

- Abdi, H., Molin, P. (2007). Lilliefors/Van Soest's test of normality. W: N. Salkind (red.), *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage.
- Ahrens, J. H., Dieter, U. (1972). Computer methods for sampling from the exponential and normal distributions. *Communication of the ACM*, 15(10), 873–882.
- Ahrens, J. H., Dieter, U. (1988). Efficient table-free sampling methods for the exponential, Cauchy, and normal distributions. *Communication of the ACM*, 31(11), 1330–1337.

- Anderson, T. W., Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23(2), 193–212. DOI: 10.1214/aoms/1177729437.
- Bell, J. R. (1968). Algorithm 334: Normal random deviates. *Communication of the ACM*, 11(7), 498.
- Bloom, G. (1958). *Statistical estimates and transformed Beta variables*. New York: Wiley.
- Box, G. E. P., Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29(2), 610–611.
- Brent, R. P. (1974). Algorithm 488: A Gaussian pseudo-random number generator. *Communication of the ACM*, 17(12), 704–706.
- Chen, J., Moon, J., Bazargan, K. (2004). Reconfigurable readback-signal generator based on a field programmable gate array. *IEEE Transactions on Magnetics*, 40(3), 1744–1750.
- Chen, W., Burford, R. L. (1981). Quality evaluation of some combinations of unit uniform random number generators and unit normal transformation algorithms. W: R. M. Huhn, E. R. Comer, F. O. Simons Jr. (red.), *ANSS'81: Proceedings of the 14th Annual Symposium on Simulation* (s. 129–149). Tampa: IEEE Press Piscataway.
- Cramér, H. (1928). On the Composition of Elementary Errors. *Scandinavian Actuarial Journal*, 1(1), 13–74.
- Cramér, H. (1958). *Metody matematyczne w statystyce*. Warszawa: PWN.
- D'Agostino, R. B. (1986). Tests for the Normal Distribution. W: R. B. D'Agostino, M. A. Stephens (red.), *Goodness-of-Fit Techniques* (s. 367–413). New York: Marcel Dekker.
- Gebhardt, F. (1964). Generating normally distributed random numbers by inverting the normal distribution function. *Mathematics of Computation*, 18(86), 302–306.
- Kabal, P. (2000). *Generating Gaussian pseudo-random deviates*. McGill University.
- Kinderman, A. J., Monahan, J. F. (1977). Computer generation of random variables using the ratio of uniform deviates. *ACM Transactions on Mathematical Software*, 3(3), 257–260.
- Knop, R. (1969). Remark on Algorithm 334 [g5]: normal random deviates. *Communication of the ACM*, 12(5), 281.
- Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, (4), 83–91.
- Kronmal, R. (1964). Evaluation of a pseudorandom normal number generator. *Journal of the ACM*, 11(3), 357–363.
- Kuiper, N. H. (1960). Tests concerning random points on a circle. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen Series A*, (63), 38–47.
- L'Ecuyer, P. (1992). Testing random number generators. W: J. J. Swain, D. Goldsman, R. C. Crain, J. R. Wilson (red.), *Proceedings of the 24th Conference on Winter Simulation* (s. 305–313). New York: ACM Press.
- L'Ecuyer, P. (2001). Software for uniform random number generation: distinguishing the good and the bad. W: *Proceedings of the 33rd Conference on Winter Simulation* (s. 95–105). Washington: IEEE Computer Society.
- L'Ecuyer, P., Simard, R. (2005). TestU01. Pobrane z: <http://simul.iro.umontreal.ca/>.
- L'Ecuyer, P., Simard, R. (2007). TestU01: AC library for empirical testing of random number generators. *ACM Transactions on Mathematical Software*, 33(4), 22.
- Leva, J. L. (1992). A fast normal random number generator. *ACM Transactions on Mathematical Software*, 18(4), 449–453.
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, 62(318), 399–402.

- Marsaglia, G. (1997). *The Diehard random number test suite*. Pobrane z: <https://github.com/nmondal/diehard.c>.
- Marsaglia, G., Bray, T. A. (1964). A convenient method for generating normal variables. *SIAM Review*, 6(3), 260–264.
- Marsaglia, G., Tsang, W. W. (1998). The Monty Python Method for Generating Random Variables. *ACM Transactions on Mathematical Software*, 24(3), 341–350.
- Marsaglia, G., Tsang, W. W. (2000). The Ziggurat Method for Generating Random Variables. *Journal of Statistical Software*, 5(8), 1–7.
- Matsumoto, M., Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1), 3–30.
- McCollum, J. M., Lancaster, J. M., Bouldin, D. W., Peterson, G. D. (2003). Hardware acceleration of pseudo-random number generation for simulation applications. W: *IEEE Southeastern Symposium on System Theory*. Morgantown: IEEE. DOI: 10.1109/SSST.2003.1194578.
- Mélard, G. (2014). On the accuracy of statistical procedures in Microsoft Excel 2010. *Computational Statistics*, 29(5), 1095–1128.
- Molle, J. W. D., Hinich, M. J., Morrice, D. J. (1992). Higher-order cumulant spectral-based statistical tests of pseudo-random variate generators. W: *Proceedings of the 24th Conference on Winter Simulation* (s. 618–625). New York: ACM Press.
- Muller, M. E. (1958). An inverse method for the generation of random normal deviates on large-scale computers. *Mathematical Tables and Other Aids to Computation*, 12(63), 167–174.
- Muller, M. E. (1959). A comparison of methods for generating normal deviates on digital computers. *Journal of the ACM*, 6(3), 376–383.
- Ohirko, I., Zaniewski, I., Ogirko, O. (2016). Modelowanie i symulacja w naukach ekonomicznych. *Autobusy: technika, eksploatacja, systemy transportowe*, 17(6), 1742–1747.
- Razali, N. M., Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Rényi, A. (1953). On the theory of order statistics. *Acta Mathematica Academiae Scientiarum Hungaricae* (4), 191–232.
- Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J., Vo, S. (2001). *A statistical test suite for random and pseudorandom number generators for cryptographic applications*. Gaithersburg: National Institute of Standards and Technology (NIST).
- Schollmeyer, M. F., Tranter, W. H. (1991). Noise generators for the simulation of digital communication systems. W: *Proceedings of the 24th Annual Symposium on Simulation* (s. 264–275). Los Alamitos: IEEE Computer Society Press.
- Shapiro, S. S., Wilk, M. B. (1965). An Analysis of Variance Test for Normality. *Biometrika*, 52(3/4), 591–611.
- Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*, (19), 279–281.
- Sulewski, P. (2000). Test trendu i sezonowości. *Wiadomości Statystyczne*, (5), 11–19.
- Thomas, D. B., Luk, W., Leong, P. H. W., Villasenor, J. D. (2007). Gaussian random number generators. *Journal ACM Computing Surveys*, 39(4), 1–38.
- Wallace, C. (2005). *Random number generators*. Technical report, Monash University. Pobrano z: <http://www.datamining.monash.edu.au/software/random/>.
- Watson, G. S. (1961). Goodness-of-Fit Tests on a Circle. *Biometrika*, 48(1/2), 109–114.

- Wetherill, G. B. (1965). An approximation to the inverse normal function suitable for the generation of random normal deviates on electronic computers. *Journal of the Royal Statistical Society Series C*, 14(2–3), 201–205.
- Wichura, M. J. (1988). Algorithm AS 241: The Percentage Points of the Normal Distribution. *Journal of the Royal Statistical Society. Series C*, 37(3), 477–484.
- Wieczorkowski, R., Zieliński, R. (1997). *Komputerowe generatory liczb losowych*. Warszawa: WNT.



## Zróźnicowanie krajów pod względem udziału krajowej wartości dodanej w eksporcie

Wirginia Doryń<sup>a</sup> 

**Streszczenie.** Powstanie globalnych łańcuchów wartości sprawiło, że pomiar korzyści z handlu zagranicznego nie może ograniczać się do pomiaru strumieni handlu czy też ich struktury. Celem artykułu jest przedstawienie zróźnicowania krajów pod względem udziału krajowej wartości dodanej w eksporcie brutto. W badaniu posłużono się analizą skupień szeregów czasowych. Dane obejmujące okres 2005–2016 zaczerpnięto z najnowszej bazy TiVA (Trade in Value Added) z grudnia 2018 r. Otrzymano cztery skupienia krajów. W skupieniu o najwyższych wartościach udziału krajowej wartości dodanej w eksporcie znalazły się gospodarki bogate w zasoby naturalne (m.in. Arabia Saudyjska, Rosja, kraje Ameryki Południowej), mające duży rynek wewnętrzny oraz zaangażowane w zaawansowane technologicznie procesy wytwórcze/usługi o wysokim nasyceniu wiedzą (m.in. USA, Japonia, Wielka Brytania). W grupie o najniższym udziale krajowej wartości dodanej w eksporcie (nieprzekraczającym 50%) znalazły się małe i otwarte gospodarki europejskie – Luksemburg i Malta.

**Słowa kluczowe:** globalne łańcuchy wartości, handel zagraniczny, krajowa wartość dodana w eksporcie, analiza skupień szeregów czasowych, baza TiVA

## Differentiation of countries in terms of the share of domestic value added in exports

**Summary.** The emergence of global value chains entails that measuring the benefits of foreign trade cannot be limited to measuring trade flows nor their structure. The article aims at presenting the results of the time series cluster analysis of the share of the domestic value added of gross exports. It is based on data from the latest TiVA database (Trade in Value Added) from December 2018, covering the period 2005–2016. Four clusters of countries were identified. The cluster containing countries with the highest values of the national share of added value in exports included economies rich in natural resources (i.e., Saudi Arabia, Russia, South American countries), with a large internal market and involved in technologically advanced manufacturing processes/knowledge intensive services (including USA, Japan, Great Britain). A group with the lowest share of the domestic value added in exports (not exceeding 50%) contained small and open European economies of Luxembourg and Malta.

**Keywords:** global value chains, foreign trade, domestic value added in exports, time series cluster analysis, TiVA database

**JEL:** F14, F15, O50

---

<sup>a</sup> Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny.

Powstanie globalnych łańcuchów wartości (global value chains – GVC) zmieniło obraz procesów globalizacyjnych, powodując, że obecnie pomiar korzyści z handlu zagranicznego nie może ograniczać się wyłącznie do oceny wielkości strumieni handlu czy też ich struktury (UNCTAD, 2013, s. 10). Dobra konsumowane na rynku docelowym zawierają wkład dostarczany przez producentów z różnych krajów, którzy z kolei zaopatrują się w innych krajach, natomiast poszczególne etapy globalnego łańcucha wartości są lokowane w warunkach (krajach) najlepiej odpowiadających potrzebom tych etapów (OECD, 2013, s. 16). Kraje o niskich i średnich dochodach od lat importowały części z krajów zaawansowanych technologicznie – zazwyczaj w celu montażu, a następnie sprzedaży na rynku krajowym (Taglioni i Winkler, 2016, s. 11). Termin *globalne łańcuchy wartości* oddaje silny trend w kierunku podziału (dyspersji) ogniw (etapów) łańcucha wartości na całym świecie (OECD, 2013, s. 14). Fragmentacja procesów produkcyjnych sprawiła, że ważne jest nie tyle to, co dany kraj eksportuje (produkt finalny), ile jego zaangażowanie w poszczególne etapy przetwarzania (OECD, 2013, s. 9).

Globalizacja procesów produkcji i związany z nią wzrost handlu dobrami pośrednimi sprawiły, że statystyki handlu zagranicznego wykorzystywane do tej pory jako wskaźniki opisujące rzeczywistość gospodarczą straciły precyzję. Wady dotychczasowego podejścia, wynikające zasadniczo z wielokrotnego liczenia wartości handlu (poprzez zliczanie wartości komponentów oraz wyrobów finalnych), skutkują utratą wiarygodności opartego na tym podejściu zobrazowania pozycji poszczególnych krajów we współczesnej gospodarce światowej; utrudnia to w szczególności podejmowanie właściwych decyzji gospodarczych i politycznych (Piotrowski, 2013). Celem artykułu jest przedstawienie zróżnicowania krajów pod względem udziału krajowej wartości dodanej w eksporcie brutto. Przyjęty wskaźnik pochodzi z bazy danych statystycznych będącej rezultatem współpracy OECD i WTO (OECD, 2019). Odzwierciedla on rzeczywisty wkład poszczególnych gospodarek w tworzenie wartości dodanej.

## POMIAR STRUMIENI HANDLU

Jak wskazują Hummels, Ishii i Yi (2001, s. 76), aby mówić o fragmentacji produkcji związanej z GVC, spełnione muszą być następujące warunki:

- dobro produkowane jest w co najmniej dwóch etapach;
- co najmniej dwa kraje tworzą wartość dodaną produkowanego dobra;
- co najmniej jeden kraj używa importowanych nakładów podczas swojego etapu produkcji, a następnie część produkcji jest eksportowana.

Z warunków tych wynika, że na wartość eksportu składa się nie tylko wartość dodana wygenerowana w kraju eksportującym towary, lecz także zagraniczna wartość dodana zawarta w nakładach pochodzących z importu. Dlatego istnienie GVC prowadzi do zawyżenia wartości strumieni eksportu – dobra pośrednie,

które wielokrotnie przekraczają granice państw, są każdorazowo wliczane do światowych obrotów handlu zagranicznego. Wielkość ta została oszacowana na ok. 28% światowego eksportu w 2010 r. (UNCTAD, 2013, s. 10).

Na całkowitą wartość dodaną w eksporcie brutto składają się następujące elementy (WTO, 2015):

- krajowa wartość dodana zawarta w finalnych/pośrednich produktach lub usługach, które są konsumowane w kraju ich importu;
- krajowa wartość dodana zawarta w pośrednich produktach lub usługach eksportowanych do kraju (krajów), który następnie przetwarza je i reeksportuje do kraju trzeciego. Zaangażowanie kraju w taką aktywność świadczy o partycypacji przedniej w globalnych łańcuchach wartości (czyli uczestnictwie w GVC „w przód”, „w dół” łańcucha – *forward GVC participation*);
- krajowa wartość dodana reimportowana w postaci nakładów i dóbr pośrednich, które uprzednio zostały wyeksportowane, a teraz powracają do kraju macierzystego w formie innych dóbr pośrednich zużywanych w produkcji na eksport;
- zagraniczna wartość dodana nakładów, które zostały importowane w celu wytworzenia dóbr pośrednich bądź finalnych przeznaczonych na eksport. Udział kraju importującego w takiej wymianie oznacza uczestnictwo w GVC „w tył” (a więc partycypację tylną, w górę łańcucha – *backward GVC participation*).

W krajowej wartości dodanej eksportu brutto można wyodrębnić następujące składniki (Taglioni i Winkler, 2016, s. 74):

- bezpośrednią wartość dodaną eksportu, która jest różnicą między eksportem brutto a wartością nakładów pochodzenia krajowego, zagranicznego oraz z reimportu;
- wartość dodaną zawartą w nakładach pochodzenia krajowego wchodzących w skład eksportu (tzw. pośrednia krajowa wartość dodana eksportu);
- krajową wartość dodaną obecną w nakładach z reimportu.

W praktyce największe znaczenie w krajowej wartości dodanej eksportu mają składniki bezpośredni i pośredni, najmniejsze zaś nakłady z reimportu (Taglioni i Winkler, 2016, s. 74).

## KONSEKWENCJE UCZESTNICTWA W GVC

Globalne łańcuchy wartości pogłębiły proces globalizacji w wymiarach geograficznym (poprzez zaangażowanie większej liczby krajów, w tym gospodarek wschodzących) oraz sektorowym i funkcjonalnym (poprzez oddziaływanie na przemysł i usługi, nie tylko na etapie produkcji i dystrybucji, lecz także prowadzenia prac badawczo-rozwojowych) (OECD, 2013, s. 14). Fragmentacja łańcuchów wartości doprowadziła na ogół do ich wydłużenia i zwiększyła jednocześnie uzależnienie krajów od zagranicznych nakładów. Należy podkreślić, że do-

datkowy import warunkuje niejako wzrost krajowej wartości dodanej (Taglioni i Winkler, 2016, s. 17).

Udział w GVC może stwarzać zagrożenia, a wręcz okazać się pułapką podziału na centrum i peryferia (*core – periphery*). Wskazuje na to malejący udział krajowej wartości dodanej w eksporcie krajów wschodzących (Taglioni i Winkler, 2016, s. 38). Głównym argumentem podnoszonym przeciwko GVC jest minimalizacja kosztów pracy oraz związane z nią głodowe pensje (*poverty wages*) i ciężkie warunki pracy (Selwyn, 2018). Uczestnictwo w GVC tworzy podział na gospodarki centrali (*headquarter economies*), które organizują sieci produkcyjne, oraz gospodarki fabryczne (*factory economies*), które zapewniają siłę roboczą (Baldwin i Lopez-Gonzalez, 2015, s. 15). Co więcej, nie cała krajowa wartość dodana jest zatrzymywana w kraju – nadwyżka operacyjna filii zagranicznych może być transferowana do kraju macierzystego spółki matki. Podział wartości dodanej w ramach sieci produkcyjnej korporacji o zasięgu ponadnarodowym zależy także od ustalonych cen transferowych (UNCTAD, 2013, s. 179).

Z drugiej strony skutek uczestnictwa w GVC kraje rozwijające się mają łatwiejszy dostęp do nowych technologii dzięki powiązaniom lokalnych dostawców z firmami zagranicznymi. Fragmentacja procesów produkcji ułatwiła również dostęp do rynków globalnych. Wykonywanie poszczególnych zadań w procesach produkcyjnych jest mniej skomplikowane i wymaga mniejszych zasobów niż stworzenie całej branży. Niezbędne do produkcji nakłady mogą pochodzić z importu – nie ma konieczności ich lokalnej produkcji (Rodrik, 2018, s. 2).

Wysokie wartości krajowej wartości dodanej w eksporcie są pożądane, ponieważ przekładają się na wartość dodaną całej gospodarki (Myszkowska, 2016, s. 22). Poszczególne składniki wartości dodanej, w tym wynagrodzenia i zyski przedsiębiorstw, powstają w kraju i są w rachunkach narodowych zaliczane do produktu krajowego brutto (Ambroziak, 2018, s. 20).

## DANE I METODA BADAWCZA

Dane wykorzystane w badaniu pochodzą z najnowszej edycji bazy danych TiVA (Trade in Value Added) z grudnia 2018 r. (OECD, 2019). Obejmują one 64 kraje, w tym wszystkie kraje członkowskie OECD, Unię Europejską, grupę G20, większość krajów Azji Wschodniej i Południowo-Wschodniej oraz część krajów Ameryki Południowej. Dane z ostatniej edycji dotyczą okresu 2005–2016. W celu zbadania i oceny poziomu oraz zmian zachodzących w krajowej wartości dodanej eksportu w badanym okresie<sup>1</sup> przeprowadzono analizę sku-

---

<sup>1</sup> Brak porównywalności z danymi z wcześniejszych lat wynika ze zmian w międzynarodowych tablicach przepływów międzygałęziowych związanych z przejściem z międzynarodowego systemu rachunków narodowych (SNA93) na nowy system SNA2008 (OECD, 2018b).

pień szeregów czasowych (*time series cluster analysis*). Przedmiotem badania była krajowa wartość dodana eksportu w relacji do eksportu brutto, co pozwoliło na zachowanie porównywalności pomiędzy poszczególnymi gospodarstwami<sup>2</sup>.

Analiza skupień szeregów czasowych jest metodą badawczą szczególnie intensywnie rozwijaną w ostatnich latach (Montero i Vilar, 2014, s. 1 i 2). Jej istota polega na pogrupowaniu badanych obiektów (w tym przypadku krajów) na zbiory homogenicznie z punktu widzenia analizowanej cechy bądź zbioru cech, dlatego kluczowy do przeprowadzenia takiej analizy jest dobór adekwatnej miary odległości (Montero i Vilar, 2014, s. 3). Po wyznaczeniu macierzy odległości dla szeregów czasowych można użyć konwencjonalnego algorytmu w celu otrzymania skupisk.

Ogół miar odległości odnoszących się do danych czasowych można podzielić na cztery kategorie (Esling i Agon, 2012, s. 17):

- miary oparte na kształcie, które są obliczane dla surowych szeregów czasowych o jednakowej długości. Do tej kategorii wlicza się w szczególności odległość Minkowskiego, która dla szeregów czasowych  $X$  i  $Y$  wyraża się wzorem:

$$d(X, Y) = \left( \sum_{i=1}^T |x_i - y_i|^p \right)^{1/p} \quad (1)$$

Formuła ta dla  $p = 1$  sprowadza się do odległości miejskiej, dla  $p = 2$  – do miary euklidesowej, natomiast dla  $p \rightarrow \infty$  – do odległości Czebyszewa (Waleśiak, 1993, s. 42). Powyższe miary zaczerpnięte są z tradycyjnej analizy skupień. Do tej grupy odległości wlicza się także miarę DTW (*Dynamic Time Wrapping*), która pozwala na wyznaczenie odległości pomiędzy szeregami czasowymi o różnej długości, a przy tym jest odporna na zniekształcenia, takie jak przesunięcie (*shifting*) i skalowanie (*scaling*) (Cassisi, Montalto, Aliotta, Cannata i Pulvirenti, 2012, s. 77);

- miary oparte na edycji, dokonujące pomiaru odległości pomiędzy szeregami na podstawie minimalnej liczby operacji potrzebnych do przekształcenia jednego szeregu w inny (Esling i Agon, 2012, s. 17);
- miary wyznaczone na podstawie charakterystyk danych pierwotnych, np. miar autokorelacji, korelacji krzyżowych czy współczynników falkowych (Montero i Vilar, 2014, s. 2);
- miary podobieństwa struktur, wyznaczone na podstawie podobieństwa modeli dopasowanych do danych (*model-based distances*) oraz na podstawie kompresji danych (*compression-based distances*) (Esling i Agon, 2012, s. 17).

<sup>2</sup> Z uwagi na specyfikę analizowanych danych podjęto decyzję o przeprowadzeniu badania na danych surowych, w szczególności zachowując ich absolutne różnice w poziomach oraz zakresach zmienności.

W niniejszym badaniu wybrano dwie metody wyznaczenia odległości:

- odległość euklidesową, która bazuje na określeniu bliskości odpowiadających sobie w czasie wartości (poszczególne pary odpowiadających sobie w czasie wartości traktowane są jako niezależne obserwacje) i dlatego jest wrażliwa na przesuwanie lub skalowanie w czasie, czyli rozciąganie lub skracanie osi czasu (Montero i Vilar, 2014, s. 4 i 5);
- odległość wykorzystującą współczynnik CORT (dalej: odległość CORT) (Chouakria i Nagabhushan, 2007), wykorzystującą nie tylko miarę bliskości, lecz także miarę czasowej korelacji pomiędzy badanymi wartościami (por. Beręsewicz i Szymkowiak, 2016):

$$CORT(X, Y) = \frac{\sum_{i=1}^{T-1} (x_{i+1} - x_i)(y_{i+1} - y_i)}{\sqrt{\sum_{i=1}^{T-1} (x_{i+1} - x_i)^2} \sqrt{\sum_{i=1}^{T-1} (y_{i+1} - y_i)^2}} \quad (2)$$

Współczynnik ten przyjmuje wartości z zakresu  $\langle -1; 1 \rangle$ , gdzie 1 wskazuje, że szeregi  $X$  i  $Y$  odnotowują takie same zmiany – zwiększają się lub zmniejszają jednocześnie o tę samą wartość, zaś wskaźnik równy  $-1$  oznacza zmiany takie same pod względem wielkości, ale zachodzące w przeciwnych kierunkach ( $X$  rośnie,  $Y$  maleje lub na odwrót). Wartość równa 0 wskazuje na brak zależności (Chouakria i Nagabhushan, 2007, s. 10).

Dla wyznaczenia odległości wykorzystano miarę DTW, która zastępuje porównanie punktowe szeregów czasowych, według poszczególnych momentów w czasie  $i$ , porównaniem parami wszystkich elementów szeregów. Procedura ta pozwala na wyznaczenie optymalnej, tzn. najmniejszej, odległości pomiędzy szeregami w sensie najlepszego dopasowania poszczególnych lokalnych wartości z porównywanych szeregów (Saeed i Adamski, 2005). Miara DTW staje się szczególnie przydatna w sytuacji, gdy porównywane szeregi są przesunięte w czasie (Beręsewicz i Szymkowiak, 2016, s. 34).

Jako metodę grupowania zastosowano metodę Warda, która tworzy wyraźnie oddzielone skupiska (por. Augustyński i Laskoś-Grabowski, 2018).

Jakość grupowania zweryfikowano na podstawie wartości trzech najczęściej używanych wskaźników (por. Basiura i Czapkiewicz, 2014): *indeksu Calińskiego-Harabasz* ( $CH$ ) (Caliński i Harabasz, 1974), *indeksu Daviesa-Bouldina* ( $DB$ ) (Davies i Bouldin, 1979) i *indeksu sylwetkowego* (*silhouette index* –  $Sil$ ) (Rousseeuw, 1987). Należą one do wskaźników wewnętrznych, czyli wykorzystujących do oceny jakości grupowania informacje pochodzące wyłącznie z analizowanych danych. Wyższe wartości indeksów *Calińskiego-Harabasz* oraz *sylwetkowego*, jak również niższe wartości w przypadku *indeksu Daviesa-Bouldina* przekładają się na wyższą jakość klasyfikacji.

Do obliczeń posłużył pakiet *dtwclust* (Sarda-Espinosa, 2018).

## WYNIKI BADANIA

Najpierw przeprowadzono analizę skupień z wykorzystaniem miary odległości Euklidesa. *Indeks sylwetkowy* przyjął w tym przypadku wartość największą, wynoszącą 0,57, co świadczy o poważnej strukturze klas (Kaufman i Rousseeuw, 1990; za: Gatnar i Walesiak, 2004, s. 342) w przypadku podziału na cztery skupienia. *Indeks Calińskiego-Harabasa* wskazał natomiast na trzy grupy, a *wskaźnik DB* – na cztery grupy (por. tabl. 1, s. 39). Klasyfikację krajów prezentując zestawienie (część A, s. 40) oraz wyk. 1A (s. 41), a w tabl. 2 (s. 40) przedstawiono wybrane statystyki opisowe poszczególnych skupień<sup>3</sup>.

Skupienie 1 obejmuje kraje o najwyższych wartościach udziału krajowej wartości dodanej w eksporcie. Znalazły się w nim kraje posiadające bogate zasoby naturalne (np. Australia, Rosja, kraje Ameryki Południowej), jak również gospodarki cechujące się dużym rynkiem wewnętrznym dóbr pośrednich i zaangażowaniem w zaawansowane technologicznie procesy (Stany Zjednoczone, Japonia) (Kowalski, Gonzalez, Ragoussis i Ugarte, 2015, s. 15). Z krajów UE zakwalifikowała się jedynie Wielka Brytania, w przypadku której większość krajowej wartości dodanej eksportu pochodzi z sektora usług, w tym finansowych (OECD, 2018a). Udział krajowej wartości dodanej w eksporcie tych krajów sięgał średnio 88,1% w latach 2005–2016, przy czym wartość najniższa wynosiła 79,6% (Turcja, dotyczy 2012 r.), a najwyższa – 97% (Arabia Saudyjska, dotyczy 2014 r.). Należy zaznaczyć, że zasadniczo im większy kraj, tym mniejszy udział zagranicznej wartości dodanej w eksporcie (Ahmad, 2013, s. 90).

Skupienie 2 to 29 krajów, w tym większość krajów UE-15, Polska wraz z Chorwacją, Cyprzem, Łotwą i Rumunią spośród gospodarek UE-13, a także Indie i Chiny – te ostatnie odnotowały znaczący wzrost udziału krajowej wartości dodanej w eksporcie z 73,7% w 2005 r. do 83,4% w 2015 r.<sup>4</sup> Udział krajowej wartości dodanej w eksporcie w tych krajach wahał się w badanym okresie w granicach 67–83%, przy średnim wskaźniku wynoszącym 75,7%, co można określić jako średnio wysoki poziom.

Skupienie 3 stanowią gospodarki, dla których średnia wartość badanego wskaźnika wynosiła 62,1%. Znalazły się w nim m.in.: większość krajów UE-13, Belgia i Irlandia z UE-15, Meksyk oraz część krajów należących do tzw. Fabryki Azja<sup>5</sup> (Baldwin, 2006): Korea Południowa, Malezja, Singapur, Tajlandia, Tajwan

<sup>3</sup> Poziom odcięcia grup wyniósł ok. 22% najdłuższego wiązania.

<sup>4</sup> W edycji TiVA z 2016 r. udział krajowej wartości dodanej w chińskim eksporcie wyniósł 61,3–69,3%.

<sup>5</sup> Terminu *Fabryka Azja* używa się w celu zobrazowania modelu powiązań produkcyjnych krajów azjatyckich produkujących i przetwarzających dobra pośrednie w ramach regionalnych sieci produkcji, z produktem finalnym kierowanym do krajów rozwiniętych poprzez globalną sieć dystrybucji (Asian Development Bank, 2013, s. 2).

i Wietnam. Udział krajowej wartości dodanej w eksporcie wahał się w tej grupie od 52,2% (Węgry, dotyczy 2011 r.) do 74,8% (Litwa, dotyczy 2007 r.).

W skupieniu 2 i 3 znalazły się kraje, które są w wysokim stopniu włączone w GVC przede wszystkim z powodu niskich kosztów pracy i które mają silne powiązania z przemysłem krajów rozwiniętych. Kraje te odnotowały znaczący wzrost udziału krajowej wartości dodanej w eksporcie w 2009 r. w następstwie kryzysu gospodarczego (por. Folfas, 2016). Niższy w porównaniu ze skupieniem 1 udział krajowej wartości dodanej w eksporcie jest również typowy dla gospodarek, w których handel zagraniczny cechuje się wysokim stopniem liberalizacji, uczestniczących w wielostronnych porozumieniach handlowych oraz goszczących bezpośrednio inwestycje zagraniczne (Mroczek, 2015, s. 3).

Grupa 4 objęła natomiast dwie małe otwarte gospodarki europejskie: Luksemburg i Maltę, o dużym imporcie dóbr pośrednich oraz udziale krajowej wartości dodanej w eksporcie wynoszącym średnio 38,4%.

Jako alternatywną metodę wyznaczenia macierzy odległości wykorzystano miarę opartą na współczynniku korelacji oraz mierze DTW (CORT). *Indeks sylwetkowy* przyjął tutaj najwyższą wartość wynoszącą 0,67, co świadczy o poważnej strukturze klas (Kaufman i Rousseeuw, 1990; za: Gatnar i Walesiak, 2004, s. 342) w przypadku podziału na cztery grupy. Wartość *indeksu Calińskiego-Harabasza* uzasadniała wyróżnienie siedmiu skupień, natomiast *wskaźnika DB* – podział na pięć grup. Aby porównać nowy podział z klasyfikacją przeprowadzoną z wykorzystaniem odległości euklidesowej, zdecydowano o podziale na cztery grupy<sup>6</sup>. Wyniki grupowania zawiera wykr. 1B (s. 41). Udział krajowej wartości dodanej w eksporcie krajów należących do poszczególnych skupień przedstawiono w tabl. 3 (s. 42).

**TABL. 1. WSKAŹNIKI JAKOŚCI GRUPOWANIA**

Wskaźniki a – odległość euklidesowa b – odległość CORT		Liczba analizowanych skupień								
		k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
<i>Sil</i> .....	a	0,557	0,539	0,575	0,450	0,445	0,377	0,361	0,316	0,313
	b	0,582	0,618	0,671	0,548	0,488	0,458	0,475	0,472	0,408
<i>CH</i> .....	a	51,46	52,01	44,26	40,84	37,84	32,60	28,82	26,81	24,80
	b	43,72	49,88	38,29	46,39	46,05	52,45	44,44	39,26	35,90
<i>DB</i> .....	a	0,833	0,650	0,490	0,638	0,749	0,787	0,875	1,227	1,240
	b	0,912	0,696	0,615	0,507	0,702	0,680	0,831	0,748	0,837
<i>DB*</i> .....	a	0,833	0,781	0,529	0,887	0,850	0,924	0,999	1,414	1,431
	b	0,912	1,057	0,725	0,688	1,189	1,182	1,543	1,413	1,776

U w a g a. *Sil* – indeks sylwetkowy, *CH* – indeks Calińskiego-Harabasza, *DB* – indeks Daviesa-Bouldina, *DB\** – według modyfikacji Kima i Ramakrishna (2005).

Ź r ó d ł o: opracowanie własne na podstawie: OECD (2019).

<sup>6</sup> Poziom odcięcia grup wyniósł ok. 20% najdłuższego wiązania.



**ZESTAWIENIE. PRZYNALEŻNOŚĆ KRAJÓW DO SKUPIEŃ**

Skupienie	Kraje
<b>A. Odległość euklidesowa</b>	
1	Arabia Saudyjska, Argentyna, Australia, Brazylia, Brunei, Chile, Indonezja, Japonia, Kazachstan, Kolumbia, Norwegia, Nowa Zelandia, Peru, Rosja, Stany Zjednoczone, Turcja, Wielka Brytania
2	Austria, Chiny, Chorwacja, Cypr, Dania, Filipiny, Finlandia, Francja, Grecja, Hiszpania, Holandia, Hongkong, Indie, Islandia, Izrael, Kambodża, Kanada, Kostaryka, Łotwa, Maroko, Niemcy, Polska, Portugalia, Republika Południowej Afryki, Rumunia, Szwajcaria, Szwecja, Tunezja, Włochy
3	Belgia, Bułgaria, Czechy, Estonia, Irlandia, Korea Południowa, Litwa, Malezja, Meksyk, Singapur, Słowacja, Słowenia, Tajlandia, Tajwan, Węgry, Wietnam
4	Luksemburg, Malta
<b>B. Odległość CORT</b>	
1	Arabia Saudyjska, Argentyna, Australia, Brazylia, Brunei, Chile, Indonezja, Japonia, Kazachstan, Kolumbia, Norwegia, Nowa Zelandia, Peru, Rosja, Stany Zjednoczone, Turcja, Wielka Brytania
2	Austria, Chiny, Chorwacja, Cypr, Dania, Filipiny, Finlandia, Francja, Grecja, Hiszpania, Holandia, Hongkong, Indie, Islandia, Izrael, Kambodża, Kanada, Kostaryka, Litwa, Łotwa, Maroko, Niemcy, Polska, Portugalia, Republika Południowej Afryki, Rumunia, Szwajcaria, Szwecja, Tunezja, Włochy
3	Belgia, Bułgaria, Czechy, Estonia, Irlandia, Korea Południowa, Malezja, Meksyk, Singapur, Słowacja, Słowenia, Tajlandia, Tajwan, Węgry, Wietnam
4	Luksemburg, Malta

Źródło: jak przy tabl. 1.

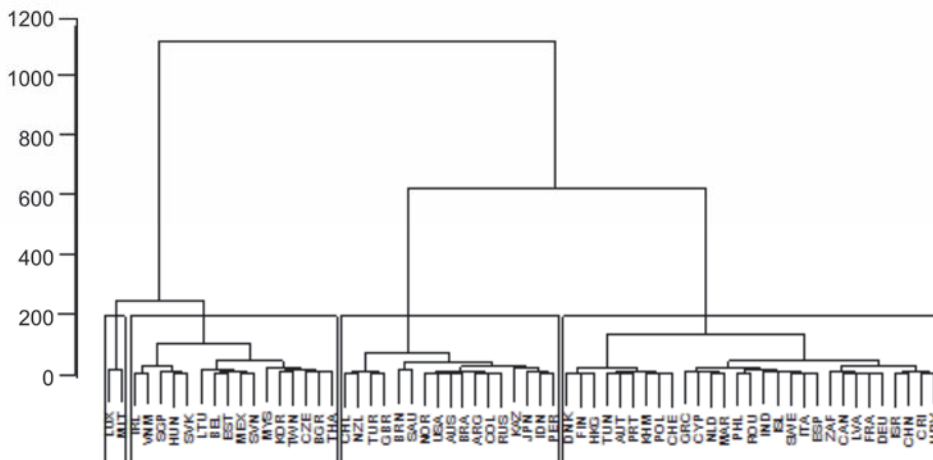
**TABL. 2. WYBRANE MIARY STATYSTYCZNE DLA SKUPIEŃ**

Miary a – odległość euklidesowa b – odległość CORT	Skupienie			
	1	2	3	4
Minimum .....	a 79,6	67,3	52,2	31,2
	b 79,6	64,5	52,2	31,2
Maksimum .....	a 97,0	84,0	74,8	49,1
	b 97,0	84,0	72,2	49,1
Średnia .....	a 88,1	75,7	62,1	38,4
	b 88,1	75,5	61,7	38,4
Odchylenie standardowe .....	a 3,7	3,5	4,8	4,3
	b 3,7	3,8	4,6	4,3
Kwartył pierwszy .....	a 85,3	73,1	58,5	35,7
	b 85,3	72,9	58,2	35,7
Mediana .....	a 88,4	75,9	62,7	38,4
	b 88,4	75,7	62,1	38,4
Kwartył trzeci .....	a 90,3	78,3	65,8	40,8
	b 90,3	78,2	65,5	40,8

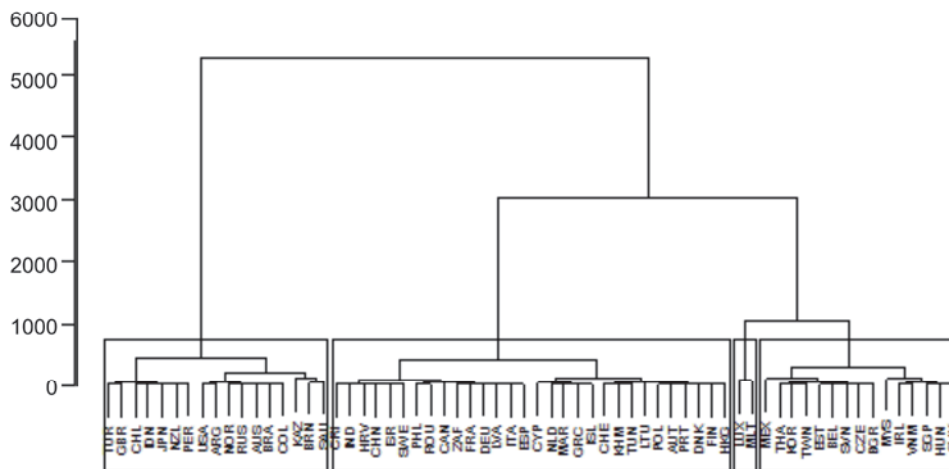
Źródło: jak przy tabl. 1.

**WYKR. 1. DENDROGRAM GRUPOWANIA KRAJÓW METODĄ WARDA**

**A. Odległość euklidesowa**



**B. Odległość CORT**



Źródło: jak przy tabl. 1.

U w a g a. ARG – Argentyna, AUS – Australia, AUT – Austria, BEL – Belgia, BGR – Bułgaria, BRA – Brazylia, BRN – Brunei, CAN – Kanada, CHE – Szwajcaria, CHL – Chile, CHN – Chiny, COL – Kolumbia, CRI – Kostaryka, CYP – Cypr, CZE – Czechy, DEU – Niemcy, DNK – Dania, ESP – Hiszpania, EST – Estonia, FIN – Finlandia, FRA – Francja, GBR – Wielka Brytania, GRC – Grecja, HGK – Hongkong, HUN – Węgry, HVR – Chorwacja, IDN – Indonezja, IND – Indie, IRL – Irlandia, ISL – Islandia, ISR – Izrael, ITA – Włochy, JPN – Japonia, KAZ – Kazachstan, KHM – Kambodża, KOR – Korea Południowa, LTU – Litwa, LUX – Luksemburg, LVA – Łotwa, MAR – Maroko, MEX – Meksyk, MLT – Malta, MYS – Malezja, NLD – Holandia, NOR – Norwegia, NZL – Nowa Zelandia, PER – Peru, PHL – Filipiny, POL – Polska, PRT – Portugalia, ROU – Rumunia, RUS – Rosja, SAU – Arabia Saudyjska, SGP – Singapur, SVK – Słowacja, SVN – Słowenia, SWE – Szwecja, THA – Tajlandia, TUN – Tunezja, TUR – Turcja, TWN – Tajwan, USA – Stany Zjednoczone, VNM – Wietnam, ZAF – Republika Południowej Afryki.

TABL. 3. UDZIAŁ KRAJOWEJ WARTOŚCI DODANEJ W EKSPORCIE

Kraje	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
	w %											
<b>Skupienie 1</b>												
Arabia Saudyjska .....	95,9	94,7	95,0	96,8	96,4	96,7	96,6	96,9	96,5	97,0	95,4	95,1
Argentyna .....	89,0	88,6	88,3	88,2	91,0	89,4	88,8	90,4	90,1	90,9	93,1	93,1
Australia .....	88,9	87,9	88,4	88,1	89,0	89,7	89,0	88,8	89,2	88,7	88,4	90,0
Brazylia .....	88,8	89,1	89,6	89,0	91,7	90,4	90,3	89,4	88,7	88,5	87,5	89,8
Brunei .....	92,5	93,0	94,2	92,7	92,5	92,5	91,0	88,7	87,5	93,2	94,6	92,3
Chile .....	82,3	84,9	83,9	80,6	84,6	86,2	84,2	84,5	85,1	85,0	84,9	87,6
Kolumbia .....	89,5	88,9	89,0	89,5	91,3	92,0	92,4	92,3	92,2	90,6	88,4	89,9
Indonezja .....	81,6	85,1	85,7	85,0	87,9	87,6	87,2	86,7	86,1	85,9	87,1	88,7
Japonia .....	89,8	87,8	86,8	84,8	89,1	87,8	85,7	86,1	84,8	84,2	86,8	88,6
Kazachstan .....	79,7	81,7	80,2	84,1	85,6	90,4	91,7	90,3	91,2	92,9	93,5	92,3
Norwegia .....	88,2	88,4	87,1	88,7	88,4	87,7	88,8	89,3	89,0	87,7	86,1	86,3
Nowa Zelandia .....	85,1	84,2	85,5	82,9	85,8	85,1	84,3	83,6	85,1	85,3	86,2	87,2
Peru .....	86,7	86,4	86,7	85,0	87,6	87,7	87,9	87,3	87,7	87,7	89,7	90,2
Rosja .....	90,1	90,7	90,7	89,6	89,5	90,3	90,6	91,3	91,1	90,6	89,2	89,8
Stany Zjednoczone ....	89,2	88,6	88,3	87,1	90,6	89,0	87,3	87,6	88,5	88,8	90,5	91,0
Turcja .....	84,6	82,5	81,9	80,4	84,3	82,8	80,6	79,6	80,9	81,8	83,2	83,5
Wielka Brytania .....	85,7	85,0	85,1	83,4	84,1	82,6	81,2	81,2	82,0	83,7	84,9	84,6
<b>Skupienie 2</b>												
Austria .....	74,5	73,5	73,4	72,8	76,7	72,0	70,3	70,1	70,5	71,3	73,5	73,4
Chiny .....	73,7	74,1	75,2	77,1	80,5	78,9	78,3	79,2	79,7	80,5	82,7	83,4
Chorwacja .....	77,7	77,6	77,3	78,2	81,0	79,3	80,6	81,2	80,3	81,0	80,0	80,8
Cypr .....	76,4	75,9	75,3	72,4	74,3	74,6	75,9	76,0	74,1	72,8	72,2	76,4
Dania .....	73,9	71,3	69,7	68,5	72,3	72,0	70,5	69,9	70,5	70,9	70,7	71,9
Filipiny .....	73,7	68,3	75,8	75,2	78,1	76,1	76,5	76,1	79,0	79,6	78,0	76,6
Finlandia .....	72,5	70,1	70,0	68,4	72,2	70,6	68,6	67,9	69,1	70,1	74,1	74,1
Francja .....	79,6	78,0	78,1	77,8	81,1	77,9	76,6	76,8	77,3	77,8	78,6	77,9
Grecja .....	80,9	78,5	76,9	79,8	83,1	77,0	74,7	70,2	70,8	71,7	75,5	78,5
Hiszpania .....	77,0	75,8	75,1	75,6	81,1	78,1	75,7	75,0	76,0	76,3	77,3	78,4
Holandia .....	77,1	77,2	76,5	76,7	78,7	76,4	73,6	73,4	73,7	73,3	72,1	73,0
Hongkong .....	72,4	70,5	69,4	67,3	70,9	68,4	67,6	68,1	69,0	70,8	73,4	75,2
Indie .....	81,2	79,4	79,3	75,5	78,2	76,4	74,9	74,9	75,3	77,1	80,9	83,9
Islandia .....	76,6	75,3	77,2	77,2	76,9	75,3	73,4	72,0	73,4	74,6	76,3	79,7
Izrael .....	74,7	73,2	76,4	75,1	78,4	77,7	79,0	79,2	81,1	80,0	81,4	82,8
Kambodża .....	70,5	71,0	71,1	72,4	74,5	73,4	73,7	72,6	73,1	72,3	73,1	71,4
Kanada .....	80,4	80,5	80,8	80,4	79,9	79,3	79,1	78,3	79,3	79,9	78,8	79,4
Kostaryka .....	77,1	77,5	78,4	76,8	81,0	80,9	80,4	80,0	81,0	81,0	83,8	84,0
Litwa .....	70,5	70,2	74,8	66,5	73,2	67,2	64,5	65,2	64,5	67,4	68,4	70,6
Łotwa .....	78,6	76,6	76,6	79,6	81,3	78,2	76,8	75,2	76,0	77,8	77,6	79,3
Maroko .....	77,3	77,8	75,2	74,5	78,6	77,0	73,8	71,5	72,7	73,1	74,3	74,7
Niemcy .....	81,4	79,6	78,9	78,6	81,9	78,5	76,8	76,9	77,6	78,3	79,0	79,7
Polska .....	75,3	72,8	72,4	72,2	75,6	73,1	71,6	72,7	72,7	72,5	73,4	73,1

**TABL. 3. UDZIAŁ KRAJOWEJ WARTOŚCI DODANEJ W EKSPORCIE (dok.)**

Kraje	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
	w %											

**Skupienie 2 (dok.)**

Portugalia .....	73,6	71,8	72,9	71,6	76,3	73,1	70,4	69,9	70,1	70,5	71,6	72,0
RPA .....	82,2	80,3	79,4	75,7	81,1	81,3	79,2	77,3	75,8	75,7	77,4	77,5
Rumunia .....	72,4	72,7	74,3	77,3	80,2	78,9	76,3	75,7	76,8	76,4	77,1	78,4
Szwajcaria .....	74,2	73,8	73,4	73,0	74,1	74,1	74,1	73,7	72,5	74,7	75,4	74,2
Szwecja .....	75,4	74,4	74,2	73,0	77,3	75,7	75,5	75,2	77,5	77,3	79,3	80,3
Tunezja .....	74,6	74,9	72,0	70,9	74,9	73,4	73,5	69,6	69,4	69,8	71,5	70,1
Włochy .....	79,5	77,3	76,9	76,1	80,2	76,0	74,5	75,3	76,6	77,3	77,8	78,0

**Skupienie 3**

Belgia .....	69,4	68,0	67,3	64,7	69,2	66,5	63,5	63,0	63,9	64,6	65,9	66,2
Bułgaria .....	67,6	62,9	60,0	59,3	68,3	65,9	62,8	62,1	62,0	62,7	63,8	67,8
Czechy .....	65,6	64,4	63,7	64,5	66,7	62,8	61,3	60,7	61,1	60,3	60,7	62,3
Estonia .....	69,6	69,7	69,2	68,2	72,2	66,1	61,6	61,4	62,1	63,1	65,2	65,5
Irlandia .....	64,3	62,9	62,7	59,8	60,9	59,3	59,9	57,2	56,8	55,3	59,8	58,3
Korea Południowa .....	67,3	65,8	65,6	58,8	62,9	61,8	57,6	58,0	60,5	62,7	67,4	69,6
Malezja .....	55,0	56,6	55,9	59,7	60,2	59,4	60,6	61,9	62,8	63,4	63,1	63,9
Meksyk .....	66,0	65,8	66,6	66,7	66,4	66,1	67,7	66,2	65,4	65,9	63,9	63,6
Singapur .....	57,2	55,4	58,6	54,8	58,0	58,7	56,5	56,2	57,2	57,0	59,1	60,6
Słowacja .....	57,0	53,5	53,9	55,0	58,1	56,1	53,3	53,4	53,2	54,1	55,2	55,5
Słowenia .....	66,7	65,0	64,0	64,8	69,4	66,1	64,5	64,9	65,8	66,1	67,5	68,5
Tajlandia .....	61,6	62,9	63,9	61,0	65,6	64,0	61,2	61,6	62,5	63,3	66,4	67,5
Tajwan .....	62,9	60,7	59,4	56,3	63,7	58,5	57,9	58,8	60,5	61,7	67,6	70,2
Węgry .....	56,0	53,8	54,1	53,3	56,5	52,5	52,2	53,0	53,9	54,0	56,9	55,9
Wietnam .....	63,9	61,9	59,2	58,5	62,8	59,5	58,2	59,2	58,3	57,6	55,5	56,4

**Skupienie 4**

Luksemburg .....	41,8	40,2	38,2	38,1	41,0	39,6	39,4	37,1	35,9	33,2	31,2	32,6
Malta .....	49,1	44,8	44,8	39,7	37,7	35,7	32,1	34,3	35,7	38,5	40,9	40,7

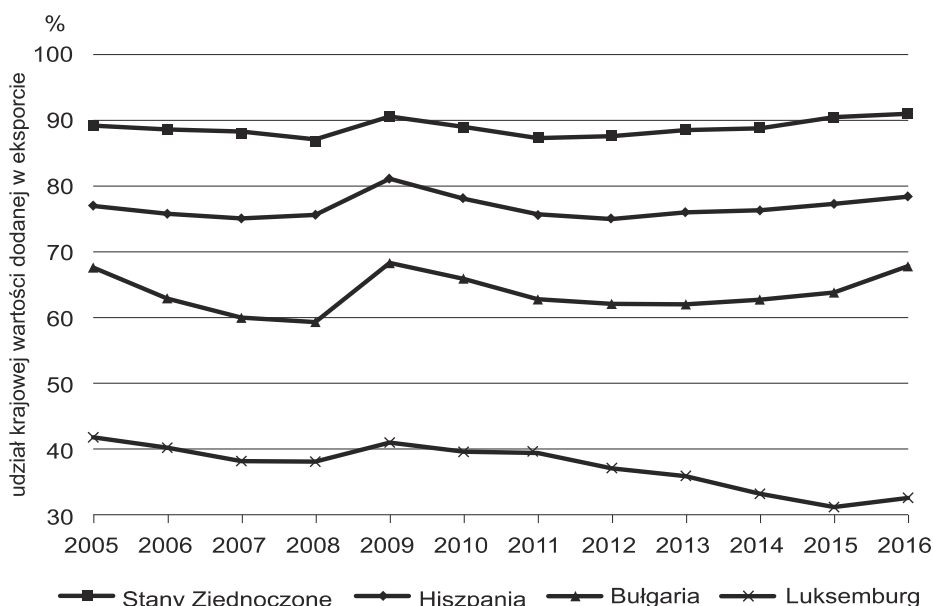
Źródło: jak przy tabl. 1.

Analiza składu otrzymanych skupień prowadzi do wniosku, że wyniki obu podziałów – według odległości euklidesowej oraz CORT – są bardzo zbliżone. Jedynym krajem zaklasyfikowanym do różnych skupień w zależności od zastosowanej metody była Litwa, która w przypadku grupowania z wykorzystaniem odległości euklidesowej znalazła się w skupieniu 3, a po przyjęciu miary CORT – w skupieniu 2. Na podstawie charakterystyki tej odległości oraz analizy składu poszczególnych skupień można wywnioskować, że wyjątek ten może wynikać z silnego powiązania gospodarki litewskiej m.in. z Łotwą, Niemcami i Polską i dlatego przy zastosowaniu miary CORT spójność czasowa zmian analizowa-

nego wskaźnika tych krajów zdecydowała o zaklasyfikowaniu Litwy do skupienia 2.

Podobieństwo uzyskanych wyników można zinterpretować w kontekście spójności w czasie zmian analizowanego wskaźnika. Na wyk. 2 zaprezentowano medoidy – reprezentatywne dla poszczególnych skupień szeregi wyznaczone z wykorzystaniem algorytmu PAM (*Partition around medoids*) (Kaufman i Rousseeuw, 1987), dla których przeciętna odległość względem pozostałych elementów skupienia jest najmniejsza. Były to odpowiednio Stany Zjednoczone (skupienie 1), Hiszpania (skupienie 2), Bułgaria (skupienie 3) oraz Luksemburg (skupienie 4). Na podstawie tego wykresu można wysnuć wniosek, że w badanym okresie kraje należące do poszczególnych skupień różniły się wysokością udziału krajowej wartości dodanej w eksporcie. Z kolei zmiany zachodzące w czasie można podsumować następująco: analizowany wskaźnik malał w latach przed kryzysem, następnie skokowo wzrósł, co należy prawdopodobnie wiązać z załamaniem handlu na przełomie 2008 i 2009 r. i renacjonalizacją łańcuchów dostaw (Gabrielczak i Serwach, 2010). W późniejszym okresie nastąpił spadek wskaźnika, a w ostatnich latach obserwowany jest wzrost udziału krajowej wartości dodanej w eksporcie, co z kolei może być związane z wyhamowaniem integracji w ramach GVC (Wojtas, 2017).

WYKR. 2. MEDOIDY SKUPIEŃ (odległość CORT)



## PODSUMOWANIE

W niniejszym badaniu dokonano analizy skupień szeregów czasowych w celu pogrupowania krajów według wysokości udziału krajowej wartości dodanej w eksporcie. W badaniu wykorzystano najnowsze dane pochodzące z bazy TiVA, obejmującej dane dla 64 krajów, w tym krajów członkowskich OECD, Unii Europejskiej, grupy G20, większości krajów Azji Wschodniej i Południowo-Wschodniej oraz części krajów Ameryki Południowej.

Wyodrębniono cztery skupienia krajów. W grupie 1 znalazły się kraje o najwyższych wartościach badanego wskaźnika (79,6–97,0%), w szczególności gospodarki o bogatych zasobach naturalnych (Arabia Saudyjska, Australia, Rosja, kraje Ameryki Południowej), kraje mające duży rynek wewnętrzny dóbr pośrednich i zaangażowane w zaawansowane technologicznie procesy (Stany Zjednoczone, Japonia) oraz Wielka Brytania, której sektor usług, w tym finansowych, generuje wysoki udział krajowej wartości dodanej w eksporcie.

Grupa 2 zawiera gospodarki o średnim udziale krajowej wartości dodanej w eksporcie rzędu 75%, w tym większość krajów UE-15, Polskę, Chorwację, Cypr, Łotwę i Rumunię z UE-13, a także Indie i Chiny.

Skupienie 3 objęło gospodarki, dla których analizowany wskaźnik przyjmował wartości przeciętnie nieznacznie przewyższające 60%. Znalazła się w nim większość krajów UE-13, Belgia i Irlandia z UE-15, Meksyk oraz część krajów należących do Fabryki Azja: Korea Południowa, Malezja, Singapur, Tajlandia, Tajwan i Wietnam.

Do grupy 4 trafiły natomiast małe i otwarte gospodarki Luksemburga i Malty, w przypadku których udział krajowej wartości dodanej w eksporcie nie przekraczał 50%.

Ponadto na podstawie analizy medoidów scharakteryzowano typowy przebieg zmian analizowanego wskaźnika w czasie: spadek w okresie przedkryzysowym, skokowy wzrost w 2009 r., a następnie ponowny spadek, zakończony wzrostem w ostatnich latach.

## BIBLIOGRAFIA

- Ahmad, N. (2013). Estimating trade in value-added: why and how. W: D. K. Elms, P. Low (red.), *Global value chains in a changing world* (s. 85–103). Geneva: World Trade Organization. Pobrane z: [https://www.wto.org/english/res\\_e/booksp\\_e/aid4tradeglobalvalue13\\_e.pdf](https://www.wto.org/english/res_e/booksp_e/aid4tradeglobalvalue13_e.pdf).
- Ambroziak, Ł. (2018). Zmiany w polskim eksporcie produktów przemysłu spożywczego według pochodzenia wartości dodanej. *Studia Ekonomiczne*, (352), 9–21.
- Asian Development Bank. (2013). *Beyond Factory Asia: Fuelling Growth in a Changing World*. Manila: Asian Development Bank.
- Augustyński, I., Laskoś-Grabowski, P. (2018). Clustering Macroeconomic Time Series. *Econometrics*, 22(2), 74–88. DOI: <https://doi.org/10.15611/eada.2018.2.06>.

- Basiura, B., Czapkiewicz, A. (2014). Badanie jakości klasyfikacji szeregów czasowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Taksonomia*, 22(327), 148–156.
- Baldwin, R. E. (2006). Multilateralising regionalism: spaghetti bowls as building blocs on the path to global free trade. *The World Economy*, 29(11), 1451–1518. DOI: <https://doi.org/10.1111/j.1467-9701.2006.00852.x>.
- Baldwin, R., Lopez-Gonzalez, J. (2015). Supply-chain trade: A portrait of global patterns and several testable hypotheses. *The World Economy*, 38(11), 1682–1721. DOI: <https://doi.org/10.1111/twec.12189>.
- Beręsewicz, M., Szymkowiak, M. (2016). Analiza skupień wybranych lokalnych rynków nieruchomości w Polsce z wykorzystaniem internetowych źródeł danych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Taksonomia*, 27(427), 30–39, DOI: 10.15611/pn.2016.427.03.
- Caliński, T., Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1), 1–27.
- Cassisi, C., Montalto, P., Aliotta, M., Cannata, A., Pulvirenti, A. (2012). Similarity measures and dimensionality reduction techniques for time series data mining. W: A. Karahoca (red.), *Advances in Data Mining, Knowledge Discovery and Applications* (s. 71–96). London: IntechOpen. DOI: <http://dx.doi.org/10.5772/49941>.
- Chouakria, A. D., Nagabhushan, P. N. (2007). Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification*, (1), 5–21. DOI: 10.1007/s11634-006-0004-6.
- Davies, D. L., Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2), 224–227. DOI: 10.1109/TPAMI.1979.4766909.
- Esling, P., Agon, C. (2012). Time-series data mining. *ACM Computing Surveys (CSUR)*, 45(1), 1–34. DOI: 10.1145/2379776.2379788.
- Folfas, P. (2016). *Handel międzynarodowy mierzony wartością brutto oraz wartością dodaną: analiza porównawcza*. Warszawa: Oficyna Wydawnicza SGH.
- Gabrielczak, P., Serwach, T. (2010). Załamanie światowych obrotów handlowych na przełomie 2008 i 2009 roku – przegląd potencjalnych przyczyn. *Ekonomia Międzynarodowa*, (1), 50–66.
- Gatnar, E. Walesiak, M. (red.). (2004). *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego.
- Hummels, D., Ishii, J., Yi, K. M. (2001). The nature and growth of vertical specialization in world trade. *Journal of International Economics*, 54(1), 75–96. DOI: [https://doi.org/10.1016/S0022-1996\(00\)00093-3](https://doi.org/10.1016/S0022-1996(00)00093-3).
- Kaufman, L., Rousseeuw, P. J. (1987). Clustering by means of medoids. W: Y. Dodge (red.), *Statistical Data Analysis Based on the L1 Norm and Related Methods* (s. 405–416). Amsterdam: Elsevier Science Pub.
- Kaufman, L., Rousseeuw, P. J. (1990). *Finding groups in data: an introduction to cluster analysis*. Nowy Jork: John Wiley & Sons.
- Kim, M., Ramakrishna, R. S. (2005). New indices for cluster validity assessment. *Pattern Recognition Letters*, 26(15), 2353–2363. DOI: 10.1016/j.patrec.2005.04.007.
- Kowalski, P., Gonzalez, J. L., Ragoussis, A., Ugarte, C. (2015). *Participation of developing countries in global value chains* (OECD Trade Policy Papers No. 179). Paris: OECD Publishing. DOI: <http://dx.doi.org/10.1787/5js331fw0xxn-en>.
- Montero, P., Vilar, J. A. (2014). Tsclust: An R package for time series clustering. *Journal of Statistical Software*, 62(1), 1–43. DOI: 10.18637/jss.v062.i01.
- Mroczek, W. (2015). Udział krajowej wartości dodanej w eksporcie nowych państw członkowskich. *Unia Europejska.pl*, 233(4), 3–6.

- Myszkowska, M. (2016). Zmiany w eksporcie polski w kategoriach wartości dodanej. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (448), 20–31. DOI: 10.15611/pn.2016.448.02.
- OECD. (2013). *Interconnected Economies: Benefiting from Global Value Chains*. Paris: OECD Publishing. DOI: <https://doi.org/10.1787/9789264189560-en>.
- OECD. (2018a). *Trade in Value Added: United Kingdom*. Pobrane z: <http://www.oecd.org/industry/ind/TIVA-2018-United-Kingdom.pdf>.
- OECD. (2018b). *Whats New? Differences between the 2018 and 2016 editions of TIVA indicators*. Pobrane z: <http://www.oecd.org/industry/ind/tiva-2018-differences-tiva-2016.pdf>.
- OECD. (2019). Trade in value added, OECD-WTO: *Statistics on Trade in Value Added* (database). DOI: <https://doi.org/10.1787/data-00648-en>.
- Piotrowski, J. (2013). Badania empiryczne na poziomie makro – przegląd literatury. W: E. Kaliszuk (red.), *Mierzenie wartości dodanej w handlu zagranicznym. Nowe koncepcje, metody i wyzwania* (s. 43–60). Warszawa: Instytut Badań Rynku, Konsumpcji i Koniunktur.
- Rodrik, D. (2018). *New technologies, global value chains, and developing economies* (NBER Working Paper No. 25164). Pobrane z: <https://www.nber.org/papers/w25164.pdf>.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, (20), 53–65. DOI: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Saeed, K., Adamski, M. (2005). *Klasyfikacja podpisu offline z wykorzystaniem metody DTW*. XIV Krajowa Konferencja Naukowa-KBIB. Pobrane z: <http://www.profuturo.agh.edu.pl/~saeed/arts/AdamsCzestochowa2005.pdf>.
- Sarda-Espinosa, A. (2018). *Package "dtwclust"*. Pobrane z: <http://cran.ma.imperial.ac.uk/web/packages/dtwclust/dtwclust.pdf>.
- Selwyn, B. (2018). Poverty chains and global capitalism. *Competition & Change*, (23), 71–97. DOI: <https://doi.org/10.1177/1024529418809067>.
- Taglioni, D., Winkler, D. (2016). *Making global value chains work for development*. Washington: The World Bank. DOI: 10.1596/978-1-4648-0157-0.
- UNCTAD. (2013). *World investment report 2013: Global value chains: Investment and trade for development*. New York: United Nations. DOI: <https://doi.org/10.18356/a3836fcc-en>.
- Walesiak, M. (1993). *Statystyczna analiza wielowymiarowa w badaniach marketingowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.
- Wojtas, M. (2017). Przyczyny spowolnienia światowego handlu. *Studia i Prace WNEiZ US*, 49(2), 75–87. DOI: 10.18276/sip.2017.49/2-06.
- WTO. (2015). *"Trade in Value-Added and Global Value Chains" profiles. Explanatory notes*. Pobrane z: [https://www.wto.org/english/res\\_e/statis\\_e/miwi\\_e/Explanatory\\_Notes\\_e.pdf](https://www.wto.org/english/res_e/statis_e/miwi_e/Explanatory_Notes_e.pdf).



## Metafory informacyjne

**Bogdan Stefanowicz**<sup>a</sup> 

**Streszczenie.** Termin *informacja* jest znany od wieków, wciąż jednak nie wyjaśniono w pełni jego znaczenia. W literaturze można znaleźć wiele definicji i interpretacji tego pojęcia, ale żadnej z nich nie sposób uznać ani za wyczerpującą, ani za powszechnie satysfakcjonującą. Dlatego kwestia rozumienia informacji i sposobów osiągnięcia jej sedna pozostaje nadal otwarta. Jednym z podejść do tego problemu jest posłużenie się środkami pomocniczymi, a spośród nich metaforami, które – choć nie zastąpią definicji – pozwalają pośrednio uchwycić sens omawianego terminu.

Celem artykułu jest pokazanie roli metafor w wyjaśnieniu niektórych cech i funkcji informacji przez analogię do innych pojęć, znanych z życia codziennego. Metafory tego rodzaju określa się mianem metafor informacyjnych. W artykule przytoczono kilka ich przykładów.

**Słowa kluczowe:** informacja, metafory informacyjne, funkcje informacji

## Information metaphors

**Summary.** The term information has been known for centuries with many different definitions and interpretations in use. But there is still lack of commonly accepted one and the question what actually stands for the term information is open. This is why some new concepts appear. Among them, the approach based on the analysis of its properties and functions enables better understanding of the phenomenon of information. In the literature one can find a suggestion that the concept of metaphor can be useful in this approach. The main aim of the paper is to give some examples of such metaphors called information metaphors, showing simultaneously their role in explaining some features of information by comparison to other concepts used in daily lives.

**Keywords:** information, information metaphors, function of information

**JEL:** A2, C80

---

<sup>a</sup> Wyższa Szkoła Informatyki Stosowanej i Zarządzania pod auspicjami Polskiej Akademii Nauk.

*Informacja* jest pojęciem często przywoływanym w różnych kontekstach i okolicznościach. Wprawdzie istnieje bogata literatura naukowa i użytkowa podejmująca próbę ostatecznego zdefiniowania tego pojęcia, trzeba jednak pamiętać, że każda definicja nakłada określone ramy interpretacyjne, które w innym kontekście mogą się okazać zbyt wąskie i jednostronne. Można wręcz postawić tezę, że dopiero wspólnie wszystkie definicje tworzą przestrzeń semantyczną pokrywającą się ze znaczeniem terminu *informacja*, przy zastrzeżeniu, że każda z nich wnosi jedynie niewielki wkład w wyjaśnienie jego sensu. Emilia Currás (2006, s. 208), autorka cenionych publikacji na temat informacji i jej wpływu na kształtowanie się społeczeństwa, w swoim artykule pisze: „[...] *informacja* jest terminem używanym przez każdego. Ale czy wiadomo, co on oznacza?”.

Poszukiwania oparte na konstruowaniu definicji jako warsztacie badawczym zasługują na to, by ich nie odrzucać, przede wszystkim jednak warto zwrócić uwagę na uzupełnienie pojęcia *informacja* na podstawie analizy cech (własności), jakie można i należy mu przypisać. Jedno z podejść w tym zakresie polega na wykorzystaniu metafor.

Metafora to figura stylistyczna oparta na skojarzeniu dwóch zjawisk poprzez dostrzeżenie ich podobieństwa i przeniesieniu nazwy jednego zjawiska na drugie (*Wielki słownik...*, 2005, s. 819). Obrazowe pokrewne znaczenie pojęcia już znanego ułatwia zrozumienie pojęcia mniej znanego lub trudnego.

Marek Hetmański, badacz teorii komunikacji, informacji i wiedzy, podkreśla, że ze względu na swobodną grę skojarzeń i oryginalnych porównań informacji z bardzo odległymi od niej zjawiskami i obiektami metafory informacyjne mają przewagę nad tradycyjnymi definicjami i opisami. W ten sposób stają się kluczem do budowania bogatych sieci powiązań pomiędzy *informacją* a występującymi w metaforach pojęciami i terminami.

Autor formułuje również interesującą tezę na temat wartości poznawczej metafor: „Im zjawisko, proces czy zdarzenie, o którym mówi dana metafora, jest mniej prawdopodobne, tym większą wartość poznawczą ma ona wówczas, można ją określić jako «informacyjną zawartość (moc) metaforę»” (Hetmański, 2015, s. 78). Stwierdzenie to wynika z kluczowej tezy Claude’a E. Shannona – im mniejsze jest prawdopodobieństwo pojawienia się jakiegoś komunikatu, tym więcej informacji wnosi jego opublikowanie (Shannon, 1948).

Celem artykułu jest pokazanie roli metafor w wyjaśnieniu niektórych cech i funkcji informacji przez analogię do innych pojęć, znanych z życia codziennego. Za Hetmańskim (2015) metafory charakteryzujące *informację* nazywane tu będą *metaforami informacyjnymi*.

## SMOG INFORMACYJNY

Smog to mgła zawierająca zanieczyszczenia powietrza atmosferycznego (*Encyklopedia PWN*). Typowy dla klimatu umiarkowanego i obserwowany głównie zimą smog przemysłowy (kwaśny) powstaje w wilgotnym powietrzu silnie zanie-

czyszczonym tzw. gazami kwaśnymi i pyłem węglowym. Smog jest zjawiskiem zagrażającym zdrowiu ludzi i zwierząt oraz szkodliwym dla środowiska.

Metafora smogu informacyjnego zrodziła się pod koniec lat 90. ub.w. w rezultacie dostrzeżenia przez specjalistów pojawiania się niekorzystnych zjawisk środowiskowych i społecznych wywołanych przez utratę kontroli człowieka nad procesami informacyjnymi i nagromadzenie się dużych ilości informacji o coraz bardziej wątpliwej jakości. Jak podaje Mariusz Szynkiewicz (2014), amerykański publicysta naukowy David Shenk użył tej metafory w 1997 r. Niezależnie od niego Ryszard Tadeusiewicz, polski automatyk i informatyk specjalizujący się w filozofii wiedzy i logice, w 1998 r. zamieścił metaforę smogu informacyjnego w tytule swojego referatu poświęconego wybranym aspektom rozwoju nowoczesnych technologii informacyjnych. Szynkiewicz zaznacza, że Tadeusiewicz uznał tę metaforę za najlepszy sposób scharakteryzowania zjawiska przeładowania informacyjnego oraz problemu oceny wiarygodności informacji w zasobach sieciowych. Jego zdaniem podobny problem przeładowania i niskiej wiarygodności dotyczył informacji w cyberprzestrzeni. Rozdrobnienie, nieuporządkowanie i nieusystematyzowanie utrudniają użytkownikom tych zasobów sprawne i efektywne wykorzystywanie zapisanych w nich treści. Stanowi to jeden z zasadniczych mankamentów współczesnej komunikacji internetowej jako źródła informacji obciążonych wielorakimi błędami i zniekształceniami, często wątpliwej wiarygodności i szkodliwych. Szynkiewicz (2014) stwierdził ponadto, że przesyt informacji paraliżuje dzisiaj rozwój i wykorzystanie technik informatycznych, a nadmiar dostępnych treści sprawia trudności w ich efektywnej selekcji i filtrowaniu.

Dodajmy, że podobnie jak w przypadku smogu atmosferycznego, winę za powstawanie smogu informacyjnego ponosi człowiek.

### BOMBA MEGABITOWA

Bomba zwykle kojarzy się z siłą niszczycielską. Tym większe wrażenie robi figura megabomby.

Stanisław Lem (1999) przywołał metaforę bomby w tytule swojej książki. Metafora ta podkreśla dynamiczny i ambiwalentny charakter informacji i zjawisk informacyjnych, przy czym – mimo istnienia pozytywnych skutków informacji – główny akcent jest położony na negatywne. Porównanie informacji do bomby megabitowej uwydatnia raptowne zwiększenie się ilości informacji i wiedzy oraz uaktywnienie procesów informacyjnych, których eksplozja stanowi uniwersalną prawidłowość. Lem zwracał w ten sposób uwagę na efekty dynamiki rozwoju – zwłaszcza niekontrolowanego – technologii informacyjnych i wywołanych przez nie szkodliwych zjawisk informacyjnych w skali globalnej, która obejmuje nie tylko cywilizację człowieka, lecz także cały wszechświat, łącznie ze światami możliwymi.

## INFORMACJA JAKO ODPADKI

Bogate zbiory informacji w cyberprzestrzeni, zasobach Big Data i innych rejestrach i repozytoriach to potencjalne źródło użytecznej wiedzy (choć o różnym stopniu przydatności). Niemniej odnalezienie w tych zbiorach konkretnych, potrzebnych użytkownikowi informacji staje się zadaniem czasochłonnym i skomplikowanym ze względu na konieczność przedarcia się przez gęszcz zupełnie zbędnych zapisów. Z perspektywy poszukującego przydatnych treści użytkownika stanowią one składowisko śmieci informacyjnych.

Lem (1999) traktuje nadmiar informacji krążących w Internecie jak odpadki i podkreśla zaśmiecenie środowiska informacyjnego ogromną ilością niedorzeczności i kłamstwa, bylejakości i szkodliwości. Podobnie pewien doświadczony informatyk porównuje zbiory informacji do składowiska śmieci, gdzie wprawdzie zaangażowani poszukiwacze znajdują czasami cenne rzeczy, a niekiedy nawet bogactwa, ale wymaga to dużo pracy, czasu i cierpliwości, a wreszcie szczęścia.

## REFLEKTOR ROZJAŚNIAJĄCY RZECZYWISTOŚĆ

Każde działanie społeczne jest poprzedzane podjęciem odpowiedniej decyzji, czasem wynikającej z głębokiej analizy, a czasem nieuświadomionej. Dotyczy to zarówno uczestnictwa człowieka w sferze publicznej i gospodarczej, jak i życia prywatnego. Podejmowanie decyzji na podstawie świadomych wyborów stanowi proces, który przypomina poruszanie się po nieznanym terenie. Aby dotrzeć do celu (czyli rozwiązać problem), należy poznać drogi, które doń wiodą. Ich nieznamość naraża na fiasko cały wysiłek wkładany w posuwanie się do przodu. W celu uniknięcia porażki trzeba zwizualizować sobie przestrzeń problemową wymagającą analizy i podjęcia decyzji – innymi słowy, trzeba mieć takie informacje, które rozjaśnią otoczenie. Im dokładniej będą one opisywać rzeczywistość, tym bardziej ograniczą niepewność co do skutków podejmowanych działań. Żadna informacja co prawda nie gwarantuje sukcesu ani nie uchroni przed błędami, ale bez tego „światła” postępowanie człowieka niechybnie byłoby narażone na przypadkowość i niepowodzenie. Docenienie roli informacji jako reflektora nie jest jednak równoznaczne z tezą, że im więcej informacji, tym większe są możliwości ustrzeżenia się przed błędami w podejmowanych decyzjach.

Peter Drucker (2010), znany jako ojciec nowoczesnego zarządzania, sformułował tezę, że nie technologia, lecz informacja stanowi podstawę działania (zarządzania). Bez niej wszelkie wysiłki człowieka byłyby skazane na podejmowanie decyzji metodą prób i błędów ze wszystkimi negatywnymi skutkami i niepotrzebnie ponoszonymi wydatkami, jak również kosztami czasu oraz zaangażowania intelektualnego.

## ŻYCIODAJNA KREW

Agnieszka Zając, specjalistka w dziedzinie systemów obliczeniowych, przywołuje słowa G. E. Milwarda: „Informacja jest życiodajną krwią w organizmie zarządzania – jest podstawą trafnych decyzji kierowniczych. Jeżeli nie można uzyskać właściwych informacji, to decyzje muszą opierać się na przypuszczaniu, odczuciach lub zgadywaniu” (Milward, 1967; za: Zając, 2009, s. 26).

Podobnie jak krew płynie w organizmie człowieka w dzień i w nocy, nieustannie pracuje jego mózg – i to właśnie funkcjonowanie mózgu stanowi wyznacznik życia człowieka. Mózg do pracy potrzebuje zaś nie tylko krwi, lecz także informacji. Specjaliści dowodzą, że gdyby człowieka pozbawić informacji (np. zamknąć w odizolowanym pomieszczeniu), to nawet przy pełnym zaopatrzeniu w pokarm i wodę groziłoby mu poważne zaburzenia.

Informacja – jak wyobrażone więzi krwi – integruje członków wspólnoty. Stopień trwałości tego układu zależy od sprawności procesów informacyjnych stanowiących istotę społecznego komunikowania się. Dowodzą tego opisy funkcjonowania społeczeństwa informacyjnego (np. Castells, 2007). Bez obiegu informacji nie może funkcjonować żaden budowany przez człowieka ład, taki jak obiekt gospodarczy, instytucja czy organizacja. Rozwijane od lat systemy informacyjne stanowią nieodłączną infrastrukturę takich ładów.

## INFORMACJA JAK OGRÓD BOTANICZNY

Dla laika w ogrodzie botanicznym rosną po prostu rośliny, z których część stanowią te kwitnące o różnych kolorach, zapachach i kształtach. Bliższy wgląd pozwala zauważyć ogromną różnorodność morfologiczną i bogactwo zarówno roślin, które wytwarzają kwiaty (nasiennych), jak i roślin zarodnikowych, takich jak mchy, paprocie i widłaki.

Podobnie jest z informacją – laikowi może się ona wydawać „sucha”. Jednakże, jak pisał Robert Poczobut (2005, s. 181): „W obiegu są takie terminy, jak informacja: ilościowa i jakościowa, aktualna i potencjalna, fizyczna, chemiczna i biologiczna, syntaktyczna, semantyczna i pragmatyczna, propozycyjalna i ostensywna, a także informacja fenomenalna”. Specjalista wyróżnia informację faktograficzną, proceduralną, kwerencyjną, imperatywną, fatyczną i wiele innych.

Znajomość tej różnorodności i specyficznych cech (własności) każdej odmiany racjonalizuje podejmowane działania związane – w pierwszym przypadku – z pielęgnacją roślin, w drugim zaś – z organizacją i realizacją procesów informacyjnych, np. budowaniem odpowiednich repozytoriów informacji czy wyszukiwaniem, gromadzeniem i wykorzystaniem danych. Specyficzne własności oznaczają także inne wymogi użytkowania informacji statystycznej, historycznej, gospodarczej, geograficznej, bankowej itd.

## OKNO NA ŚWIAT

„Człowiek nie tylko widzi świat, lecz także czyta go i słucha. [...] Dźwięk słowa wywołuje u nas obraz, rzeczywistość, marzenia, wspomnienie, plan. Słowo przypomina klucz otwierający wrota pamięci” – pisał rabin Philippe Haddad (2012, s. 87). Autor dodaje, że słowa służą człowiekowi do komunikacji, wyrażania myśli i ich zapisywania. „Jedność rodzaju ludzkiego zależy od dobrych słów, dobrych wypowiedzi, od słów właściwych i szczerych” – konkluduje.

Słowa niosą ze sobą treści, komunikaty, informacje. Treść książek wręcz przenosi czytelnika w opisywane światy, podobnie jak obrazy w telewizji włączają widza w prezentowane zdarzenia.

## LUSTRO ŚWIATA

Bliższa analiza funkcji informacji pozwala dostrzec w niej pewną analogię do lustra – w jednym i w drugim przypadku otrzymujemy pewne odzwierciedlenie wycinka rzeczywistości. Można tu znaleźć kilka podobieństw:

- informacja, podobnie jak lustro, daje obraz pewnej rzeczywistości. Lustro odbija jednak przedmioty materialne, podczas gdy informacja oddaje zarówno rzeczywistość materialną, jak i niematerialną;
- lustro musi być czyste – w przeciwnym wypadku obraz obiektu będzie skażony. Informacja powinna być czysta pod względem jakościowym, aby spełniała kryteria według uznania odbiorcy;
- zniekształcenie rzeczywistości może zachodzić w lustrze przez zniekształcenia fizyczne, w informacji zaś – na skutek błędów w uchwyceniu opisywanego obiektu przez zniekształcenia terminologiczne itp.;
- każde lustro, tak jak każda informacja, może być wykorzystane do różnych celów, zależnie od pomysłowości użytkownika.

## NADMIAR INFORMACJI – ZAWAŁ MÓZGU

Zdaniem Stefana Garczyńskiego (1984, s. 99) nadmierna ilość informacji bynajmniej nie gwarantuje redukcji niepewności co do skutków podejmowanych działań. Czasem taki nadmiar wręcz paraliżuje człowieka, utrudniając podjęcie racjonalnej decyzji. Nie tyle rozjaśnia problem, ile oślepia decydenta, który w konsekwencji nie jest w stanie rozsądnie przeanalizować wielu otrzymanych wiadomości. Pojawiają się wówczas odpadki informacyjne i smog, które utrudniają selekcję potrzebnych danych, a nawet mogą wprowadzić w błąd. Podobnie jak w fotografii może nastąpić przejawaskrawienie obrazu.

Na ogół nie sposób zebrać kompletu informacji, które są związane z danym problemem i potencjalnie mogłyby być przydatne; zawsze można dowieść, że jeszcze nie dowiedzieliśmy się wszystkiego. W praktyce zdobycie kompletnej wiedzy jest niemożliwe z powodu kosztów i czasu związanych ze zgromadzeniem informacji, a także ich niewyczerpywalności. Garczyński pisze, że bezwzględne domaganie się całości informacji w wielu sytuacjach jedynie paraliżuje działania. Nikt w takich warunkach nie podjąłby żadnej pracy (bo a nuż znajdzie się lepsza), nikt nie opowiedziałby się za żadną partią polityczną czy nikt by się nigdy nie ożenił.

### PODSUMOWANIE

Podane w artykule przykłady metafor informacyjnych nie wyczerpują całej listy. Można ją rozwinąć i uzupełnić o metafory informacyjne w statystyce. Ale już te przykłady utwierdzają w przekonaniu, że informacja jest pojęciem wieloznacznym i bogatym w treści. Oczywiście żadna z cząstkowych metafor nie odzwierciedla pełnego pola semantycznego tego pojęcia, pozwalając dostrzec tylko niewielki wycinek rzeczywistości. Stanowi jednak krok do jego poznania.

Metafora opiera się na analogii pomiędzy rozpatrywanym (poznawanym) pojęciem a pojęciem lub obiektem, które stanowi odniesienie do znanej rzeczywistości. Jak każda analogia pozwala tylko w przybliżeniu uchwycić cechy poznawanego obiektu lub pojęcia, czasem prowadząc do zbyt daleko wybiegających i mylnych wniosków. Jednak zaletą metafor jest to, że ułatwiają one przynajmniej wstępne zrozumienie trudnych pojęć, a takim jest informacja. Potem zostaje już tylko analiza i weryfikacja, czy brana pod uwagę metafora zasługuje na uwagę i czy stanowi dobry punkt wyjścia poznania badanego obiektu.

Połączenie definicji informacji z opisem jej własności i funkcji oraz związanymi z nią metaforami umożliwia pełniejsze zrozumienie istoty tego terminu. Warto jednak pamiętać o niepokoju kryjącym się w filozoficznym pytaniu Currás: czym jest informacja? Metafory nie odpowiadają na tego rodzaju pytania, ponieważ nie definiują pojęcia. Ułatwiają za to wyraźniejsze dostrzeżenie własności informacji i unaocznienie funkcji, jakie może ona pełnić w życiu człowieka – jak woda, światło, krew albo świat zaśmiecony przez nieroztropność, chciwość i inne wady człowieka.

### BIBLIOGRAFIA

- Castells, M. (2007). *Spółczesność sieci*. Warszawa: Wydawnictwo Naukowe PWN.
- Currás, E. (2006). Informationism and neural information assimilation. *Informationswissenschaft*, 4(57), 203–210.
- Drucker, P. (2010). *Zarządzanie XXI wieku – wyzwania*. Warszawa: New Media.
- Encyklopedia PWN*. Pobrane z: <https://encyklopedia.pwn.pl>.

- Garczyński, S. (1984). *Z informacją na bakier*. Warszawa: Instytut Wydawniczy Związków Zawodowych.
- Haddad, Ph. (2012). *Mądrości rabinów*. Warszawa: Wydawnictwo Cyklady.
- Hetmański, M. (2015). *Świat informacji*. Warszawa: Difin.
- Kuraś, M., Zając, A. (1966). Metafory organizacyjne – narzędzie badania i diagnozowania potrzeb użytkowników. *Zeszyty Naukowe Akademii Ekonomicznej w Krakowie*, (474), 5–27.
- Lem, S. (1999). *Bomba megabitowa*. Kraków: Wydawnictwo Literackie.
- Poczobut, R. (2005). Od informacji fizycznej do informacji fenomenalnej. W: M. Heller J. Mączka (red.), *Informacja a rozumienie* (s. 177–193), Kraków: Polska Akademia Umiejętności, Ośrodek Badań Interdyscyplinarnych „Biblios”.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3–4).
- Stefanowicz, B. (2013). *Informacja. Wiedza. Mądrość*. Biblioteka Wiadomości Statystycznych, t. 66. Warszawa: Główny Urząd Statystyczny.
- Szynkiewicz, M. (2014). Metafora smogu informacyjnego a procesy informacyjne. Pobrane z: [https://repozytorium.amu.edu.pl/bitstream/10593/13777/1/Studia\\_metod\\_32\\_2014\\_Mariusz\\_Szynkiewicz.pdf](https://repozytorium.amu.edu.pl/bitstream/10593/13777/1/Studia_metod_32_2014_Mariusz_Szynkiewicz.pdf).
- Wielki słownik wyrazów obcych PWN*. (2005). Warszawa: Wydawnictwo Naukowe PWN.
- Zając, A. (2009). Wykorzystanie metafor do identyfikacji potrzeb informacyjnych. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, (770), 23–35.



## Demograficzny obraz ziem polskich w przededniu I wojny światowej<sup>1</sup>

Józef Pociecha<sup>a</sup> 

**Streszczenie.** W 2018 r. Główny Urząd Statystyczny, wydawca roczników statystycznych, obchodził setną rocznicę powstania. W niniejszym artykule przedstawiono zawartość książki będącej bezpośrednim poprzednikiem roczników statystycznych – wydanej w 1915 r. *Statystyki Polski* autorstwa Adama Krzyżanowskiego i Kazimierza Władysława Kumanieckiego, wybitnych statystyków i ekonomistów polskich. Celem artykułu jest prezentacja sytuacji demograficznej na ziemiach polskich w przededniu wybuchu I wojny światowej, co zapoczątkowuje analizę procesu odzyskania niepodległości prowadzoną przez pryzmat sytuacji demograficznej i społeczno-gospodarczej kraju. Przedstawiono liczbę ludności zamieszkującej poszczególne dzielnice ziem polskich ok. 1910 r., szacunki dotyczące liczby ludności polskiej na tych ziemiach oraz dane o skali emigracji i ruchu naturalnego ludności historycznych ziem polskich. Informacje te stanowią przyczynek do poznania historii odrodzenia niepodległości Polski oraz historii statystyki polskiej.

**Słowa kluczowe:** historia Polski, historia statystyki polskiej, ludność ziem polskich, I wojna światowa

## Demographic picture of the Polish lands before the outbreak of the First World War

**Summary.** In 2018, Statistics Poland, the publisher of statistical yearbooks celebrated its 100th anniversary. The purpose of this work is to present a book which is the immediate predecessor of the Polish statistical yearbooks. The work, published in 1915, entitled "Polish Statistics", was elaborated by Adam Krzyżanowski and Kazimierz Władysław Kumaniecki, eminent Polish statisticians and economists. Based on this work, we can reconstruct the demographic picture of the Polish lands before the outbreak of the First World War, which initiates the analysis of the process of independence restoration through demographic and socio-economic situation of the country. The number of population on historical Polish lands around the year 1910 is shown. At the same time, the estimates of the number of Polish population on these lands with the information on the scale of emigration and vital statistics is presented. Such information contributes to the knowledge of the history of the rebirth of Polish independence and the history of statistics.

**Keywords:** history of Poland, history of Polish statistics, population of Polish lands, First World War

**JEL:** C80, J10

---

<sup>1</sup> Artykuł został opracowany na podstawie referatu wygłoszonego na II Kongresie Statystyki Polskiej, który odbył się w Warszawie w dniach 10–12 lipca 2018 r.

<sup>a</sup> Uniwersytet Ekonomiczny w Krakowie, Wydział Zarządzania.

W ubiegłym roku świętowano obchody stulecia odzyskania przez Polskę niepodległości. Ustanowienie niepodległego państwa polskiego stało się możliwe dzięki niezwykle sprzyjającemu rezultatowi I wojny światowej, wysiłkowi zbrojnemu Polaków oraz trafnemu formułowaniu celów strategicznych przez Józefa Piłsudskiego. W kontekście tej ważnej rocznicy obowiązkiem statystyka i demografa jest przedstawienie sytuacji demograficznej oraz ekonomicznej społeczeństwa polskiego w tamtym czasie.

Celem niniejszej pracy jest prezentacja sytuacji demograficznej na ziemiach polskich w przededniu wybuchu I wojny światowej, co zapoczątkowuje tak zarysowaną analizę procesu odzyskania niepodległości. Jej punktem wyjścia jest przedstawienie obrazu statystycznego ziem polskich przed rokiem 1914 na podstawie publikacji *Statystyka Polski*, będącej pierwszym owocem działalności Polskiego Towarzystwa Statystycznego (PTS), założonego w Krakowie w 1912 r.

Kolejnym krokiem w analizie statystycznej tego procesu jest zdanie sprawy ze strat poniesionych na ziemiach polskich w wyniku działań wojennych i dewastacji znajdującego się tam majątku. Dopiero połączenie informacji o sytuacji społeczno-gospodarczej ziem polskich przed I wojną światową i o stratach poniesionych w czasie jej trwania w pełni obrazuje sytuację terenów, które znalazły się w granicach Polski u progu odzyskania przez nią niepodległości. Jednakże przedstawienie statystycznego zarysu rozmiarów zniszczeń i strat ludności spowodowanych działaniami wojennymi na terytorium przyszłej Rzeczypospolitej wykracza poza ramy tego artykułu.

## POWSTANIE POLSKIEGO TOWARZYSTWA STATYSTYCZNEGO I PIERWSZY EFEKT JEGO DZIAŁALNOŚCI

Początek XX w. to w krajach europejskich czas rozwoju statystyki publicznej. Zakres i szczegółowość ówczesnych danych statystycznych mogą zadziwić współczesnego statystyka. W rozwoju statystyki publicznej przodowały Niemcy i Austria. Niemiecki urząd statystyczny Kaiserliches Statistisches Amt publikował obszerny *Statistisches Jahrbuch für das Deutsche Reich*, a jego austriacki odpowiednik K. K. Statistische Zentralkommission – roczniki *Oesterreichischen Statistik*. Również w Cesarstwie Rosyjskim po rewolucji 1905 r. nastąpiła modernizacja państwa, w tym służb statystycznych. Centralny Komitet Statystyczny wydawał publikacje *Statistika Rossijskoj Imperii* oraz *Statisticzeskij Jeżegodnik Rossiji* (Berger, 2008).

Służby statystyczne były zorganizowane na poziomie nie tylko państwa, lecz także niższych jednostek terytorialnych, takich jak prowincje i większe miasta. We Lwowie funkcjonowało operujące na obszarze Galicji Krajowe Biuro Statystyczne, które wydawało od 1900 r. *Podręcznik statystyki Galicji*, stanowiący rodzaj rocznika statystycznego. W Królestwie Polskim działał Warszawski Komi-

tet Statystyczny, publikujący *Trudy Warszawskiego Statystycznego Komitetu* – także mający charakter rocznika statystycznego (Domański, 2004). Rady miejskie niektórych dobrze zarządzanych miast tworzyły własne biura statystyczne. Przykładem może być Kraków, w którym istniało Miejskie Biuro Statystyczne (MBS); w ostatnich latach przed I wojną światową kierował nim doc. dr Kazimierz Władysław Kumaniecki (Berger, 2008).

Jak widać, dla polskich ziem i ludności polskiej istniały w końcowym okresie zaborów dość szczegółowe i wiarygodne dane demograficzne, społeczne i ekonomiczne, jednak rozproszone w publikacjach statystycznych państw zaborczych. Ze względu na brak własnego państwa niemożliwe było publikowanie rocznika statystycznego ziem polskich. W tym kontekście powstała idea założenia towarzystwa statystycznego, którego jednym z głównych zadań byłoby scalenie informacji statystycznych zamieszczanych w publikacjach urzędów państw zaborczych i wydawanie własnych prac na kształt roczników statystycznych ziem polskich i ludności polskiej. W Europie żywo działał wtedy ruch społeczny towarzystw statystycznych; wspomagały one organy statystyki publicznej i organizowały szczegółowe badania nieobjęte ewidencją statystyki publicznej. Przykładem może być *Societe de statistique de France* czy *Deutschen Statistisches Gessellschaft*. Do tych wzorców odnosiła się idea utworzenia towarzystwa statystycznego na ziemiach polskich (Pociecha, 2011).

Od 1861 r. Galicja posiadała autonomię w ramach monarchii austriackiej, dzięki czemu Kraków stał się intelektualną i symboliczną stolicą Polski, nieistniejącej od 1795 r. jako byt państwowy. W Krakowie odbywały się uroczystości patriotyczne i funkcjonował – ponownie spolonizowany – Uniwersytet Jagielloński (UJ), który na przełomie wieków przeżywał rozkwit. Tutaj w 1872 r. powstała Akademia Umiejętności, skupiająca polskich uczonych ze wszystkich zaborów, a także działała Akademia Sztuk Pięknych, której profesorem był Jan Matejko. Kraków, obok Lwowa, stanowił największe skupisko polskiej inteligencji (Bieniarzówna i Małecki, 1979).

Właśnie w Krakowie na początku 1912 r. grono statystyków i ekonomistów podjęło projekt powołania odrębnego stowarzyszenia statystyków polskich. Było to możliwe jedynie w Galicji, cieszącej się autonomią i względnie liberalnymi prawami dotyczącymi zakładania i działalności stowarzyszeń. Realizację powierzono Kumanieckiemu, który z pomocą Władysława Studnickiego oraz z poparciem inicjatywy przez ówczesnego Prezydenta Miasta Krakowa prof. dr. Juliusza Leo przygotował projekt statutu stowarzyszenia. 29 marca 1912 r. Kumaniecki złożył do Cesarsko-Królewskiego (C.K.) Namiestnictwa we Lwowie podanie o zezwolenie na działalność stowarzyszenia pod nazwą Polskie Towarzystwo Statystyczne z siedzibą w Krakowie. C.K. Namiestnictwo pismem z 9 kwietnia 1912 r. oznajmiło, że „nie zakazuje zawiązania tego stowarzyszenia”, i zatwierdziło projekt jego statutu. Datę tego pisma przyjmuje się za początek istnienia PTS (Pociecha, 2011).

Siedzibą stowarzyszenia stał się lokal krakowskiego MBS. Na czele PTS stała dwunastoosobowy zarząd, zwany wówczas Wydziałem, wybrany przez walne zgromadzenie członków. Zgodnie ze statutem co roku zmieniano jedną trzecią składu zarządu. Wydział PTS wybierał ze swego grona prezesa, sekretarza i skarbnika oraz ich zastępców. Pierwszym prezesem Towarzystwa (i jedynym w okresie jego działalności w Krakowie) został prof. dr Juliusz Leo, profesor skarbowości na UJ, pełniący również funkcję Prezydenta Miasta Krakowa. Na sekretarza PTS wybrano Kumanieckiego. Nie jest znana liczba członków Towarzystwa – nie zachowały się materiały archiwalne dotyczące działalności PTS z tego okresu (Pocięcha, 2011).

Z działalności publikacyjnej, szeroko przewidzianej w statucie, PTS przygotowało przed wybuchem I wojny światowej tylko jedną pozycję, a mianowicie *Statystykę Polski*, wydrukowaną w 1915 r. w drukarni UJ (Krzyżanowski i Kumaniecki, 1915). Opracowanie jej sfinansowało kilka krakowskich instytucji, w znacznej mierze Akademia Umiejętności. Autorami *Statystyki Polski* wymienionymi na karcie tytułowej byli, później bardzo znani, A. Krzyżanowski i K. Kumaniecki. Oprócz nich w redagowaniu publikacji brali udział również, wchodzący w skład wybranego przez Wydział PTS Komitetu Redakcyjnego, prof. dr Franciszek Bujak (przewodniczący), prof. dr Stanisław Surzycki i prof. dr Edward Grabowski, który opracował rozdział poświęcony źródłom danych statystycznych. Współpracownikami Komitetu Redakcyjnego byli m.in. dr Marcin Nadobnik ze Lwowa oraz Michał Römer z Wilna, radą zaś i „życzliwym poparciem” (jak to określono w przedmowie) służyli tacy znani i szanowani profesorowie, jak: prof. dr Władysław L. Jaworski, prof. dr Stanisław Kutrzeba, prof. dr Michał Rostworowski i dr Franciszek Stefczyk. Zestawień tabelarycznych dokonali pracownicy MBS (Krzyżanowski i Kumaniecki, 1915).

## ZAKRES TERYTORIALNY I CZASOWY STATYSTYKI POLSKI

*Statystyka Polski* została wydana w wersji trójjęzycznej: polskiej, niemieckiej i francuskiej, co wskazuje, że zamiarem autorów było rozpowszechnienie przygotowanej pracy w środowisku międzynarodowym. W przedmowie autorzy zwracali uwagę, że ich publikacja stanowi realizację głównego celu założenia PTS:

Polskie Towarzystwo Statystyczne, zawiązane przed kilku laty w Krakowie, uważało za główne swe zadanie zaradzić złemu, wynikającemu z rozproszenia materiału statystycznego, oświetlającego życie i rozwój polskiego społeczeństwa. Trudności statystycznego badania stosunków polskich pochodzą nie tylko z rozproszenia, ale także z różnorodności rozporządzalnego materiału. Zakres i metoda zestawień statystycznych są odmienne w każdym z trzech państw, które dziś władają ziemią polskimi. Zachodzą różnice językowe, różnice w jednostkach mierniczych powierzchni ciężaru,

objętości, stosowanych w obliczeniach statystycznych. W Austrii i w Niemczech system metryczny jest daleko bardziej rozpowszechniony, niż w Rosji. Tym rozlicznym trudnościom pragnie zaradzić wydawnictwo obejmujące całość ziem polskich, które obecnie oddajemy w ręce Sz. Czytelników w nadziei, że tym sposobem przyczynimy się do pogłębienia naszego społecznego samopoznania (Krzyżanowski i Kumaniecki, 1915, s. XXIV).

Komitet Redakcyjny *Statystyki Polski* tak określił w przedmowie cel jej wydania:

Pracę naszą podjęliśmy jako podręcznik [podręcznik oznaczał wówczas zbiór informacji z pewnego okresu, np. zbiór roczników z kilku lat – przyp. aut.]. Nie mogło być mowy ani o wyczerpującym zestawieniu wszystkich cyfr statystycznych ani o wydawnictwie, któreby czyniło zbędnym sięganie do źródeł. Chodziło tylko o ich wskazanie oraz ułatwienie rozejrzenia się w obszernym materiale statystycznym przez skupienie i treściowe przedstawienie najważniejszych obliczeń (Krzyżanowski i Kumaniecki, 1915, s. XXV).

Ciekawie rozważano problem zakresu terytorialnego publikacji. Autorzy pisali w przedmowie:

Określenie zakresu naszych zestawień w przestrzeni nasuwało dużo wątpliwości. Chcąc dać obraz Polski w cyfrach statystycznych, musieliśmy przede wszystkim odpowiedzieć na pytanie, które obszary terytorialne staną się przedmiotem naszego badania? Mieliśmy do wyboru dwie drogi: historyczną i etnograficzną. Mogliśmy wyświetlać statystycznie stosunki na ziemiach, wchodzących ongi w skład państwa polskiego albo na ziemiach, dziś przez Polaków zamieszkałych. Wybraliśmy drogę pośrednią w mniemaniu, że tym sposobem ominiemy jednostronność związane z wyłącznym oparciem się na zasadzie historycznej czy etnograficznej. Objęliśmy naszymi zestawieniami przede wszystkim ziemie dawnej Rzeczypospolitej: Galicyę, Królestwo, Poznańskie i Prusy Królewskie, oraz Litwę, Białoruś i trzy gubernie, stanowiące dawniej ruskie województwa: Podole, Wołyń, Kijowszczyznę; nie pominęliśmy krajów, zamieszkałych w znacznej mierze przez Polaków, które już od dawna przestały być częścią państwa polskiego. Nie pominęliśmy ani rejencji olsztyńskiej ani okręgów cieszyńskiego i opolskiego na Śląsku. Z powodu trudności zebrania wiarygodnego materiału nie uwzględniliśmy Polski na wychodźstwie, licznego zastępu Polaków zamieszkałych w granicach trzech państw zaborczych, ale poza obszarami, które wciągnęliśmy w zakres naszych obliczeń, na Bukowinie, w prowincjach nadreńskich, w Berlinie i Petersburgu, a nadto w Ameryce Północnej i Południowej. Zestawienia nie obejmują całego polskiego społeczeństwa, nie obejmują ogółu ludzi, których językiem ojczystym jest język polski, albowiem nie chcieliśmy sięgać poza granice terytorialne wytknięte w naszej pracy. Wciągnęliśmy w jej zakres całą ludność zamieszkałą w tych granicach, a więc nie ograniczyliśmy się wyłącznie do statystyki polskiego społeczeństwa. Na obszarze, opracowanym statystycznie w naszej książce, mieszkają obok Polaków ludzie, których językiem ojczystym nie jest język polski. Wyodrębnienie Polaków z ogółu ludności, zamieszkującej ziemie dawnej Rzeczypospolitej, Śląsk opolski i cieszyński, rejencję olsztyńską nie dałoby się przeprowadzić, a nawet nie byłoby pożądanem (Krzyżanowski i Kumaniecki, 1915, s. XXIV–XXV).

Przytoczony fragment ukazuje poglądy elit polskich na temat zakresu terytorialnego przysłego odrodzonego państwa polskiego. W ich przekonaniu powinno ono obejmować terytorium Rzeczypospolitej Obojga Narodów w granicach z 1772 r. (przedrozbiorowych), powiększone o Śląsk Górny, Opolski i Cieszyński oraz Południowe Mazury, wchodzące oprócz Warmii w skład rejencji olsztyńskiej. Historia pokazała, że w 1918 r. ustanowienie państwa polskiego w takich granicach nie było możliwe.

Co do przedziału czasowego uwzględnionego w *Statystyce Polski*, autorzy twierdzili, że chodzi im o „zobrazowanie współczesnego stanu rzeczy”, czyli okresu przed I wojną światową. Zaznaczyli jednakże:

Chcąc ułatwić zrozumienie i ocenę teraźniejszości, sięgnęliśmy wstecz w granicach wskazanych charakterem i rozmiarami pracy. Najwięcej statystyki retrospekcyjnej znajdzie Sz. Czytelnik w części pierwszej, obejmującej stan i ruch ludności. W tym dziale historyczny materiał statystyczny jest obfitszy, niż w innych. Nie cofaliśmy się wstecz poza rok 1815 – granica narzucona stanem rozporządzalnego materiału (Krzyżanowski i Kumaniecki, 1915, s. XXV).

Nadmienili również:

Zagraniczne wydawnictwa obfitują w diagramy i mapy. Nie stać nas było na ich wykonanie i wydrukowanie. Cieszymy się, że, o ile nam wiadomo, równocześnie dochodzi do skutku wydawnictwo *Geograficznego i statystycznego atlasu Polski* pod redakcją Prof. Dr. Eugeniusza Romera. Jedna praca uzupełnia drugą (Krzyżanowski i Kumaniecki, 1915, s. XXVI).

Chodzi tutaj o publikację Romera z 1916 r. Należy podkreślić, że zarówno *Statystyka Polski*, jak i *Geograficzno-statystyczny atlas Polski* odegrały istotną rolę w polskiej argumentacji w sprawie określenia granic odrodzonego państwa polskiego na konferencji pokojowej w Paryżu, zakończonej podpisaniem Traktatu Wersalskiego w 1919 r.<sup>2</sup>

Autorzy *Statystyki Polski* podali ponadto, ile czasu zajęła im praca nad tą publikacją:

Wykonanie zestawień rozpoczęliśmy w styczniu 1914 r. W sześć miesięcy później praca była bliska ukończenia. [...] A jednak praca doznała przerwy w chwili wybuchu wojny. Dopiero wiosną 1915 r. podjęliśmy ją na nowo z zamiarem jak najrychlejszego doprowadzenia do końca, choćby w rozmiarach mniej wyczerpujących niż pierwotnie zamierzaliśmy (Krzyżanowski i Kumaniecki, 1915, s. XXVI).

Przedmowę opatrzyli miejscem i datą: „Kraków, wrzesień 1915 r.”.

---

<sup>2</sup> Trzeba dodać, że *Statystyka Polski* stanowi podstawowe źródło danych dotyczących początku XX w. na ziemiach polskich, publikowanych w wielu pracach statystyczno-historycznych, jak np. *Historia Polski w liczbach* (GUS, 1994).

*Statystyka Polski* jest obszernym dziełem, liczącym łącznie 348 stron, w tym: przedmowa i spis źródeł pracy, numerowane cyframi łacińskimi (31 stron) oraz część tabelaryczna, numerowana cyframi arabskimi (317 stron i 315 tablic). Praca składa się z czterech części: I. Ludność – 104 strony i 74 tablice; II. Stosunki gospodarcze i społeczne – 156 stron i 196 tablic; III. Oświata i szkolnictwo – 22 strony i 18 tablic; IV. Wybory do ciał ustawodawczych – 24 strony i 26 tablic.

W niniejszej pracy przytoczone i zinterpretowane zostaną jedynie te dane spośród gąszczu informacji dotyczących sytuacji demograficznej i społeczno-gospodarczej historycznych ziem polskich, które zawarto w części I *Statystyki Polski*. Składają się one na obraz demograficzny tych ziem w przededniu wybuchu I wojny światowej.

### LUDNOŚĆ HISTORYCZNYCH ZIEM POLSKICH

Część I *Statystyki Polski* zawiera bogate informacje dotyczące stanu ludności w poszczególnych dzielnicach ziem polskich, zmian liczby ludności w XIX w., gęstości zaludnienia, ludności w podziale na miasto i wieś oraz struktury ludności według płci i wieku, wyznania i języka ojczystego.

Według danych Warszawskiego Komitetu Statystycznego, *Jeżegodnika Rosji* za 1911 r. oraz spisów powszechnych ludności z 1910 r. w Austrii i Prusach ludność ziem polskich w granicach przyjętych przez autorów *Statystyki Polski* (teren o powierzchni 753,198 tys. km<sup>2</sup>) wynosiła 52,5 mln osób, a przeciętna gęstość zaludnienia – 70 os./km<sup>2</sup>. Szczegółowe dane przedstawia tabl. 1. Należy zauważyć, że powierzchnia Rzeczypospolitej Obojga Narodów przed rozbiorami (w 1771 r.) wynosiła 733 tys. km<sup>2</sup>, a liczba ludności – 12,3 mln, co dawało gęstość zaludnienia 17 os./km<sup>2</sup> (GUS, 2012, s. 102). Obszar odrodzonej Rzeczypospolitej zajmował 388,634 tys. km<sup>2</sup>, czyli 51,6% terytorium uznanego przez autorów *Statystyki Polski* za historyczne ziemie polskie.

**TABL. 1. LUDNOŚĆ ZIEM POLSKICH WEDŁUG ICH HISTORYCZNYCH CZĘŚCI**

Historyczne części ziem polskich	Liczba ludności w tys.	Udział w ogólnej liczbie ludności w %	Gęstość zaludnienia w os./km <sup>2</sup>
<b>Ziemie polskie</b> .....	<b>52537,0</b>	<b>100,0</b>	<b>70</b>
Ziemie pod panowaniem rosyjskim <sup>a</sup> .....	37521,8	71,4	63
Królestwo Polskie .....	12476,0	23,7	101
Litwa i Białoruś .....	12709,2	24,2	42
Rusć .....	12336,6	23,5	75
Ziemie pod panowaniem austriackim <sup>b</sup> .....	8460,4	16,1	105
Galicja .....	8025,7	15,3	102
Śląsk Cieszyński .....	434,8	0,8	196
Ziemie pod panowaniem pruskim <sup>b</sup> .....	6554,8	12,5	76
Wielkie Księstwo Poznańskie .....	2099,8	4,0	72
Prusy Królewskie .....	1703,5	3,3	67
Rejencja olsztyńska .....	543,5	1,0	45
Rejencja opolska .....	2208,0	4,2	167

a Dane za 1911 r. b Wyniki spisu powszechnego z 1910 r.

Źródło: Krzyżanowski i Kumaniecki (1915, s. 2, tabl. 1) oraz obliczenia własne.

*Statystyka Polski* zawiera również dane na temat stopnia urbanizacji historycznych ziem polskich. O ile Królestwo Polskie było stosunkowo wysoko zurbanizowane, co wiązało się głównie z rozwojem przemysłu włókienniczego w Łodzi i okolicach oraz przemysłu ciężkiego w Zagłębiu, a także z rozwojem Warszawy jako ówczesnej metropolii, o tyle pozostałe ziemie polskie, z wyjątkiem Śląska Górnego i Cieszyńskiego, miały charakter wiejski.

Warszawa, największe miasto historycznych ziem polskich w 1913 r. liczyła według *Statystyki Polski* 845 tys. mieszkańców, przy czym w omawianym okresie nie należały do niej jeszcze Czyste, Brudno<sup>3</sup>, Mokotów i Młociny. Łącznie z tymi miejscowościami liczba ludności Warszawy wynosiłaby 1 mln 36 tys. mieszkańców<sup>4</sup>. Drugim co do wielkości miastem była Łódź, w której w 1913 r. mieszkało 459 tys. osób; wraz z Radogoszczą, wtedy jeszcze nienależącą do Łodzi, miasto liczyłoby 584 tys. mieszkańców. Trzecim pod względem wielkości miastem Królestwa Polskiego był Sosnowiec ze 114 tys. mieszkańców w 1913 r. Zgodnie z danymi za 1912 r. największym miastem na Litwie było Wilno (186 tys. mieszkańców), na Białorusi – Mińsk (113 tys. mieszkańców), a na Ukrainie – Kijów (413 tys. mieszkańców). Według danych za 1910 r. największym miastem Galicji była jej stolica – Lwów, w którym mieszkało 207 tys. osób, podczas gdy w Krakowie – 154 tys. osób. Na ziemiach pod panowaniem pruskim największym miastem był Gdańsk (170 tys. mieszkańców w 1910 r.); ustępował mu Poznań ze 157 tys. ludności.

## LUDNOŚĆ POLSKA WEDŁUG HISTORYCZNYCH ZIEM POLSKICH

W czasie opracowywania publikacji nie badano liczby ludności pod względem narodowości. Narodowość określano przez pryzmat języka ojczystego (którym posługiwano się w domu i w towarzystwie) oraz wyznania. *Statystyka Polski* zawiera tego rodzaju informacje, pochodzące ze spisów powszechnych przeprowadzonych w 1910 r. w Cesarstwie Austriackim i Cesarstwie Niemieckim oraz z szacunków urzędowych wykonanych w 1909 r. w Cesarstwie Rosyjskim na okoliczność wprowadzania ziemstw. Jeżeli za kryterium przynależności narodowej uznać język ojczysty, to liczba Polaków zamieszkujących historyczne ziemie polskie przed I wojną światową wynosiła ok. 18,5 mln osób (szczegółowe dane przedstawiono w tabl. 2). Liczba ta nie obejmuje Polaków przebywających na emigracji w Ameryce Północnej i Południowej oraz w Europie poza ziemiami polskimi, jak również licznych skupisk Polaków w stolicach państw zaborczych: Petersburgu, Berlinie i Wiedniu.

<sup>3</sup> Pisownia oryginalna (tabl. 13, s. 24). Obecnie – Bródno.

<sup>4</sup> Warszawa była trzecim co do wielkości miastem Cesarstwa Rosyjskiego, po Petersburgu (2 mln 36 tys. mieszkańców w 1912 r.) oraz Moskwie (1 mln 617 tys. mieszkańców w 1912 r.).



TABL. 2. LUDNOŚĆ POLSKA WEDŁUG HISTORYCZNYCH ZIEM POLSKICH

Historyczne części ziem polskich	Liczba ludności w tys.	Udział w ogólnej liczbie ludności w %
<b>Ziemie polskie</b> .....	<b>18441,1</b>	<b>35,1</b>
Ziemie pod panowaniem rosyjskim <sup>a</sup> .....	10076,7	26,9
Królestwo Polskie .....	8957,8	71,8
Litwa i Białoruś .....	711,8	5,6
Ruś .....	407,1	3,3
Ziemie pod panowaniem austriackim <sup>b</sup> .....	4941,8	58,4
Galicja .....	4703,1	58,6
Śląsk Cieszyński .....	238,7	54,9
Ziemie pod panowaniem pruskim <sup>b</sup> .....	3422,6	52,2
Wielkie Księstwo Poznańskie .....	1291,4	61,5
Prusy Królewskie .....	604,7	35,5
Rejencja olsztyńska .....	267,9	49,3
Rejencja opolska .....	1258,6	57,0

a Dane za 1911 r. b Wyniki spisu powszechnego z 1910 r.

Źródło: Krzyżanowski i Kumaniecki (1915, s. 39–42, tabl. 33) oraz obliczenia własne.

W *Statystyce Polski* podano tylko fragmentaryczne dane na temat polskich emigrantów. O zakresie emigracji z ziem polskich mogą świadczyć dane Urzędu Imigracyjnego USA, według których emigracja Polaków do Stanów Zjednoczonych w latach 1909–1912 wynosiła 362,5 tys. osób, Rusinów (Ukraińców) – 83,4 tys. osób, a Żydów – 313,6 tys. osób, co daje łącznie liczbę 759,5 tys. osób opuszczających ziemię polską. Według *The Canada Year Book (Rocznika Statystycznego Kanady)* za 1911 r. liczba Polaków emigrujących do tego państwa w latach 1908–1911 wynosiła 11,6 tys. osób. W *Statystyce Polski* nie uwzględniono porównywalnych danych dotyczących emigracji do Ameryki Południowej czy innych krajów.

Podobnie brakuje kompletnych danych odnośnie do liczby osób emigrujących z poszczególnych części ziem polskich. Możemy znaleźć jedynie informacje częściowe, np. w tabl. 72 podano informację, że w latach 1880–1910 z Galicji wyemigrowało 857 tys. osób, w tym z Galicji Zachodniej (z przewagą ludności polskiej) – 468 tys. osób, a z Galicji Wschodniej (o przewadze ludności niepolskiej) – 389 tys. osób.

Oprócz emigracji długookresowej, głównie za ocean, częsta w tym czasie była emigracja sezonowa do Niemiec, „na saksy”. Według danych niemieckiej Arbeiterzentrale emigracja sezonowa Polaków z zaboru rosyjskiego w okresie 1912/13<sup>5</sup> wyniosła 285,8 tys. osób, Polaków z zaboru austriackiego – 75,1 tys. osób, a Rusinów z Austro-Węgier – 91,4 tys. osób. Emigracja sezonowa nierzadko stanowiła początek emigracji długookresowej.

<sup>5</sup> Dane roczne za okres od 1 października 1912 r. do 30 września 1913 r.

Nie ma wiarygodnych informacji na temat migracji wewnętrznych, w szczególności ze wsi do miast, przede wszystkim rozrastających się ośrodków przemysłowych Królestwa Polskiego. Można jednak przyjąć, że wzrost liczby ich mieszkańców odbywał się głównie dzięki migracji ze wsi. Tylko dla Galicji, w zamieszczonej w *Statystyce Polski* tabl. 73, zatytułowanej „Wpływ wędrowek na stan ludności w powiatach Galicji”, można przeczytać, że dodatnie saldo migracji w latach 1880–1910 obserwowano jedynie we Lwowie (63 tys.), w Krakowie (46 tys.), Stanisławowie (24 tys.) oraz w powiecie lwowskim (16 tys.). Dodatkowo saldo migracji (powyżej 1000 osób) zarejestrowano także w Przemyślu, Podgórzu (w 1915 r. włączonym do Krakowa) oraz Drohobyczu i Turce, czyli w galicyjskim zagłębiu naftowym.

Liczbę 18,5 mln Polaków zamieszkałych na historycznych ziemiach polskich można w przybliżeniu przyjąć za wiarygodną. Pozostaje do oszacowania liczba Polaków żyjących w tym czasie na wychodźstwie. W pracach historyczno-demograficznych można znaleźć wiele rozbieżnych danych na temat emigracji z ziem polskich. Dotyczą one ogółu ludności, a z ziem polskich emigrowali nie tylko Polacy, lecz także Żydzi i Rusini (Ukraińcy), nie licząc innych narodowości. Ponadto w danych tych zazwyczaj nie rozróżniano emigracji stałej (długookresowej) i sezonowej. Problematyczna jest także ocena procesu asymilacji, w efekcie której o polskim pochodzeniu świadczyłoby jedynie polskie brzmienie nazwiska. W opinii autora niniejszego artykułu w polskiej literaturze istnieje tendencja do zawyżania liczebności Polaków. Uwzględnienie tych zastrzeżeń, jak również informacji zawartych w pracy *Emigracja z ziem polskich w czasach nowożytnych i najnowszych* (Piłch, 1984) pozwala oszacować prawdopodobną liczbę Polaków żyjących na całym świecie w przededniu I wojny światowej na ok. 22 mln osób.

#### RUCH NATURALNY LUDNOŚCI NA HISTORYCZNYCH ZIEMIACH POLSKICH

W *Statystyce Polski* szeroko scharakteryzowano ruch naturalny ludności przez podanie liczby oraz intensywności zawieranych małżeństw, urodzeń i zgonów oraz płci ludności na ziemiach polskich w poszczególnych zaborach. Dane uwzględniają także podział na miasto i wieś, różnice wyznaniowe i niekiedy narodowościowe. Z danych za rok 1911<sup>6</sup> dotyczących intensywności urodzeń (36,4 na 1000 ludności) i zgonów (21,5 na 1000 ludności) oraz przyrostu naturalnego (14,9 na 1000 ludności) na historycznych ziemiach polskich wynika, że na początku XX w. ludność tych ziem znajdowała się u progu procesu przejścia demograficznego (Okólski, 1990). Zróżnicowanie intensywności zawierania mał-

<sup>6</sup> Za 1910 r. w przypadku braku danych za rok 1911.

żeństw w poszczególnych częściach ziem polskich nie było duże. Przy średniej intensywności wynoszącej 6,8 na 1000 ludności, najwyższą – 7,9 na 1000 ludności obserwowano na ziemiach ukraińskich należących do Rosji, a najniższą, wynoszącą 6,4 na 1000 ludności, na Litwie i Białorusi.

Zróżnicowanie współczynników urodzeń i zgonów było wyższe w ujęciu regionalnym. Najwyższa intensywność urodzeń obserwowana była na Górnym Śląsku (40,2) i w Galicji (39,8), a najniższa na Mazurach – w rejencji olsztyńskiej (34,0) oraz na Litwie i Białorusi – 34,1. Jeśli chodzi o intensywność zgonów, to współczynnik zgonów brutto osiągał najwyższą wartość (27,7) w Galicji, a najniższą (18,4) w rejencji olsztyńskiej (Mazury). Z kolei przyrost naturalny najwyższy był w Wielkim Księstwie Poznańskim (19,1) i ogólnie na ziemiach pod panowaniem pruskim (18,2), a najniższy (13,5) – na ziemiach ukraińskich. Pierwsza faza przejścia demograficznego osiągnęła zatem największe zaawansowanie na ziemiach pod panowaniem pruskim, szczególnie w Wielkopolsce i na Górnym Śląsku – obszarach najwyżej rozwiniętych gospodarczo, a zwłaszcza pod względem rolnictwa, zapewniającego produkcję żywności dla rosnącej liczby ludności. Szczegółowe dane zawiera tabl. 3.

**TABL. 3. RUCH NATURALNY LUDNOŚCI NA ZIEMIACH POLSKICH W 1911 R.<sup>a</sup>**

Historyczne części ziem polskich	Małżeństwa	Urodzenia	Zgony	Przyrost naturalny
	na 1000 ludności			
<b>Ziemie polskie</b> .....	<b>6,8</b>	<b>36,4</b>	<b>21,5</b>	<b>14,9</b>
Ziemie pod panowaniem rosyjskim	7,0	35,5	20,8	14,7
Królestwo Polskie .....	7,2	37,5	20,3	14,2
Litwa i Białoruś .....	6,4	34,1	19,0	15,1
Ruś .....	7,9	37,4	23,9	13,5
Ziemie pod panowaniem austriackim	7,5	38,4	25,3	14,1
Galicja .....	7,5	39,8	25,7	14,1
Śląsk Cieszyński .....	7,6	37,2	22,3	14,9
Ziemie pod panowaniem pruskim ...	6,9	37,8	19,7	18,2
Wielkie Księstwo Poznańskie .....	7,0	38,1	19,0	19,1
Prusy Królewskie .....	7,1	37,8	20,2	17,7
Rejencja olsztyńska .....	6,8	34,0	18,4	15,6
Rejencja opolska .....	6,7	40,2	21,3	18,9

a W 1910 r. w przypadku braku danych za rok 1911.

Źródło: Krzyżanowski i Kumaniecki (1915, s. 72–80, tabl. 43–50) oraz obliczenia własne.

W procesie przejścia demograficznego szczególną rolę odgrywa przyrost naturalny ludności. W tabl. 4 przedstawiono wielkość przyrostu naturalnego ludności w poszczególnych regionach historycznych ziem polskich w latach 1861–1910 w przedziałach dziesięcioletnich. Dane te potwierdzają, że ziemie polskie znajdowały się w pierwszej fazie procesu przejścia demograficznego.

**TABL. 4. ŚREDNIOROCZNY PRZYRÓST NATURALNY NA 1000 LUDNOŚCI W POSZCZEGÓLNYCH DZIELNICACH ZIEM POLSKICH**

Historyczne części ziem polskich	1861–1870	1871–1880	1881–1890	1891–1900	1901–1910
Królestwo Polskie .....	14,4	13,8	14,4	15,6	15,4
Litwa i Białoruś .....	14,7	–	14,7	16,6	14,7
Ruś .....	15,3	–	17,8	17,4	16,2
Galicja .....	11,5	7,6	11,3	14,5	15,6
Śląsk Cieszyński .....	–	6,3	12,0	8,1	14,9
Wielkie Księstwo Poznańskie ...	13,8	16,2	16,3	19,4	19,7
Prusy Królewskie .....	13,6	15,1	16,0	17,9	18,1
Rejencja olsztyńska .....	–	–	–	–	16,5
Rejencja opolska .....	–	15,0	14,4	20,0	18,5

Źródło: Krzyżanowski i Kumaniecki (1915, s. 80, tabl. 50) oraz obliczenia własne.

Wielkość przyrostu naturalnego regulowana była w tych czasach przede wszystkim umieralnością spowodowaną epidemiami oraz klęskami głodu związanymi z nieurodzajami płodów rolnych. Szczegółowe dane zamieszczone w *Statystyce Polski* wskazują na duże zróżnicowanie przyrostu naturalnego w poszczególnych dekadach, a zwłaszcza w poszczególnych regionach. Najniższy przyrost naturalny, wynoszący 5,7‰, odnotowano w 1905 r. w Guberni Kowieńskiej, najwyższy – 24,3‰ – w 1902 r. w Guberni Mohylewskiej.

Rozpatrując przyrost naturalny w czasie, można zauważyć jego ciągły wzrost, od ponad 13‰ do ok. 20‰. Wyróżniała się pod tym względem Wielkopolska (Wielkie Księstwo Poznańskie), ze wzrostem przyrostu naturalnego od 13,8‰ w latach 1861–1870 do 19,7‰ w latach 1901–1910; podobnie proces ten przebiegał na Pomorzu (w Prusach Królewskich). Przedstawione dane świadczą o najwyższym postępie w zakresie ochrony zdrowia i rolnictwa na ziemiach polskich pod zaborem pruskim. Natomiast na terenach pod zaborem rosyjskim, szczególnie w Królestwie Polskim oraz na Litwie i Białorusi, przyrost naturalny nie ulegał większym zmianom. Należy też zauważyć, że pierwsza dekada XX w. była pod względem rozwoju demograficznego gorsza niż ostatnia dekada XIX w.

Dane zamieszczone w *Statystyce Polski* nie zawierają informacji o płodności (dzietności) kobiet na ziemiach polskich, z jednym wyjątkiem przedstawionym w tabl. 5.

**TABL. 5. PŁODNOŚĆ POLEK I NIEMEK W REJENCJACH POD PANOWANIEM PRUSKIM WEDŁUG SPISU POWSZECHNEGO Z 1910 R.**

Narodowość	Liczba dzieci na 100 mężatek, wdów i rozwódek								
	0	1	2	3	4	5 lub 6	7 lub 8	9 i więcej	brak danych
Polki .....	4,7	6,6	7,6	8,2	8,9	18,4	15,8	20,1	9,7
Niemki .....	6,9	10,2	11,5	10,7	9,9	16,1	11,3	12,9	10,5

Źródło: Krzyżanowski i Kumaniecki (1915, s. 98, tabl. 67).

Dane przedstawione w tabl. 4 i 5 należy interpretować w świetle teorii przejścia demograficznego, w którym zauważa się trzy lub cztery fazy takiego przejścia. W podstawowym ujęciu czterofazowym wyróżnia się fazy:

- pierwszą, nazywaną fazą równowagi prymitywnej (rozrzutnej), charakteryzującą się wysoką umieralnością i wysoką, naturalną płodnością;
- drugą, obejmującą okres obniżania się umieralności, przy naturalnej rozrodzności, do momentu pojawienia się spadkowego trendu płodności; w fazie tej obserwuje się wzrostową tendencję przyrostu naturalnego;
- trzecią, w której następuje dalsza, jednak coraz słabsza tendencja do obniżania umieralności, przy coraz szybszej tendencji spadkowej płodności; w fazie tej obserwuje się spadkową tendencję przyrostu naturalnego;
- czwartą – równowagi nowoczesnej (oszczędnej), cechującą się niską umieralnością i płodnością oraz niskim (zerowym lub ujemnym) przyrostem naturalnym (Kurkiewicz, 1998).

Zgodnie z założeniami tej teorii można stwierdzić, że ludność zaboru rosyjskiego, a zwłaszcza Królestwa Polskiego, znajdowała się w pierwszej fazie przejścia demograficznego o stabilnym przyroście naturalnym. Ludność Galicji i Śląska Cieszyńskiego od początku XX w. zaczęła wchodzić w fazę drugą. Proces transformacji demograficznej był najbardziej zaawansowany na ziemiach polskich pod panowaniem niemieckim, gdzie już w drugiej połowie XIX w. następowała druga faza przejścia. Dane tabl. 5 wskazują jednak, że społeczność niemiecka na tym obszarze wchodziła w proces transformacji demograficznej szybciej niż społeczność polska.

## ZAKOŃCZENIE

28 lipca 1914 r. Austro-Węgry wypowiedziały wojnę Serbii, co na zasadzie domina pociągnęło za sobą wypowiedzenie wojny między państwami ententy a państwami centralnymi i doprowadziło do ponadczteroletniego światowego konfliktu. Dotychczasowy porządek polityczny, trwający od kongresu wiedeńskiego 1815 r., upadł. Wynikiem wojny była klęska trzech zaborców; Polska odzyskała niepodległość. Powstały instytucje niepodległego państwa polskiego, w tym utworzony 3 lipca 1918 r. Główny Urząd Statystyczny, nie było więc potrzeby wydawania polskich roczników statystycznych siłami społecznymi. PTS w Krakowie zaprzestał działalności w okresie wojennym i nie podjął jej po odzyskaniu niepodległości. Pierwszy *Rocznik Statystyki Rzeczypospolitej Polskiej* został wydany przez GUS w dwóch częściach: pierwszej w 1921 r. i drugiej w 1923 r. (GUS, 2013). Część pierwsza zawierała informacje na temat meteorologii, podziału administracyjnego i ludności. W części drugiej, opartej już częściowo na danych z przeprowadzonego w 1921 r. spisu powszechnego, zamieszczono informacje o użytkowaniu ziemi i produkcji rolniczej. *Rocznik Statystyki Rzeczypospolitej Polskiej* z 1921 r. był pierwszym rocznikiem statystycznym

Polski, jednak za najwcześniejszy faktyczny rocznik statystyczny ziem polskich trzeba uznać *Statystykę Polski*, wydaną przez PTS w 1915 r.

Dane statystyczne zawarte w *Statystyce Polski* zdezaktualizowały się wskutek ogromnych wojennych strat ludności oraz strat gospodarczych. Nastąpiły też duże wojenne i powojenne przemieszczenia ludności. Niemniej dane te stanowią punkt wyjścia szacowania strat wojennych na ziemiach polskich oraz odniesienie dla statystycznego badania rozwoju demograficznego i gospodarczego niepodległej Polski.

#### BIBLIOGRAFIA

- Berger, J. (2008). Rys historyczny powstania GUS. *Wiadomości Statystyczne*, (1), 2–17.
- Bieniarzówna, J., Małecki, J. M. (1979). *Dzieje Krakowa*, t. 3. Kraków: Wydawnictwo Literackie.
- Domański, Cz. (2004). Jubileusz Polskiego Towarzystwa Statystycznego. W: A. Zeliaś (red.), *Tradycje i obecne zadania statystyki w Polsce* (s. 9–22). Kraków: Wydawnictwo Akademii Ekonomicznej w Krakowie.
- GUS. (1994). *Historia Polski w liczbach*. Warszawa: Główny Urząd Statystyczny.
- GUS. (2012). *Zarys historii Polski w liczbach*. Warszawa: Główny Urząd Statystyczny.
- GUS. (2013). *Historia, stan obecny i aktualne wyzwania wobec statystyki publicznej*. Warszawa: Główny Urząd Statystyczny.
- Krzyżanowski, A., Kumaniecki, K. Wł. (1915). *Statystyka Polski*. Kraków: Polskie Towarzystwo Statystyczne.
- Kurkiewicz, J. (1998). *Modele przemian płodności w wybranych krajach europejskich w świetle drugiego przejścia demograficznego*. Kraków: Zeszyty Naukowe AE w Krakowie, Seria Specjalna: Monografie.
- Okólski, M. (red.). (1990). *Teoria przejścia demograficznego*. Warszawa: PWE.
- Pilch, A. (red.). (1984). *Emigracja z ziem polskich w czasach nowożytnych i najnowszych*. Warszawa: PWN.
- Pocięcha, J. (2011). Powstanie Polskiego Towarzystwa Statystycznego w Krakowie. J. Leo – pierwszy prezes Towarzystwa. *Wiadomości Statystyczne*, (10), 1–9.
- Pocięcha, J. (2018). "Statistics of Poland" – the First Yearbook of Polish Lands: Authors and Content. *Acta Universitatis Lodzensis. Folia Oeconomica*, 5(338), 229–245, DOI: <http://dx.doi.org/10.18778/0208-6018.338.14>.
- Romer, E. (1916). *Geograficzno-statystyczny atlas Polski*. Warszawa i Kraków: Gebethner i Wolff.

## Profesor dr hab. Andrzej Stanisław Barczak, dr h.c. (1939–2019)

**P**rofesor dr hab. Andrzej Stanisław Barczak urodził się w Nowym Bytomiu 4 lutego 1939 r. Począwszy od studiów w latach 1957–1962 związany był z Uniwersytetem Ekonomicznym w Katowicach (UEK). Początkowo pracował jako wolontariusz w kierowanej przez prof. Maksymiliana Ziomka Katedrze Statystyki, a następnie, po ukończeniu studiów, w tej samej katedrze pod kierunkiem prof. dr hab. Zbigniewa Pawłowskiego. Tam obronił w 1968 r. pracę doktorską pt. *Ekonometryczna analiza kosztów produkcji na przykładzie przemysłu piwowarsko-słodowniczego*.



Początki kariery naukowej  
Fot. Archiwum UEK

W 1976 r. na podstawie rozprawy habilitacyjnej *Makromodele ekonometryczne a planowanie gospodarki narodowej* uzyskał stopień doktora habilitowanego. W 1985 r. otrzymał tytuł naukowy profesora i od 1993 r. do przejścia na emeryturę pracował na stanowisku profesora zwyczajnego na UEK. W 2005 r. uzyskał godność doktora honoris causa Uniwersytetu Ekonomicznego we Wrocławiu. Ponadto w latach 1996–2003 był kierownikiem Zakładu Mikroekonomii Instytutu Zarządzania Wydziału Zarządzania i Komunikacji Społecznej Uniwersytetu Jagiellońskiego w Krakowie oraz w latach 2009–2011 rektorem Śląskiej Wyższej Szkoły Zarządzania im. gen. Jerzego Ziętka w Katowicach.

Na UEK pełnił m.in. funkcje prodziekana Wydziału Przemysłu (1975–1981), dziekana Wydziału Przemysłu (1984–1987), kierownika Katedry Ekonometrii, prorektora (1981–1984), przewodniczącego Uczelnianej Komisji Wyborczej (1993–1996) i rzecznika dyscyplinarnego (1993–1996). 29 września 2017 r., po 50 latach od nadania Mu stopnia naukowego doktora, na UEK odbyła się uroczystość odnowienia Jego doktoratu.

Profesor Barczak był kontynuatorem i współtwórcą zapoczątkowanej przez Pawłowskiego katowickiej szkoły ekonometrycznej. W latach 70. wykładał w Letniej Szkole Ekonometrycznej w Ustroniu Wielkopolskim, której celem było przybliżenie zagadnień ekonometrii ekonomistom reprezentującym inne specjalności. Był także stałym i aktywnym uczestnikiem oraz współorganizatorem Kon-

ferencji Statystyków, Ekonometryków i Matematyków Uczelni Ekonomicznych Polski Południowej, nie opuszczając chyba żadnej z dotychczasowych 55 konferencji, łącznie z ostatnią, która odbyła się w kwietniu br.

Wypromował 17 doktorów (ostatniego 6 marca 2019 r.), ponad 600 magistrów, był recenzentem 75 prac doktorskich, ponad 50 prac habilitacyjnych, 28 wniosków o tytuł naukowy oraz czterech wniosków o tytuł doktora honoris causa.

Pełnił wiele ważnych i zaszczytnych funkcji z wyboru: w Radzie Głównej Szkolnictwa Wyższego (1987–1998), w Centralnej Komisji ds. Stopni Naukowych i Tytułu Naukowego przy Prezesie Rady Ministrów oraz w Komisji Konkursowej Komitetu Badań Naukowych (1994–2006). Od 1976 r. był wybierany do Komitetu Statystyki i Ekonometrii PAN, od 1993 r. będąc członkiem Prezydium, a w latach 2007–2010 przewodniczącym. Ponadto w latach 2006–2014 sprawował funkcję Prezesa Oddziału Katowice Polskiego Towarzystwa Ekonomicznego oraz członka Prezydium Zarządu Krajowego.

Został powołany do Naukowej Rady Statystycznej GUS (1981–2005), Komisji Matematycznej GUS (1976–1981) i Komisji Nagród Naukowych przy Prezesie Rady Ministrów (1994–1997). Był członkiem rad programowych czasopism naukowych: „Przeglądu Statystycznego”, „Gospodarki Narodowej”, „Śląskiego Przeglądu Statystycznego” i „Journal of Economics & Management”. Należał do wielu towarzystw naukowych, m.in.: The Econometric Society, The European Economic Association, The Royal Economic Society, Polskiego Towarzystwa Statystycznego, Polskiego Towarzystwa Prakseologicznego, Towarzystwa Naukowego Organizacji i Kierowania, Polskiej Akademii Umiejętności w Krakowie oraz Towarzystwa Popierania i Krzewienia Nauki.

Za swe osiągnięcia został odznaczony Krzyżem Oficerskim i Kawalerskim Orderu Odrodzenia Polski, Złotym i Srebrnym Krzyżem Zasługi, Medalem Komisji Edukacji Narodowej; otrzymał również pięć nagród Ministra Nauki i Szkolnictwa Wyższego, a w 2010 r. prestiżową nagrodę za całokształt działalności w dziedzinie nauk ekonomicznych.

Jego zainteresowania naukowe początkowo koncentrowały się wokół problemów funkcji produkcji, ekonometrycznej analizy kosztów produkcji oraz budowy modeli gospodarki Polski, a w szczególności związków makromodeli z planowaniem gospodarki narodowej oraz wykorzystaniem modeli ekonometrycznych w prognozowaniu wykonania planów gospodarczych. Następnie zajmował się zagadnieniami regionalnych rachunków narodowych. W latach 80. zainteresował się także analizami społecznymi – badaniami warunków życia, ekonomicznych skutków alkoholizmu oraz śląskiego rynku pracy.





Uroczystość odnowienia doktoratu 29 września 2017 r.  
Fot. Archiwum UEK

Profesor Barczak z wielkim zaangażowaniem zajmował się popularyzacją wiedzy ekonomicznej. Napisał wiele artykułów i komentarzy prasowych, często występował w radiu i telewizji.

Miał dwoje dzieci: Stanisława i Aleksandrę. Często podkreślał ogromną rolę, jaką odegrały w Jego życiu trzy kobiety: matka Jadwiga, żona Mirosława i córka Aleksandra.

Lubił muzykę klasyczną, wędrowki po górach, podróże, kuchnię francuską i dobrą kawę.

Był człowiekiem pogodnym, obdarzonym dużym poczuciem humoru i niezwykle przyjaznym, a do tego urodzonym gawędziarzem. Cieszył się nie tylko ogromnym szacunkiem, lecz także wielką sympatią kolegów i młodzieży.

Zmarł 21 maja.

Odszedł znakomity uczyony, szanowany nauczyciel akademicki, wychowawca wielu pokoleń studentów, człowiek prawy, życzliwy współpracownikom i studentom. Dla polskiej nauki i środowiska akademickiego, a zwłaszcza UEK, jego śmierć stanowi niepowetowaną stratę.

**Zofia Mielecka-Kubień** (Uniwersytet Ekonomiczny w Katowicach)

**A**ndrzej Stanisław Barczak był wybitnym ekonomistą i cenionym ekonometrykiem. Jego kariera akademicka zaczęła się w czasie studiów w Wyższej Szkole Ekonomicznej w Katowicach, gdzie w latach 1960–1962 podjął pracę jako asystent na zasadzie wolontariatu w Katedrze Statystyki. Po skończeniu studiów w 1962 r. został tam zatrudniony. W 1968 r. zdobył stopień doktora, a w 1976 r. na podstawie rozprawy habilitacyjnej *Makromodele ekonometryczne a planowanie gospodarki narodowej* – doktora habilitowanego na Wydziale Przemysłu Akademii Ekonomicznej im. Karola Adamieckiego w Katowicach. W lutym 1985 r. otrzymał tytuł profesora nadzwyczajnego. 14 kwietnia 2005 r. uzyskał godność doktora honoris causa Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.



W swoim gabinecie w UEK  
Fot. Archiwum UEK

W dowód uznania autorytetu naukowego Profesora od 1976 r. był on wybierany do Komitetu Statystyki i Ekonometrii PAN. Sprawował funkcje z wyboru w Radzie Głównej Szkolnictwa Wyższego (1987–1998), Centralnej Komisji ds. Stopni Naukowych i Tytułu Naukowego przy Prezesie Rady Ministrów oraz Komisji Konkursowej Komitetu Badań Naukowych (1994–2006). W latach 2006–2014 pełnił funkcję Prezesa Oddziału Katowice Polskiego Towarzystwa Ekonomicznego; był również członkiem Prezydium Zarządu Krajowego. Działal także w wielu innych towarzystwach naukowych oraz radach programowych czasopism naukowych.

Był aktywnym członkiem Polskiego Towarzystwa Statystycznego (PTS), a w latach 1981–2010 przewodniczył Radzie Oddziału Katowickiego PTS. W tym okresie inicjował wiele wspólnych przedsięwzięć PTS i służb statystyki publicznej, biorąc w nich czynny udział. W 2007 r. przeprowadził konsultacje naukowe na temat metodologii obliczeń produktu krajowego brutto w przekroju terytorialnym na potrzeby badania Urzędu Statystycznego w Katowicach pn. *1.67.07 Produkt krajowy brutto*

*i jego elementy w ujęciu regionalnym.* Współpracował z urzędem także podczas realizacji projektu *Wsparcie w zakresie rozwijania zestawu wskaźników do monitorowania narodowych strategii zrównoważonego rozwoju – Rozwój i wdrożenie polskiego zestawu wskaźników zrównoważonego rozwoju w latach 2009–2011* jako ekspert w zakresie teoretycznych podstaw konstrukcji wskaźników zrównoważonego rozwoju, metod ich obliczania i interpretacji.

Profesor był nieustannie aktywny na polu statystyki. Interesował się informacjami statystycznymi z różnych obszarów tematycznych, od wyników finansowych po budownictwo, także w ostatnich latach.



Z wojewodą śląskim Zygmuntem Łukaszczykiem  
Fot. Archiwum UEK

Oprócz działalności naukowo-dydaktycznej angażował się w rozwiązywanie społeczno-gospodarczych problemów regionu. Był moderatorem Programu Regionalnej Polityki Gospodarczej Górnego Śląska oraz głównym negocjatorem strony społecznej Kontraktu Regionalnego dla województwa katowickiego. Przewodniczył Radzie Nadzorczej Funduszu Górnośląskiego S.A. w latach 1997–1999, a także zasiadał w radach nadzorczych kilku spółek, w tym Kompanii Węglowej S.A. w Katowicach, Eurofaktora S.A. w Mysłowicach i Akcept S.A. w Mysłowicach. W 2008 r. został powołany w skład rady programowej Radia Katowice i wybrany na jej przewodniczącego.

Był doradcą wojewody śląskiego i prezydenta Katowic. Zajmował się również popularyzacją wiedzy ekonomicznej w prasie, radiu i telewizji.

## Konferencja naukowa *230 lat Statystyki Publicznej w Polsce*

Wkrótce po obchodach Dnia Statystyki Polskiej<sup>1</sup>, który przypada 9 marca, i w związku z tym wydarzeniem, 18 i 19 marca 2019 r. odbyła się w Łodzi ogólnopolska konferencja naukowa *230 lat Statystyki Publicznej w Polsce – Od Spisu Ludności Rzeczypospolitej Obojga Narodów do Narodowego Spisu Powszechnego 2021*. Honorowy patronat nad konferencją objęli: Wojewoda Łódzki prof. dr hab. Zbigniew Rau, Marszałek Województwa Łódzkiego Grzegorz Schreiber, Prezydent Miasta Łodzi Hanna Zdanowska oraz JM Rektor Uniwersytetu Łódzkiego (UŁ) prof. dr hab. Antoni Różalski. Organizatorem był Urząd Statystyczny (US) w Łodzi, współorganizatorami – Główny Urząd Statystyczny (GUS), Polskie Towarzystwo Statystyczne (PTS) oraz Komitet Statystyki i Ekonometrii Polskiej Akademii Nauk (KSiE PAN), a partnerem – Centralna Biblioteka Statystyczna im. Stefana Szulca w Warszawie (CBS).

W komitecie programowym zasiedli: prezes GUS dr Dominik Rozkrut, prezes PTS prof. dr hab. Czesław Domański, przewodniczący KSiE PAN prof. dr hab. Krzysztof Jajuga, prof. dr hab. Włodzimierz Okrasa z Uniwersytetu Kardynała Stefana Wyszyńskiego w Warszawie, dr inż. Janusz Dygaszewicz z GUS, dyrektor CBS dr Bożena Łazowska, dr hab. Jerzy Tadeusz Kowaleski, prof. UŁ, prof. dr hab. Cezary Kukło z Uniwersytetu w Białymstoku, prof. dr hab. Józef Pocięcha z Uniwersytetu Ekonomicznego w Krakowie (UEK) oraz dr hab. Konrad Wnęk z Uniwersytetu Jagiellońskiego.

Patronat medialny nad wydarzeniem sprawowały TVP3 Łódź –Telewizja Polska S.A. oraz Radio Łódź S.A.

Termin konferencji nawiązywał do rocznicy pierwszego ogólnokrajowego spisu ludności na ziemiach polskich pod nazwą *Lustracja dymów i podanie ludności*, przeprowadzonego w marcu 1789 r. Zamierzeniem organizatorów było nie tylko uczczenie tego historycznego wydarzenia, lecz przede wszystkim próba oceny procesu doskonalenia metod gromadzenia informacji i poszukiwania nowych źródeł danych na podstawie historii, rozwoju i osiągnięć statystyki publicznej w zakresie spisów powszechnych.

Konferencji towarzyszyły wystawa *230 lat Spisów Ludności na Ziemiach Polskich*, gala wręczenia nagród w konkursie na najlepszą pracę magisterską i doktorską z zakresu statystyki oraz walne zgromadzenie PTS.

---

<sup>1</sup> Dzień Statystyki Polskiej uchwalony został przez Radę Główną PTS, KSiE PAN oraz Prezesa GUS 2 grudnia 2008 r. Jego obchody mają na celu promocję i popularyzację statystyki jako nauki oraz statystyki publicznej, a zwłaszcza stają się okazją do informowania opinii publicznej o wynikach naukowych badań statystycznych oraz o rozwoju wiedzy statystycznej.

W konferencji uczestniczyło ponad 100 osób, m.in. już wymienieni przedstawiciele statystyki publicznej: Dominik Rozkrut, Czesław Domański, Włodzimierz Okrasa, Janusz Dygaszewicz i Bożena Łazowska. Licznie reprezentowane było środowisko akademickie; udział wzięli także przedstawiciele Instytutu Geodezji i Kartografii (IGiK), Powiatowego Urzędu Pracy w Kutnie oraz Urzędu Miasta Łodzi.

W imieniu dyrektora US w Łodzi dr. Piotra Ryszarda Cmeli oraz własnym uczestników powitał dr Artur Mikulec, p.o. zastępcy dyrektora US w Łodzi. Podczas otwarcia głos zabrali również Dominik Rozkrut, Czesław Domański i Antoni Różalski.

Tematyka konferencji – podczas której wygłoszono 22 referaty – obejmowała zagadnienia takie, jak:

- prehistoria spisów powszechnych, od staropolskich spisów ludności i spisów organizowanych przez władze kościelne po rolę Sejmu Czteroletniego w rozwoju statystyki publicznej;
- statystyka publiczna w Polsce, w tym osiągnięcia prekursorów statystyki publicznej, działalność towarzystw i instytucji statystycznych oraz powstanie GUS;
- osiągnięcia i wyzwania statystyki publicznej w XXI w., a zwłaszcza nowe źródła danych w statystyce publicznej, nowoczesne metody i narzędzia statystyczne, rozwój badań statystycznych, nowe technologie w procesie badawczym, innowacje do zastosowania w Narodowym Spisie Powszechnym 2021 i rozwój statystyki międzynarodowej.

Podczas pierwszej sesji, której przewodniczył prof. dr hab. Mirosław Krzyśko z Uniwersytetu im. Adama Mickiewicza w Poznaniu, Dominik Rozkrut w wystąpieniu *Statystyka Publiczna jako element Narodowej Platformy Cyfrowej* omówił rolę statystyki publicznej w dobie zmieniającego się „krajobrazu rynku informacji”, sposób funkcjonowania statystyki publicznej w warunkach transformacji cyfrowej oraz związane z tym wyzwania. Stwierdził, że GUS ze swoimi zasobami stanowi rdzeń infrastruktury informacyjnej państwa, ponieważ zarządza informacjami i danymi na poziomie krajowym i ponadnarodowym, monitoruje procesy społeczno-gospodarcze, a także świadczy usługi informacyjne dla społeczeństwa. Cyfryzacja sektora publicznego stawia jednak nowe wymagania. Aby im sprostać, statystyka publiczna musi nie tylko dbać o swoje fundamenty, takie jak transparentność, niezależność i użyteczność, lecz także podejmować wyzwania związane z *big data*, tworzyć wysokiej jakości dane i informacje, poszukiwać nowych źródeł danych, wchodzić w strategiczne partnerstwa z innymi gestorami danych, budować „ekosystem danych”, a wreszcie wykorzystywać nowoczesne narzędzia informatyczne, w tym *open source*, i dbać o publiczny ład informacyjny. Ponadto zadaniem statystyki publicznej pozostaje wspieranie i rozwijanie edukacji statystycznej na różnych poziomach i w różnych formach,

jak również współpraca ze środowiskiem naukowym w ramach Naukowej Rady Statystycznej i Komisji Metodologicznej czy publikowanie opracowań naukowych i organizacja konferencji.

Z kolei Włodzimierz Okrasa w prezentacji zatytułowanej *Spisy ludności w USA 1790–2020: perspektywa metodologiczno-historyczna. Statystyka imigracji z Polski* skontekstualizował problemy pojawiające się w historii amerykańskich spisów, od obliczania liczby mieszkańców w okresie przed zniesieniem niewolnictwa, po współczesne sposoby uwzględniania w takich rachunkach niezarejestrowanych imigrantów. Zwrócił też uwagę na sprawy związane ze zmianą sposobów realizacji spisu, od „naturalnego” do bazujących na nowoczesnych technologiach, np. internetowych. Drugi nurt rozważań poświęcony był szacunkom dotyczącym mieszkańców polskiego pochodzenia przed spisem w 1920 r., czyli tych, którzy przybyli do Stanów Zjednoczonych jako obywatele krajów zabornych, oraz w spisach późniejszych. Na koniec prelegent podjął próbę wskazania pytań badawczych, które mogłyby stanowić agendę badań statystyczno-historycznych nad Polonią amerykańską, w pełni wykorzystujących informacje ze spisów USA.

Czesław Domański w wystąpieniu *Statystycy I Rzeczypospolitej Obojga Narodów* przedstawił sylwetki przedstawicieli statystyki administracyjnej: Jana Długosza (*Księga uposażeń diecezji krakowskiej*) i Jana Łaskiego (*Księga uposażeń*) oraz statystyki opisowej: Macieja Miechowity (*Traktat o obojgu Sarmacjach*) i Marcina Bielskiego (*Kronika wszystkiego świata*). Omówił również pracę Macieja Strykowskiego *Opis europejskiej Sarmacji*, poświęconą procesom ludnościowym i gospodarczym Rzeczypospolitej Obojga Narodów.

W pierwszym dniu konferencji odbyło się ponadto walne zgromadzenie członków PTS, poświęcone strategii towarzystwa na następne lata. Wybrano nowych przedstawicieli organów PTS w związku z zakończeniem kadencji 2014–2018. Ważnym punktem spotkania była uroczystość wręczenia dr. Tadeuszowi Jurkowi, członkowi Rady Głównej PTS w ostatniej kadencji, Odznaki Honorowej za Zasługi dla Statystyki Rzeczypospolitej Polskiej. Aktu dekoracji dokonał prezes GUS Dominik Rozkrut.

W drugiej sesji, prowadzonej przez dr Agnieszkę Zgierską z GUS, Sylwia Filas-Przybył (US w Poznaniu) wygłosiła referat *Metodyka klasyfikowania jednostek administracyjnych w oparciu o stopień urbanizacji – DEGURBA*. Omówiła w nim ideę tytułowej klasyfikacji, uwzględniającej kryteria ludnościowe w układzie odniesienia, którym jest siatka kilometrowych kwadratów. Nowa klasyfikacja przyporządkowuje jednostki statystyczne LAU 2 (*Local Administrative Units*) do jednego z trzech typów obszarów: gęsto zaludnionego, pośredniego lub słabo zaludnionego. Ma to zapewnić porównywalność poszczególnych kategorii jednostek administracyjnych na obszarze Unii Europejskiej (UE). Autorka przedstawiła porównania podstawowych charakterystyk jednostek administracyjnych kraju według rejestru TERYT i klasyfikacji DEGURBA.

Dr Wiesława Gierańczyk z US w Bydgoszczy w wystąpieniu *Obszary rynku pracy – nowe ujęcie prezentowania statystyk na przykładzie demografii i rynku pracy* przedstawiła koncepcję obszarów rynku pracy LMA (Labour Market Areas), początkowo rozwijaną w Wielkiej Brytanii i we Włoszech. LMA należy rozumieć jako zintegrowane gospodarczo terytoria, innymi słowy obszary funkcjonalne, w których mieszkańcy mogą znaleźć zatrudnienie w rozsądnej odległości od miejsca zamieszkania albo zmienić pracę bez konieczności przeprowadzki. Autorka omówiła wyniki badań, w których podjęto próbę wyodrębnienia lokalnych obszarów rynku pracy w Polsce odzwierciedlających relacje pomiędzy miejscem pracy a miejscem zamieszkania pracowników. W rezultacie delimitacji zidentyfikowano 339 lokalnych obszarów rynku pracy w Polsce. Analizie poddano zależność pomiędzy sytuacją na rynku pracy wyrażoną stopą bezrobocia a „zachowaniami demograficznymi”. Zaznaczono użyteczność LMA do kształtowania strategii rozwoju regionalnego.

Tomasz Józefowski z US w Poznaniu w referacie *Estymacja pośrednia liczby pracujących na obszarach funkcjonalnych miast wojewódzkich* przybliżył możliwości wykorzystania metod statystyki małych obszarów na potrzeby badań rynku pracy – dla nietypowych przekrojów, których nie przewidziano na etapie planowania badania reprezentacyjnego. Głównym celem referatu było przedstawienie możliwości wykorzystania estymatora typu SPREE i jego uogólnienia (GLSM, GLSMM) do oszacowania liczby pracujących na obszarach funkcjonalnych miast wojewódzkich z uwzględnieniem płci i grup wieku. Wspomniane estymatory przenoszą strukturę asocjacji dla kategorii zmiennej pomocniczej (zmiennych pomocniczych) i badanych domen znaną z rejestrów administracyjnych bądź spisów powszechnych, przy jednoczesnym zachowaniu spójności z szacunkami na wyższym poziomie agregacji. W analizie wykorzystano dane z Badania Aktywności Ekonomicznej Ludności, a w charakterze zmiennych pomocniczych – informacje ze źródeł administracyjnych, m.in. ZUS i KRUS.

W sesji równoległej, pod przewodnictwem Dominiki Rogalińskiej z GUS, dr Wanda Nowara, dr Arleta Olbrot-Brzezińska i Artur Owczarkowski z US w Poznaniu w wystąpieniu *1918–2018. Poznaniacy w statystyce publicznej* przybliżyli sylwetki poznańskich statystyków, naukowców, urzędników i działaczy społecznych oraz gospodarczych, którzy w okresie ostatnich stu lat tworzyli statystykę publiczną w Wielkopolsce. Byli to m.in.: Marcin Nadobnik, Zygmunt Zaleski, Stanisław Waszak i Zbigniew Czerwiński. Trudno się nie zgodzić z poglądem autorów, że rozwój statystyki, w tym publicznej, można rozpatrywać przez pryzmat rozwoju myśli statystycznej, włącznie z metodami badań zjawisk i procesów społeczno-ekonomicznych, powstawania i działalności biur oraz urzędów statystycznych, ośrodków akademickich i PTS, a przede wszystkim – pracy statystyków.

Dr Krzysztof Markowski z US w Lublinie w prezentacji *Migracja gospodarcza (ekonomiczna) podmiotów gospodarczych i ludności z terenu województwa lu-*

*belskiego* przypominał wymiary migracji gospodarczej i zarobkowej, uwarunkowania rozwoju przedsiębiorstw oraz społeczno-gospodarcze konsekwencje migracji. Tło empiryczne prowadzonych rozważań stanowiły dane dotyczące skali migracji gospodarczej ludności i firm z woj. lubelskiego w odniesieniu do innych województw i całego kraju. Wyszczególniono także skutki zjawiska migracji dla regionu.

Dr Dorota Wyszowska z US w Białymstoku w referacie *Zmiany demograficzne a nowe obszary badawcze statystyki publicznej* poruszyła zagadnienie srebrnej gospodarki. Statystyka publiczna podjęła się, w reakcji na niezaspokojone potrzeby informacyjne z zakresu srebrnej gospodarki, badań rozwojowych, których celem było wprowadzenie tej tematyki w obręb swoich prac. W latach 2017 i 2018 w ramach prac eksperymentalnych zrealizowano badanie strony podażowej srebrnej gospodarki w Polsce. W wystąpieniu zaprezentowano wybrane wyniki tego badania i przedstawiono zamierzenia dalszych studiów nad szeroko rozumianą gospodarką senioralną.

Trzecią sesję, której przewodniczył Jerzy Kowaleski, otworzyła prezentacja Konrada Wnęka *Ludność Krakowa w XVIII w. na podstawie spisu z 1779 r. Po między epoką prestatystyczną i powszechnych spisów ludności*. Badacz zreferował wyniki analizy tego mało znanego spisu ludności Krakowa – pierwszego spisu ludności miasta, który nie ograniczał się do posesjonatów lub podatników i zachował się do czasów współczesnych. Wprawdzie nie był to jeszcze w pełni spis imienny, ponieważ przeprowadzono go według gospodarstw domowych, niemniej na jego podstawie można było wykonać badania nad strukturą demograficzną, zawodową (z pogrupowaniem kategorii zawodowych według klasyfikacji HISCO, uwzględniającym płeć) i społeczną miasta. Zaprezentowano społeczne podziały źródłowe w porównaniu do sumariusza poprzedniego spisu z 1777 r. Podjęto też temat pracy zawodowej kobiet, która bynajmniej nie pojawiła się w wyniku przeobrażeń związanych z rewolucją przemysłową, lecz występowała szeroko w środowisku miejskim już w drugiej połowie XVIII w.

Referat Bożeny Łazowskiej *Osiągnięcia i porażki I powszechnego spisu ludności Rzeczypospolitej Polskiej w 1921 r.* dotyczył warunków, w jakich odbywał się tamten spis. Były one tak trudne, że część komisarzy spisowych rezygnowała z pracy, zaś mieszkańcy reagowali nieufnością, niechętnie udzielając odpowiedzi. Autorka omówiła największe osiągnięcia pierwszego spisu ludności RP, tj. dokonanie pełnego bilansu ludności i stanu gospodarki państwa polskiego po I wojnie światowej i 123 latach zaborów. Pierwsze wyniki opublikowano już w trzy miesiące po spisie, natomiast szczegółowe dane dopiero około sześciu lat później. Informacje obejmujące zagadnienia demograficzne, rynku pracy, zdrowia, edukacji, narodowości, języka i wyznania były istotnym materiałem poznawczym, którego wartość dla zarządzania państwem trudno przecenić.

Dr hab. Jan Paradysz, prof. UEP, reprezentujący US w Poznaniu, w wystąpieniu *Respondent grupowy jako główna przyczyna błędów nielosowych w NSP*



2011 i, nie daj Boże, w NSP 2021 podjął się próby oceny tego rodzaju błędów na tle poprzednich spisów ludności. W erze nowoczesnych spisów powszechnych, do jakich niewątpliwie należał NSP 2011, który szeroko wykorzystywał administracyjne źródła danych i nowoczesne techniki estymacji pośredniej, można nie zauważyć poważnych błędów spowodowanych instytucją respondenta grupowego. Rosnące rozmiary błędów nielosowych wynikają m.in. z coraz większej mobilności przestrzennej (w tym międzynarodowej) społeczeństwa, przybierającego na sile przewrażliwienia osób na punkcie ochrony swoich danych osobowych, ze wzrostu odsetka odmów i innych rodzajów nieobecności podczas spisu, a także z upowszechnienia się nieformalnych związków.

W trakcie tej sesji odbyła się gala wręczenia nagród w konkursie pod patronatem Prezesa GUS na najlepszą pracę magisterską i doktorską z zakresu statystyki obronioną w roku akademickim 2017/18. Pierwsze miejsce w kategorii prac magisterskich zdobyła Justyna Skibińska za pracę *Zastosowanie zaawansowanych metod uczenia maszynowego do detekcji raka trzustki na podstawie obrazowania chemicznego w podczerwieni (IR)*, napisaną pod opieką dr. Tomasza P. Wróbla. Laureatka obroniła się w Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie.

Drugie miejsce zajęła Wioleta Zgliczyńska, której praca *Zastosowanie estymatorów odpornych do oceny ryzyka spółek notowanych na GPW w Warszawie* powstała pod opieką dr. hab. Witolda Orzeszki na Uniwersytecie Mikołaja Kopernika w Toruniu (UMK). Za nią uplasował się Łukasz Leja, autor pracy *Asymetryczna  $\beta$  na rynku nieruchomości – rola czynników egzogenicznych w kształtowaniu anomalii niskiej zmienności funduszy REIT*. Promotorem była dr Małgorzata Snarska, a miejscem obrony – UEK.

Laureatką I nagrody za najlepszą pracę doktorską została Karolina Górna za rozprawę *Metody ekonometrii przestrzennej w analizie konwergencji gospodarczej regionów Unii Europejskiej*, napisaną pod kierunkiem dr hab. Elżbiety Szulc na UMK.

Drugie miejsce zajął Krzysztof Kil z UEK, autor pracy *Stabilność finansowa banków spółdzielczych w Polsce w świetle pokryzysowych zmian regulacyjnych*, której promotorem była dr hab. Ewa Miklaszewska, prof. UEK, zaś promotorem pomocniczym był dr Marcin Idzik. Trzecią nagrodę otrzymała Anna Markowska, której rozprawa *Użyteczność kartograficznych anamorfoz powierzchniowych* powstała na Uniwersytecie Warszawskim, pod opieką dr. hab. Jacka Paślawskiego i promotor pomocniczej dr Izabeli Gołębiowskiej.

Pierwszy dzień konferencji zakończył koncert w wykonaniu skrzypaczki Agnieszki Flis, absolwentki Akademii Muzycznej w Łodzi.

Drugi dzień rozpoczął się dwiema sesjami równoległymi. W sesji toczącej się pod przewodnictwem prof. dr hab. Grażyny Trzpiot z UE w Katowicach dr Bohdan Wyżnikiewicz z GUS wygłosił referat *CESTAT – współpraca statystyków krajów Grupy Wyszehradzkiej przed akcesją do Unii Europejskiej*. Przedstawił

w nim okoliczności powołania w czerwcu 1991 r. przez urzędy statystyczne Węgier, Czechosłowacji i Polski, która była inicjatorem, organizacji współpracy urzędów statystycznych CESTAT (Central European Cooperation in Statistics). Omówił cele działania organizacji, skoncentrowane wokół publikacji kwartalnego biuletynu statystycznego. Do CESTAT sukcesywnie dołączały wszystkie poza Maltą kraje akcesyjne, a organizacja prowadziła działalność do momentu rozszerzenia UE w maju 2004 r.

Następnie Janusz Dygaszewicz w referacie *Współczesne oblicza spisów* omówił różne sposoby przeprowadzania spisów ludności i mieszkań na świecie. Zwrócił uwagę na trzy główne i zgodne z zaleceniami Europejskiej Komisji Gospodarczej (UNECE) kategorie spisu: tradycyjną, opartą na rejestrach i mieszaną, dodając, że państwa członkowskie UE mają możliwość wykorzystania różnych źródeł danych. Prelegent wyraził pogląd, że w planowanej rundzie spisów 2020 większość krajów będzie odchodzić od spisu tradycyjnego. W Polsce planowana jest realizacja pełnego spisu w trybie mieszanym z wykorzystaniem różnych kanałów pozyskiwania danych w obrębie metod CAXI, ze szczególnym uwzględnieniem samospisu internetowego. To podejście pozwoli zebrać komplet danych od respondentów poprzez weryfikację informacji z rejestrów i uzupełnienie tych, których pobranie z rejestrów jest niemożliwe. Nadchodzący spis w 2021 r. będzie zatem dobrą okazją do sprawdzenia mieszanego sposobu realizacji spisu oraz zastanowienia się nad jego znaczeniem w kontekście użyteczności danych spisowych, ponoszonych kosztów i funkcji, jaką sposób mieszany pełni na drodze do przejścia do najbardziej rentownego spisu opartego na rejestrach. Będzie to również sposobność dokonania analizy zalet i wad różnych wariantów zastosowania trybu mieszanego.

W równoległej sesji, prowadzonej przez dr hab. Iwonę Markowicz, prof. US, Tomasz Piasecki z US w Łodzi poruszył, zgodnie z tytułem wystąpienia, *Problem komunikacji wyników obarczonych niepewnością na przykładzie szybkich szacunków dochodu publikowanych przez Eurostat*. Szybkie szacunki wskaźników nierówności dochodowych i ubóstwa wykonywane są przez Eurostat i publikowane jako dane eksperymentalne; wiąże się to z dość długim czasem oczekiwania na pełne wyniki regularnego badania EU-SILC dla wszystkich krajów UE. W referacie przedstawiono ewolucję podejścia do prezentacji wyników w sposób zbliżony do koncepcji przedziału ufności – z uwzględnieniem niepewności wyniku, w sposób przybliżony, bez fałszywego wrażenia o jego dokładności i przy zachowaniu prostoty przekazu, bez ujawniania oszacowania punktowego.

Wystąpienie dr Anny Markowskiej z IGIK pt. *Bitwa map – prezentacja wyników wyborów w wybranych państwach* dotyczyło nowych sposobów prezentowania wyników wyborów. Dawniej wyniki przedstawiano za pomocą kartogramów w przypadku danych względnych lub kartodiagramów w przypadku danych bezwzględnych. Rozwój technologii, w szczególności oprogramowania GIS oraz metod i narzędzi statystycznych, umożliwił zastosowanie bardziej złożonych

form prezentacji danych na mapach statystycznych, np. kartogramów bryłowych (3D) czy anamorfoz powierzchniowych. Stosowanie nietypowych form kartograficznych pozwala zaprezentować te same dane w innym ujęciu, a tym samym przekazać odbiorcom dodatkowe informacje. Na podstawie przeglądu nietypowych form prezentacji wyników wyborów w USA oraz Wielkiej Brytanii opracowano i porównano mapy wyników wyborów prezydenckich w Polsce na poziomie województw, powiatów i gmin.

Ostatnie dwie sesje również przeprowadzono równolegle. W sesji pod przewodnictwem dr. Marka Cierpień-Wolana z US w Rzeszowie dr Maciej Beręsewicz z US w Poznaniu i UEP, dr Tomasz Klimanek i Jakub Ratajczyk z US w Poznaniu w prezentacji *Klasyfikacja produktów na potrzeby koszyka inflacyjnego z wykorzystaniem uczenia maszynowego* poruszyli temat możliwości wykorzystania danych z sieci handlowych, portali internetowych czy e-paragonów do kalkulacji indeksów cen, a w szczególności do pomiaru inflacji. Przedstawili wyniki wstępnych prac badawczych polegających na klasyfikacji różnych nazw produktów stosowanych w ogólnopolskich sieciach handlowych według COICOP (czyli klasyfikacji spożycia indywidualnego według celu) na potrzeby konstrukcji koszyka inflacyjnego. Poruszono kwestie przetwarzania tekstu i zastosowania uczenia maszynowego do automatycznej klasyfikacji produktów. Wykorzystanie algorytmów uczenia maszynowego pozwala na bardzo dobrą kategoryzację produktów po nazwie, zgodnie z COICOP.

Dr hab. Jacek Białek, prof. UŁ, reprezentujący także GUS, i Anna Bobel z GUS wygłosili referat *Dane skanowane w pomiarze CPI – nowe możliwości i wyzwania*. Omówili w nim rolę tego relatywnie nowego rodzaju źródła w pomiarze inflacji. Dostępność danych skanowanych stale wzrasta, a ich główną zaletą jest kompletność – informacje zarówno o cenach produktów, jak i o wartości ich sprzedaży można uzyskać na najniższym poziomie agregacji. Kolejnymi plusami tych danych są m.in. ich względnie niski koszt oraz mnogość obserwacji. Jednakże dane skanowane mają też wady i ograniczenia, a ich ewentualne zastosowanie w pomiarze CPI nastrocza wiele problemów metodologicznych. W prezentacji omówiono także formuły indeksów cen dla danych skanowanych wraz z przykładem empirycznym oraz wyniki dotychczasowych i planowanych prac badawczych.

Wystąpienie dr Magdaleny Mojsiewicz z Uniwersytetu Szczecińskiego oraz US w Szczecinie pt. *Zatrudnieni, pracujący, personel. Wyzwania dla statystyki nauki i techniki* dotyczyło nowego podejścia w statystyce publicznej do pomiaru liczby i czasu pracy osób zaangażowanych w działalność badawczą i rozwojową. Prelegentka omówiła główne wyzwania, przed jakimi stoi statystyka publiczna krajów UE w świetle rekomendacji podręcznika metodycznego *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities*, wydanej przez OECD w 2015 r.

W sesji prowadzonej przez Jana Paradysza Leszek Kozłowski i Agnieszka Piórek z US w Bydgoszcy zajęli się tematem *Administracyjnych źródeł danych w badaniach statystycznych rynku pracy w Polsce*. Autorzy wskazali zarówno na doświadczenia statystyki publicznej w zakresie wykorzystania administracyjnych źródeł danych w badaniach rynku pracy, jak i na wyzwania przed nią stojące. Omówili aktualne uwarunkowania korzystania z rejestrów urzędowych – prawno-organizacyjne, techniczne i metodyczne – w tym rodzaje i strukturę tych rejestrów oraz zakres pozyskiwanych informacji. Wnioski dotyczyły perspektyw wykorzystania administracyjnych źródeł danych w badaniach statystycznych rynku pracy realizowanych przez US w Bydgoszcy oraz możliwych sposobów użycia danych z rejestrów w kolejnych spisach powszechnych.

Temat kontynuowała Renata Rechnio z GUS, która poświęciła swój referat *Wykorzystanie danych administracyjnych w badaniach rynku nieruchomości* doświadczeniom statystyki publicznej w zakresie wykorzystania danych pochodzących z Rejestru Cen i Wartości Nieruchomości (RCiWN) w badaniach rynku nieruchomości. Przedstawiła uwarunkowania sposobu prowadzenia tego rejestru wraz z ich konsekwencjami dla możliwości wykorzystania gromadzonych danych w badaniach statystycznych. Wskazała na szanse i ograniczenia RCiWN jako źródła danych o transakcjach zawieranych na rynku nieruchomości. Ponadto omówiła działania podejmowane przez statystykę publiczną w celu poprawy użyteczności informacji pochodzącej z RCiWN oraz wdrażane procedury przetwarzania danych, których celem jest podniesienie jakości uzyskiwanych wyników. Zarysowała także kierunki dalszego rozwoju statystyki rynku nieruchomości.

Sławomir Konczak (US w Łodzi) w wystąpieniu *TransGUS – dane administracyjne z rozproszonych systemów informacyjnych* przekazał podstawowe informacje na temat pozyskiwania zbiorów danych administracyjnych za pomocą bezpiecznego kanału teleinformatycznego TransGUS. Przedstawił rodzaje danych pozyskiwanych z rozproszonych systemów informacyjnych zgodnie z PBSSP 2018 i PBSSP 2019 oraz harmonogram ich przekazywania, a także trudności napotkane podczas przesyłania danych od gestorów.

Konferencji towarzyszyła wystawa *230 lat Spisów Ludności na Ziemiach Polskich*, eksponująca plakaty spisowe narodowych spisów powszechnych i powszechnych spisów rolnych z lat 1950–2011 oraz publikacje spisowe z ostatnich stu lat, pochodzące ze zbiorów CBS.

Wybrane prezentacje uczestników oraz zdjęcia i podsumowanie z konferencji są dostępne na stronie internetowej [konferencja230.stat.gov.pl](http://konferencja230.stat.gov.pl).

## Wydawnictwa GUS. Czerwiec 2019

W czerwcowej ofercie wydawniczej warto zwrócić uwagę na publikację okolicznościową ***Polska w Unii Europejskiej. Portret statystyczny***, opracowanie cykliczne ***Bilans przepływów międzygałęziowych w bieżących cenach bazowych w 2015 r.*** oraz raport ***Zastosowanie metodologii UE do zdefiniowania obszarów rynku pracy w Polsce***.



***Polska w Unii Europejskiej. Portret statystyczny*** to pozycja wydana z okazji piętnastej rocznicy przystąpienia Polski do Unii Europejskiej (UE). Jubileusz jest dobrą okazją do podsumowania zmian społecznych i makroekonomicznych, jakie w tym czasie zaszły w naszym kraju i innych państwach członkowskich.

Wydawnictwo w oparciu o dane statystyczne opisuje wybrane zagadnienia społeczne, gospodarcze i środowiskowe, ukazując drogę, jaką Polska i UE wspólnie przeszły od 2004 r. Publikacja wskazuje również niektóre wyzwania rozwojowe, jakie stoją dziś przed Polską i innymi krajami Unii.

Przemiany społeczno-gospodarcze przedstawiono na podstawie zasobów informacyjnych gromadzonych w ramach Europejskiego Systemu Statystycznego (ESS). Danym liczbowym i opisom najciekawszych tendencji towarzyszą krótkie informacje o niektórych badaniach prowadzonych w ramach tego systemu. Publikacja stanowi ciekawe źródło informacji na temat integracji europejskiej i zachęca do dalszego zapoznawania się z bogactwem zasobów ESS, którego częścią jest GUS.

Publikację wydano w wersji polsko-angielskiej; jest ona dostępna również na stronie internetowej GUS.



***Bilans przepływów międzygałęziowych w bieżących cenach bazowych w 2015 r.*** to publikacja wydawana co pięć lat, prezentująca wyniki badań z zakresu rachunków narodowych w Polsce. Zawiera informacje o całości zasobów gospodarki polskiej w podziale na pochodzące z produkcji krajowej i z importu oraz o wykorzystaniu tych zasobów (w podziale na zużycie pośrednie, spożycie przez gospodarstwa domowe, instytucje rządowe i samorządowe oraz instytucje niekomercyjne, nakłady brutto na środki trwałe, przyrost rzeczowych środków obrotowych i aktywów o wyjątkowej wartości oraz na eksport).

Tablice prezentowane w publikacji stanowią podstawę do dalszych analiz strukturalnych gospodarki. Obejmują one zarówno strukturę gospodarki, jak i wpływ wywierany na gospodarkę przez zmiany w finalnym popycie. Podstawę, na której zostały opracowane tablice, stanowił Europejski System Rachunków Narodowych i Regionalnych. Dane przedstawione w bilansie przepływów międzygałęziowych są zgodne z odpowiednimi danymi w rachunkach narodowych według sektorów instytucjonalnych. Mogą jednak ulec zmianie po opracowaniu zrewidowanej wersji rachunków narodowych według sektorów instytucjonalnych za 2015 r. i lata następne.

Opracowanie ukazało się w wersji polsko-angielskiej i dostępne jest na stronie internetowej GUS. Wszystkie tablice zamieszczono także w formacie MS Excel.



Raport **Zastosowanie metodologii UE do zdefiniowania obszarów rynku pracy w Polsce** jest rezultatem realizacji grantu Komisji Europejskiej – Eurostatu przez Urząd Statystyczny w Bydgoszczy. Celem projektu było zastosowanie unijnej metodologii do zdefiniowania obszarów rynku pracy w Polsce. Zaproponowano nową metodę wyznaczania optymalnych wartości parametrów wejściowych dla Polski, wykorzystując dorobek polskiej taksonomii.

W raporcie podsumowano wyniki projektu, napotkane problemy i przyszłe wyzwania. Wykorzystując algorytm EU-TTWA wyznaczono 339 obszarów rynku pracy w Polsce, scharakteryzowanych m.in. pod względem liczby pracujących w faktycznym miejscu pracy według płci i sekcji PKD w latach 2011–2014. Rezultaty projektu wskazują, że przebieg granic administracyjnych często różni się od tych wyznaczonych przez czynniki społeczno-gospodarcze.

Publikacja jest dostępna wyłącznie w języku angielskim na stronie Urzędu.

W czerwcu br. ukazały się ponadto:

- *Badanie Aktywności Ekonomicznej Ludności* (folder dla rodzin biorących udział w badaniu aktywności ekonomicznej ludności) – *IV kwartał 2018 r.*;
- *Bezrobocie rejestrowane I kwartał 2019 r.*;
- „Biuletyn statystyczny” nr 5/2019;
- *Ceny robót budowlano-montażowych i obiektów budowlanych* (kwiecień 2019 r.);
- *Ceny w gospodarce narodowej w latach 2013–2017*;
- *Efektywność wykorzystania energii w latach 2007–2017*;
- *Energia 2019* (folder);
- *Koniunktura w przetwórstwie przemysłowym, budownictwie, handlu i usługach 2000–2019* (czerwiec 2019);
- *Nakłady i wyniki przemysłu – I kwartał 2019 r.*;

- *Popyt na pracę w 2018 r.;*
- *Produkcja ważniejszych wyrobów przemysłowych w maju 2019 r.;*
- *Przemysł – wyniki działalności w 2018 r.;*
- *Rachunki kwartalne produktu krajowego brutto w latach 2014–2018;*
- *Sytuacja społeczno-gospodarcza kraju w maju 2019 r.;*
- *Sytuacja społeczno-gospodarcza województw Nr 1/2019;*
- *Turystyka w 2018 r.;*
- *Zatrudnienie i wynagrodzenia w gospodarce narodowej w I kwartale 2019 r.;*
- *Warunki pracy w 2018 r.;*
- *Wybrane wskaźniki przedsiębiorczości w latach 2013–2017;*
- *„Wiadomości Statystyczne” nr 6/2019.*

**Justyna Gustyn** (Główny Urząd Statystyczny)

## Do Autorów To the Authors

(for information go to: [stat.gov.pl/en/journals/statistical-news](http://stat.gov.pl/en/journals/statistical-news))

### **Szanowni Państwo!**

- W „Wiadomościach Statystycznych. The Polish Statistician” publikowane są artykuły o charakterze naukowym poświęcone teorii i praktyce statystycznej, prezentujące wyniki oryginalnych badań teoretycznych lub analitycznych wykorzystujących metody statystyki matematycznej, opisowej lub ekonometrii. W miesięczniku zamieszczane są również artykuły przeglądowe, popularno-naukowe, recenzje publikacji naukowych oraz inne opracowania informacyjne. Prezentowany w artykule naukowym problem badawczy powinien być jednoznacznie zdefiniowany oraz istotny dla oceny zjawisk społecznych lub gospodarczych. Wyniki studiów przeprowadzanych w artykułach winny oddziaływać na rozwój myśli statystycznej oraz edukacji, wnosząc oryginalny wkład do tej dziedziny.

Czasopismo publikuje także artykuły i opracowania prezentujące informacje o teorii i praktyce statystycznej, jak również o problemach edukacji statystycznej. Dotyczą one: programów badań statystycznych statystyki publicznej, systemu zbierania i udostępniania informacji statystycznych, zastosowań informatyki w statystyce, informacji o konferencjach naukowych, działalności organów doradczych Prezesa GUS oraz edukacji statystycznej.

- Artykuły 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 winni wyraźnie określić cel artykułu oraz jasno przedstawić uzyskane wyniki przeprowadzonej analizy. W przypadku prezentacji badań prowadzonych przez autorów należy opisać zastosowaną w nich metodę. Przy prezentacji nowatorskich metod analizy pożądane jest podanie przykładu pokazującego ich zastosowanie w praktyce statystycznej.
- Artykuły zamieszczane w „WS” powinny wyrażać opinie własne autorów. 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 na łamach miesięcznika.
- Po wstępnej ocenie przez redakcję „WS” tematyki artykułu pod względem zgodności z profilem czasopisma i ocenie formalnej artykuły mające charakter naukowy przekazywane są do recenzji osobom specjalizującym się w poszczególnych dziedzinach, które w ocenie kierują się kryterium oryginalności i jakości opracowania, w tym treści i formy, a także potencjalnego zainteresowania czytelników.
- Recenzowanie artykułów odbywa się zgodnie z jednym z dwóch systemów rekomendowanych przez MNiSW – stosowana jest zasada double-blind.



Szczegółowe informacje dotyczące procedury recenzowania, kryteria oceny oraz wzór karty recenzji artykułu znajdują się na stronie internetowej „Wiadomości Statystycznych. The Polish Statistician” w zakładce *Recenzowanie artykułów*.

- Autorzy artykułów, które otrzymały pozytywne recenzje, wprowadzają zasugerowane przez recenzentów poprawki i dostarczają redakcji zaktualizowaną wersję opracowania. Autorzy poświadczają w przysłanym piśmie uwzględnienie wszystkich poprawek. Jeśli pojawi się różnica zdań co do zasadności proponowanych zmian, należy wyjaśnić, które poprawki zostały uwzględnione, a w przypadku ich nieuwzględnienia przedstawić motywy swojego stanowiska.
- Kontroli poprawności stosowanych przez autorów metod statystycznych dokonują redaktorzy statystyczni.

**Redakcja „WS” zastrzega sobie prawo dokonywania w artykułach zmian tytułów, skrótów i przeredagowania tekstu i tablic bez naruszenia zasadniczej myśli autora.**

**Redakcja „WS” oświadcza, że nie wypłaca autorom honorariów za opracowanie artykułów zamieszczanych na łamach naszego czasopisma oraz nie pobiera opłat za ich publikację.**

**Redakcja „WS” informuje, że istnieje możliwość publikacji na łamach miesięcznika artykułów w języku angielskim.**

Uprzejmie informujemy, że od 2007 r. „Wiadomości Statystyczne. The Polish Statistician” znajdują się na liście polskich punktowanych czasopism naukowych Ministerstwa Nauki i Szkolnictwa Wyższego.

W komunikacie MNiSW z dnia 18 grudnia 2015 r. w sprawie wykazu czasopism naukowych wraz z liczbą punktów przyznawanych za publikację w tych czasopismach, miesięcznikowi „Wiadomości Statystyczne. The Polish Statistician” przyznano **12 punktów**.

Miesięcznik posiada konto w Polskiej Bibliografii Naukowej. Jest indeksowany w następujących bazach: Pol-index, CEJSH (Central European Journal of Sciences and Humanities), Agro oraz BazEkon.

### **Zasady etyki publikacyjnej czasopisma „Wiadomości Statystyczne. The Polish Statistician”**

Redakcja „WS” podejmuje wszelkie starania w celu utrzymania najwyższych standardów etycznych obowiązujących w praktyce wydawniczej oraz wykorzystuje wszystkie możliwe środki mające na celu zapobieżenie nadużyciom i nierzetelności autorskiej. Przyjęte zasady postępowania obowiązujące redakcję, autorów, recenzentów i wydawcę przygotowano na podstawie wytycznych Komitetu ds. Etyki Publikacyjnej (COPE), dostępnych na stronie internetowej [www.publicationethics.org](http://www.publicationethics.org).

## Publikacja i autorstwo

- Decyzję o publikacji artykułu podejmuje Kolegium Redakcyjne „WS”. Podstawą tej decyzji jest wynik dyskusji dotyczącej zgłoszonego artykułu, w której uwzględniane są opinie przedstawione w recenzjach, i zależy całkowicie od oceny wartości artykułu, jego oryginalności i jasności przekazu, a także od ścisłego związku z obszarami tematycznymi.
- Artykuły są publikowane po wyrażeniu przez autorów zgody na przeniesienie autorskich praw majątkowych.
- Na autorach spoczywa obowiązek zapewnienia pełnej oryginalności przedłożonych prac, które nie mogą w żadnej części stanowić plagiatu ani zawierać fałszywych lub nieuczciwie podanych informacji. Wykorzystanie w tekście fragmentów dzieł lub stwierdzeń innych autorów należy opatrzyć właściwym przypisem lub oznaczyć jako cytaty.
- Autorzy są zobowiązani do wzięcia udziału w procesie wzajemnej recenzji (peer review).
- Opublikowane artykuły powinny zawierać listę wykorzystanych materiałów oraz (w przypadkach, których to dotyczy) informację na temat dofinansowania badań.
- Autorzy nie mogą składać do publikacji w „WS” artykułów, które zostały wcześniej opublikowane w tym samym kształcie w innym wydawnictwie. Jeżeli doszło do opublikowania podobnych materiałów lub zaprezentowania ich podczas konferencji lub sympozjum naukowego, to podczas składania tekstu autor zobowiązany jest podać ten fakt.
- Redakcja „WS” przestrzega zasady nietolerowania przejawów nierzetelności naukowej autorów artykułów polegającej na:
  - o nieujawnianiu współautorów, mimo że wnieśli oni istotny wkład w powstanie artykułu, określanemu w języku angielskim terminem *ghostwriting*;
  - o podawaniu jako współautorów osób o znikomym udziale lub niebiorących udziału w opracowaniu artykułu, określanemu w języku angielskim terminem *guest authorship*.
- Stwierdzone przypadki nierzetelności naukowej w tym zakresie mogą być ujawniane. W celu przeciwdziałania zjawiskom *ghostwriting* i *guest authorship* należy dołączyć do przesłanego artykułu oświadczenie, którego wzór zamieszczono na stronie internetowej czasopisma (link do załącznika znajduje się w zakładce *Do Autorów*).
- Główną odpowiedzialność za rzetelność przekazanych informacji, łącznie z informacją na temat wkładu poszczególnych współautorów w powstanie artykułu, ponosi zgłaszający artykuł.
- Jeżeli autorzy odkryją w swoim rękopisie lub tekście już opublikowanym błędy, nieścisłości lub niewłaściwe dane, powinni o tym niezwłocznie poinformować redakcję w celu dokonania korekty lub wycofania tekstu.
- Autorzy zobowiązani są do podania wszelkich źródeł finansowania badań będących podstawą treści artykułu lub udziału instytucji naukowo-badawczych, a także do wymienienia publikacji, osób lub ustaleń prawnych, które wpłynęły na charakter i wyniki ich prac.

## Odpowiedzialność redakcji

- Redakcja nie może pozostawać w jakimkolwiek konflikcie interesów w odniesieniu do przyjmowanych artykułów.
- Redakcja jest odpowiedzialna za obsługę procedury wyboru recenzentów; zapewnia przy tym anonimowość i brak konfliktu interesów.
- Redakcja może przekazywać informacje dotyczące rękopisu wyłącznie autorowi, recenzentom, wydawcy lub innym doradcom redakcyjnym.
- W przypadku podjęcia decyzji o niepublikowaniu przesłanego materiału redakcja nie może go w żaden sposób wykorzystać bez pisemnej zgody autora. Jeśli autor postanowi odwołać się od decyzji o nieopublikowaniu jego artykułu, ostateczna decyzja w tej sprawie należy do redaktora naczelnego. Redaktor naczelny może, po skonsultowaniu się z redaktorami i recenzentami oceniającymi daną pracę, zmienić początkową decyzję o niepublikowaniu artykułu.

## Odpowiedzialność recenzentów

- Recenzenci są zobligowani do zachowania obiektywności i poufności oraz powstrzymania się od osobistej krytyki. Recenzenci zawsze powinni uzasadnić swoją ocenę, przedstawiając stosowną argumentację.
- Recenzenci są zobligowani do zadeklarowania, że nie istnieje konflikt interesów w odniesieniu do badań, autorów ani instytucji finansujących badania.
- W ramach przeglądu rękopisu, w stosownych przypadkach, recenzenci powinni wskazać ważne dla wyników badań opublikowane prace, które w ich ocenie powinny zostać przywołane.
- W przypadku stwierdzenia wysokiego poziomu zbieżności treści recenzowanej pracy z innymi opublikowanymi materiałami recenzenci są zobowiązani poinformować o tym redakcję.
- Po ukończeniu recenzji recenzenci powinni zwrócić wszystkie materiały do redakcji. Ich przechowywanie przez recenzentów (w jakiegokolwiek formie) jest niedozwolone.

## Zasady dotyczące procesu wydawniczego

- Wersja elektroniczna czasopisma jest jego wersją pierwotną.
- Dostęp do pełnej zawartości czasopisma jest otwarty.
- Datą publikacji numeru „Wiadomości Statystycznych. The Polish Statistician” w otwartym dostępie w Internecie jest dzienna data jego zamieszczenia na stronie internetowej czasopisma.
- Wersja artykułów zamieszczonych w otwartym dostępie jest ich wersją ostateczną.
- **Materiały zamieszczone w „Wiadomościach Statystycznych. The Polish Statistician” są chronione prawem autorskim. Przedruk tekstu może nastąpić wyłącznie za zgodą redakcji. Treści cytowane z „WS” powinny być opatrzone dokładną informacją o źródle ich pochodzenia.**

- Wydawca deklaruje gotowość do opublikowania poprawek, wyjaśnień oraz przeprosin. Plagiat i świadome preparowanie danych są niedozwolone. Standardy intelektualne i etyczne zawsze mają pierwszeństwo przed potrzebami biznesowymi wydawcy.

### **Informacje dotyczące wymaganej formy oraz kompletności artykułów przesyłanych do „Wiadomości Statystycznych. The Polish Statistician”**

- Artykuły należy przysyłać pocztą elektroniczną pod adresem: **redakcja.ws@stat.gov.pl**.
- Tytuł powinien być podany w językach polskim i angielskim. Konieczne jest dołączenie skróconej informacji (streszczenia) treści artykułu (ok. 10 wierszy) w językach polskim i angielskim. **Streszczenie powinno być utrzymane w formie bezosobowej i zawierać: zwięźle sprecyzowany cel badania, przybliżony jego zakres i przyjętą metodologię oraz ważniejsze wnioski.**
- Prosimy o podanie słów kluczowych, w językach polskim i angielskim, przybliżających zagadnienia w artykule. Słowa kluczowe powinny być zawarte w streszczeniu.
- Prosimy również o podanie kodów klasyfikacji JEL (Journal of Economic Literature).
- Razem z artykułem należy przesłać skan oświadczenia o oryginalności pracy, niezłożeniu jej w innym wydawnictwie i zgodzie na przeniesienie autorskich praw majątkowych, zawierającego też numer ORCID i dane kontaktowe autora oraz afiliację zgłaszanego artykułu.
- **Redakcja rozpoczyna postępowanie kwalifikujące artykuł do opublikowania po spełnieniu przez autora warunku przesłania oświadczenia.**
- Pytania dotyczące przesłanego artykułu, co do jego aktualnego statusu itp., należy kierować do redakcji pod adresem: **redakcja.ws@stat.gov.pl** lub telefonicznie: 22 608 32 25.

### **Wymogi czasopisma dotyczące przygotowania artykułu**

Artykuł powinien mieć optymalną objętość (łącznie z wykresami, tablicami i literaturą) 10–20 stron przygotowanych zgodnie z poniższymi wytycznymi:

1. Tekst zapisany alfabetem łacińskim. Prosimy o transliterację nazw własnych, tytułów itp. oryginalnie zapisanych innym alfabetem.
2. Edytor tekstu – Microsoft Word, format \*.doc lub \*.docx.
3. Czcionka:
  - o tytuł opracowania – Arial, wyśrodkowany, 16 pkt.,
  - o autor – Arial, wersalik, wyrównanie do lewej, 12 pkt.,
  - o tytuły rozdziałów (podrozdziałów) – Arial, wyśrodkowany, 14 (12) pkt.,
  - o tekst główny – Arial, normalny, wyjustowany, 12 pkt.,
  - o przypisy – Arial, 10 pkt.
4. Marginesy przy formacie strony A4 – 2,5 cm z każdej strony.

5. Odstęp między wierszami półtorej linii oraz interlinia przed tytułami rozdziałów.
6. Pierwszy wiersz akapitu wcięty o 0,4 cm, enter na końcu akapitu.
7. Wyszczególnianie rozmaitych kategorii należy zacząć od kropek, a numerowanie od cyfr arabskich.
8. Strony powinny być ponumerowane automatycznie.
9. Wykresy, mapy i schematy powinny być zamieszczone w artykule oraz koniecznie przesłane w oddzielnym oryginalnym pliku, np. Excelu lub Corelu. Należy także przekazać dane, na podstawie których opracowano wykresy i schematy. Wskazówki dotyczące opracowywania map znajdują się w publikacji *Mapy statystyczne. Opracowanie i prezentacja danych*, dostępnej na stronie internetowej GUS w zakładce *Publikacje*: [stat.gov.pl/statystyka-regionalna/publikacje-regionalne/podreczniki-atlasy/podreczniki/mapy-statystyczne-opracowanie-i-prezentacja-danych,1,1.html](http://stat.gov.pl/statystyka-regionalna/publikacje-regionalne/podreczniki-atlasy/podreczniki/mapy-statystyczne-opracowanie-i-prezentacja-danych,1,1.html).
10. Tablice – koniecznie w formie edytowalnej – należy zamieszczać w tekście zgodnie z treścią artykułu. W tablicach nie należy stosować rastrów, cieniowania, pogrubiania czy też podwójnych linii itp.
11. Pod wykresami i tablicami należy podać informacje dotyczące źródła opracowania.
12. Oznaczenia literowe należy wyróżniać następująco: macierze – wersalik, proste, pogrubione (np. **P**, **N<sub>ij</sub>**); wektory – małe litery, kursywa, pogrubione (np. **w**, **x**); pozostałe zmienne – małe lub wielkie litery, kursywa, bez pogrubienia (np. *w*, *x<sub>ij</sub>*, *Z*).
13. Stosowane są skróty: tablica – tabl., wykres – wykr.
14. Przypisy do tekstu należy umieszczać na dole strony.
15. W przypadku cytowania fragmentu pracy należy podać numer strony, z której pochodzi.
16. Wykaz literatury załącznikowej i przytoczenia konkretnych prac w treści artykułu należy przygotować według stylu APA (American Psychological Association).

### **Zasady przywoływania pracy w tekście**

- a. 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, iż... (Iksiński, 2001).

- b. 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, iż... (Iksiński i Nowak, 1999).

- c. 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 można zastosować określenie „i współpracownicy” (w przypadku umieszczenia przywołania nazwisk w strukturze zdania) lub „i in.” (w przypadku gdy nazwiska autorów nie stanowią części struktury zdania).

Przykład zapisu:

Przywołanie po raz pierwszy:

Jak sugerują Nowak, Iksiński i Jankiewicz (2003)...

Badania (Nowak, Iksiński i Jankiewicz, 2003) wskazują, iż...

Kolejne przywołania:

Badania Nowaka i współpracowników (2003)...

Badania te wskazują, iż... (Nowak i in., 2003).

- d. Sześciu i więcej autorów: wymienić należy tylko nazwisko pierwszego autora, zarówno gdy praca przywoływana jest po raz pierwszy, jak i w późniejszych przywołaniach, natomiast pozostałych autorów należy zastąpić skrótem „i in.” (gdy nazwiska nie stanowią części struktury zdania). W literaturze cytowanej należy umieścić nazwiska wszystkich autorów pracy.

Przykład zapisu:

Nowakowski i współpracownicy twierdzą, iż... (1997).

Pierwsze badania na ten temat (Nowakowski i in., 1997) sugerują...

- e. Przywoływanie jednocześnie kilku prac: należy wymienić je alfabetycznie, według nazwiska pierwszego autora. Przywołania kolejnych prac muszą być oddzielone średnikiem i umieszczone w nawiasie. Lata wydania prac tego samego autora/autorów muszą być oddzielone przecinkiem.

Przykład zapisu:

(Iksiński, 2001; Nowak i Iksiński, 1999)

(Iksiński, 1997, 1999, 2004a, 2004b; Nowak i Iksiński, 1999).

- f. Przywoływanie pracy za innym autorem: stosujemy w tekście, natomiast w literaturze cytowanej umieszczamy jedynie pracę czytaną.

Przykład zapisu:

Jak wykazał Nowakowski (1990; za: Zieniecka, 2007)...

Badania sugerują, iż ... (Nowakowski, 1990; za: Zieniecka, 2007).

17. Wykaz literatury powinien być zamieszczony na końcu opracowania. Prace należy zapisać alfabetycznie według nazwiska pierwszego autora. W przypadku dwóch lub więcej prac tego samego autora/autorów należy je uporządkować według roku publikacji. Jeśli kilka prac tego samego autora/autorów zostało opublikowanych w tym samym roku, należy wstawić litery a, b, c itd. po roku publikacji, porządkując prace alfabetycznie według tytułu.

Zapis dotyczący każdej nowej pracy należy zacząć bez wcięcia, wyrównanie do lewego marginesu, a w kolejnych wierszach zapisu stosować wcięcie 0,4 cm.

### **Zasady zapisu literatury załącznikowej**

Poniżej znajdują się schematy zapisów bibliograficznych podstawowych źródeł (artykułów i książek). Sposoby zapisu innych, rzadziej przywoływanych źródeł są szczegółowo opisane w szóstym wydaniu *Publication Manual of the American Psychological Association*.

- a. 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., Nazwisko2, X. Y., Nazwisko3, Z. (rok). Tytuł artykułu. *Tytuł Czasopisma, rocznik* (zeszyt), strona początku–strona końca.
  - b. 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., Nazwisko2, X. Y., Nazwisko3, Z. (rok). Tytuł artykułu. *Tytuł Czasopisma, rocznik*, strona początku–strona końca.
  - c. Jeśli artykuł ma numer DOI (*Digital Object Identifier*), należy podać go na końcu zapisu bibliograficznego:  
Nazwisko, X., Nazwisko2, X. Y. (rok). Tytuł artykułu. *Tytuł Czasopisma, rocznik*, strona początku–strona końca. DOI: xxxxx.
  - d. Książka:  
Nazwisko, X., Nazwisko2, X. Y. (rok). *Tytuł książki*. Miejsce wydania: Wydawnictwo.
  - e. Książka napisana pod redakcją:  
Nazwisko, X. (red.). (rok). *Tytuł książki*. Miejsce wydania: Wydawnictwo.
  - f. Rozdział w pracy zbiorowej:  
Nazwisko, X. (rok). Tytuł rozdziału. W: Y. Nazwisko, B. Nazwisko2 (red.), *Tytuł książki* (s. strona początku–strona końca). Miejsce wydania: Wydawnictwo.
  - g. 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 tekst:  
Nazwisko, X. (rok). *Tytuł tekstu*. Pobrane z: adres strony internetowej.
18. W wykazie literatury należy zamieścić wyłącznie pozycje przytoczone w artykule.
  19. Opracowanie przygotowane w sposób niezgodny z powyższymi wskazówkami będzie odesłane do autora z prośbą o dostosowanie jego formy do wymagań redakcji.

**Zakres tematyczny**  
**działów „Wiadomości Statystycznych. The Polish Statistician”**  
**Thematic scope**  
**of sections of ”Wiadomości Statystyczne. The Polish Statistician”**  
(for information go to: [stat.gov.pl/en/journals/statistical-news](http://stat.gov.pl/en/journals/statistical-news))

**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.

**INFORMACJE. RECENZJE. DYSKUSJE**

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.