Cena zł 12,00
(VAT 5%)

Indeks 381306
PL ISSN 0043-518X

WIADOMOŚCI STATYSTYCZNE

GŁÓWNY URZĄD STATYSTYCZNY

POLSKIE TOWARZYSTWO STATYSTYCZNE

MIESIĘCZNIK
ROK LVIII
WARSZAWA
KWIECIEŃ 2013

4

w numerze m.in.:

HENRYK GURGUL, MARCIN SUDER
Rozkład prawdopodobieństwa dziennych wypłat z bankomatów

TOMASZ GÓRECKI, EWA ŁAŹNIEWSKA
Funkcjonalna analiza składowych głównych PKB

DOMINIK ŚLIWICKI, MAREK RĘKLEWSKI
Ekonometryczna analiza czynników aktywności zawodowej kobiet i mężczyzn

1918 • GŁÓWNY URZĄD STATYSTYCZNY • 2013
95 lat GUS

TPS

KOLEGIUM REDAKCYJNE:

prof. dr hab. Tadeusz Walczak (redaktor naczelny, tel. 22 608-32-89, t.walczak@stat.gov.pl), dr Stanisław Paradysz (zastępca red. nacz.), prof. dr hab. Józef Zegar (zastępca red. nacz., tel. 22 826-14-28), inż. Alina Świderska (sekretarz redakcji, tel. 22 608-32-25, a.swiderska@stat.gov.pl), mgr Jan Berger (tel. 22 608-32-63), dr Marek Cierpiał-Wolan (tel. 17 853-26-35), mgr inż. Anatol Kula (tel. 0-668 231 489), mgr Wiesław Łagodziński (tel. 22 608-32-93), dr Grażyna Marciniak (tel. 22 608-33-54), dr hab. Andrzej Młodak (tel. 62 502-71-16), prof. dr hab. Bogdan Stefanowicz (tel. 0-691 031 698), dr inż. Agnieszka Zgierska (tel. 22 608-30-15)

REDAKCJA

al. Niepodległości 208, 00-925 Warszawa, gmach GUS, pok. 353, tel. 22 608-32-25
http://www.stat.gov.pl/pts/16_PLK_HTML.htm

Elżbieta Grabowska (e.grabowska@stat.gov.pl)

Wersja internetowa jest wersją pierwotną czasopisma.

RADA PROGRAMOWA:

dr Halina Dmochowska (przewodnicząca, tel. 22 608-34-25), mgr Ewa Czumaj, prof. dr hab. Czesław Domański, dr Jacek Kowalewski, mgr Krzysztof Kurkowski, mgr Izabella Zagoździńska

ZAKŁAD WYDAWNICTW STATYSTYCZNYCH

al. Niepodległości 208, 00-925 Warszawa, tel. 22 608-31-45.
Informacje w sprawach nabywania czasopism tel. 22 608-32-10, 608-38-10.
Zbigniew Karpiński (redaktor techniczny), Ewa Krawczyńska (skład i łamanie),
Wydział Korekty pod kierunkiem Bożeny Gorczycy, mgr Andrzej Kajkowski (wykresy).

Indeks 381306

Prenumerata realizowana przez RUCH S.A.:

Zamówienia na prenumeratę w wersji papierowej i na e-wydania można składać bezpośrednio na stronie www.prenumerata.ruch.com.pl
Ewentualne pytania prosimy kierować na adres e-mail: prenumerata@ruch.com.pl lub kontaktując się z Infolinią Prenumeraty pod numerem: 22 693 70 00 — czynna w dni robocze w godzinach 7^{00}—17^{00}.
Koszt połączenia wg taryfy operatora.

Zam. 126/2013 — nakład 440 egz.

NR 4 (623)
KWIECIEŃ 2013

WIADOMOŚCI STATYSTYCZNE

CZASOPISMO GŁÓWNEGO URZĘDU STATYSTYCZNEGO
I POLSKIEGO TOWARZYSTWA STATYSTYCZNEGO

STUDIA METODOLOGICZNE

Henryk GURGUL, Marcin SUDER

Rozkład prawdopodobieństwa dziennych wypłat z bankomatów

Do wczesnych lat 80. ub. wieku system wpłat i wypłat był stosunkowo prosty i składał się z gotówki, czeków, banków oraz przelewów. Wraz z wprowadzeniem kart debetowych i kredytowych ten system zmienił się zasadniczo.

Jednym z istotnych elementów systemu bankowego stały się sieci bankomatów (*Automated Teller Machines* — ATM). Historia bankomatów sięga końca lat 70. XX w., kiedy to pierwszy bankomat udostępniono klientom nowojorskiego oddziału Chemical Bank. Począwszy od lat 80. ub. wieku liczba bankomatów zaczęła stale rosnąć, zaś liczba okienek bankowych odpowiednio spadać.

W ostatnich latach liczba automatów w krajach wysoko rozwiniętych ustabilizowała się, a nawet zaczęła spadać — i to pomimo realnego ciągłego wzrostu gotówki. Przyczyną tego mogą być banki, które są odpowiedzialne za dystrybucję gotówki, ponieważ operują częścią bankomatów. Jednak należy zaznaczyć, że banki bezpośrednio nie korzystają finansowo na dystrybucji gotówki (korzyść odnosi tu bank centralny), wprost przeciwnie — dystrybucja gotówki jest dla banków źródłem poważnych kosztów. Chociaż banki mogą pokryć część tych kosztów poprzez różne opłaty i taryfy, stopa zwrotu netto z gotówki jest ujemna. Co więcej, użycie gotówki zmniejsza depozyty i w ten sposób zmniejszają się zyski banków. Jednak banki nie mogą wstrzymać dostaw gotówki, ponieważ znaczna część klientów ciągle jej jeszcze używa. Jeśliby więc bank nie miał możliwości prowadzenia usług gotówkowych, to mógłby utracić istotną część klientów.

W Polsce pierwszy bankomat pojawił się w drugiej połowie lat 80. ub. wieku w III Oddziale Banku Polska Kasa Opieki SA w Warszawie (dziś Pekao SA). Jednak dopiero w drugiej połowie lat 90. ub. wieku bankomaty zaczęły odgrywać istotną rolę w kształtowaniu całego systemu. Od tego czasu ich liczba zaczęła wyraźnie wzrastać, jednocześnie wzrastała liczba operacji bankomatowych i obroty w bankomatach. Zwiększała się również liczba operatorów bankomatów.

Badania wykazały, że istnieje istotna i odporna zależność między liczbą bankomatów (w relacji do ludności) i liczbą sieci (operatorów) bankomatów. Jedną z nich jest Euronet, który powstał w 1994 r. w Budapeszcie, a już w roku 1996 wprowadził pierwsze bankomaty w Polsce.

Od tego czasu rozwój sieci bankomatowych był bardzo dynamiczny. Na wykresie zobrazowano zmiany liczby bankomatów wszystkich operatorów w Polsce w latach 1999—2011.

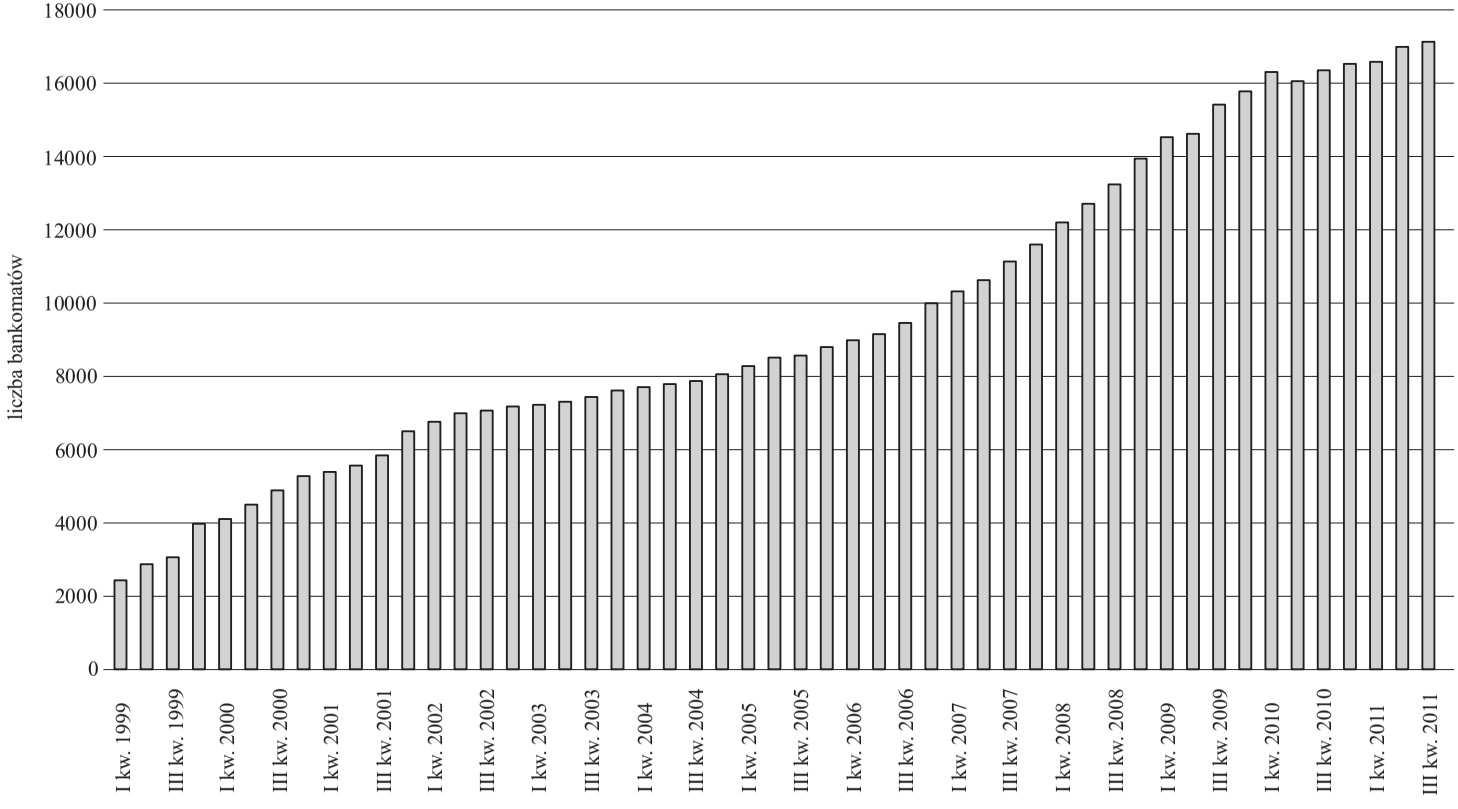
W Polsce działa prawie 18 tys. bankomatów, w których np. w trzecim kwartale 2011 r. dokonano prawie 180 mln operacji (wypłaty, depozyty, przelewy, zakup towarów lub usług) na kwotę ponad 70 mld zł. Są to kwoty, które znacznie przewyższają sumy operacji pieniężnych w oddziałach bankowych.

Tak duża ilość operacji oraz ogromny przepływ gotówki powodują, że zarządzanie obsługą sieci bankomatów ma ogromne znaczenie dla operatora tej sieci. Błędne decyzje związane z zarządzaniem bankomatami, którego elementami są m.in.: lokalizacja bankomatu, wielkość i moment doładowania oraz logistyka konwoju mogą nieść za sobą istotne zwiększenie kosztów obsługi. Operatorzy zatem poszukują rozwiązań mogących zoptymalizować obsługę sieci bankomatów, którymi zarządzają.

Jednym z istotniejszych elementów zarządzania bankomatem lub siecią bankomatów jest prognoza wielkości wypłat z danej maszyny. Możliwość dobrej predykcji wielkości wypłaty z bankomatu pozwala na zbudowanie optymalnego systemu zarządzania minimalizującego koszty obsługi poprzez fakt, że w bankomacie nie będzie za dużo pieniędzy, które nie byłyby w obrocie, a w rezultacie nie przynosiłyby zysku. Ponadto trafne prognozowanie wielkości wypłat minimalizuje ryzyko, że dany bankomat będzie pusty, co mogłoby powodować niezadowolenie klientów.

Z punktu widzenia obniżki kosztów obsługi bankomatów istotną rolę mogą odgrywać także ogólne prognozy zapotrzebowania klientów bankomatów na gotówkę. Prognozy te mogą być użyteczne z punktu widzenia zarządzania gotówką, określania poziomu ryzyka, a przede wszystkim mogą być podstawą przyjęcia przez operatora odpowiedniej strategii napełniania bankomatów gotówką. Są to złożone zagadnienia, bowiem wyniki badań (prognoz) nie mogą być oparte tylko na przeszłych zachowaniach klientów, ale powinny brać pod uwagę dodatkowe czynniki, jak np. aktualną liczbę i strukturę ludności obszaru, na którym znajdują się bankomaty. W modelach badawczych należy uwzględniać nie tylko występowanie składnika systematycznego, ale i efektów losowych.

LICZBA BANKOMATÓW W POLSCE



Ź r ó d ł o: opracowanie własne.

Aby możliwa była predykcja szeregu czasowego, w pierwszym etapie należy poznać własności statystyczne danych empirycznych pochodzących z maszyny bankomatowej.

W artykule podjęto próbę dopasowania rozkładu sumy dziennych wypłat z bankomatów. Autorzy mają na celu zbadanie, czy wybrane przez nich rozkłady mogą skutecznie opisywać strukturę statyczną szeregu wypłat z bankomatów. Dodatkowo sprawdzimy, czy rodzaj rozkładu, jaki dopasowano do danych ma związek z miejscem, w którym dany bankomat się znajduje (województwo, miejscowość, typ lokalizacji).

Określenie rozkładu teoretycznego dla danych empirycznych ma istotne znaczenie w prognozowaniu. Znajomość prawdopodobieństwa wystąpienia danych o niskich oraz wysokich wartościach może istotnie wpływać na jakość prognozy. Niektóre szeregi czasowe wykazują tzw. „efekt kalendarza", tzn. ich wartości mogą zależeć od pory roku, miesiąca, dnia tygodnia, pory dnia itd. W sytuacji gdy szereg czasowy jest nim obarczony, to jeśli znany jest teoretyczny rozkład prawdopodobieństwa dla analizowanych danych, jesteśmy w stanie zweryfikować, czy prognozy „odstające" są dopuszczalne czy też należy je skorygować.

Dodatkowo przedstawimy i porównamy wyniki badań podstawowych własności statystycznych wielkości wypłat z bankomatów według różnych typów ich lokalizacji.

PRZEGLĄD LITERATURY

Zagadnienie optymalizacji zarządzania siecią bankomatową jest problemem dość znanym. Istnieje wiele firm oferujących rozwiązania w tym zakresie. Jednak w literaturze przedmiotu trudno napotkać na publikacje naukowe związane z badaniami za pomocą modeli ekonometrycznych wielkości wypłat gotówkowych z bankomatów. Wynika to głównie z braku dostępu badaczy do tego typu danych. Operatorzy bankomatów, szukając optymalnych rozwiązań w zakresie zarządzania, zwracają się zazwyczaj do firm komercyjnych, a nie do pracowników naukowych. W związku z tym wszelkie analizy tych szeregów są utajnione przez te firmy.

Jedną z najbardziej trafnych prac statystycznych na omawiany temat jest analiza transakcji za pomocą kart kredytowych autorstwa Handa i Blunt (2001). Jednakże w literaturze ekonomicznej można znaleźć też pozycje zajmujące się związkiem pomiędzy liczbą bankomatów i popytem na gotówkę. Amromin i Chakravorti (2007) przeprowadzili analizę za pomocą regresji liniowej danych panelowych w celu porównania krajów, głównie pod kątem wzrostu transakcji przy użyciu kart debetowych w punktach sprzedaży. Boeschoten (1998) oraz Snellman i Viren (2006) mieli podobny cel, ale swoje wnioski sformułowali na podstawie wyników zastosowania modeli opisujących indywidualne zachowania konsumenta.

Boeschoten analizował dane dotyczące konsumentów holenderskich z okresu 1990—1994 stosując deterministyczny model inwentarzowy. Zakładano w nim, że klienci utrzymują stały poziom gotówki i zużywają ją ze stałą stopą, przy czym, jeśli jej ilość spadnie poniżej określonego poziomu, uzupełniają zasoby gotówkowe. Wymienieni badacze wykazali, że klienci bankomatów mają na ogół niższe zapasy gotówki niż osoby niekorzystające z bankomatów, ponieważ koszt uzupełnienia gotówki przez tych pierwszych jest na ogół niższy niż tych drugich.

Związek pomiędzy kosztem uzyskania gotówki i liczbą bankomatów modelowali wspomniani Snellman i Viren w postaci deterministycznego problemu optymalizacyjnego. Zakładali oni, że koszty są proporcjonalne do odległości od bankomatu. Z kolei Simutis i in. (2007) podjęli próbę optymalizacji zarządzania gotówką w sieci bankomatowej poprzez wybór odpowiedniego modelu predykcji wypłat. Autorzy ci wykazują, że zastosowanie sieci neuronowych do prognozowania wypłat z bankomatów pozwala na obniżenie kosztów obsługi sieci bankomatowej o ok. 15—20%. Z kolei Armenise i in. (2010) jako metodę prognozowania wielkości wypłat z bankomatów zaproponowali zastosowanie algorytmów genetycznych.

Pomimo tego, że wymienieni naukowcy stosowali modele wypłat gotówki przez klientów bankomatów z uwzględnieniem czasu, to jednak przyjęte przez nich uproszczenia uniemożliwiały sformułowanie wniosków na poziomie makroekonomicznym, jeśli chodzi o łączne zapotrzebowanie na gotówkę. Autorzy podkreślili, że wnioski mogłyby być uściślone, gdyby badania były przeprowadzone na poziomie indywidualnych wypłat.

W badaniu zapotrzebowania na gotówkę w bankomatach mogą być istotne takie aspekty, jak ilość i wielkość wypłat, ich rozkłady empiryczne, autokorelacja i efekty czasowe. Istotną rolę w badaniach odgrywają rozkłady. W literaturze dotyczącej rozkładów wypłat występują generalnie dwa typy. Pierwszy rodzaj to rozkłady opisujące sytuację, w której wypłata nie zależy od historii wypłat (rozkład wielomianowy oraz rozkład Dirichleta efektu losowego). W drugim typie modeli nie specyfikuje się rozkładu składnika losowego, lecz aproksymuje się go za pomocą rozkładu empirycznego.

W przypadku różnych danych mających postać szeregu czasowego często zakłada się, że rozkład normalny w przybliżeniu dobrze pasuje do tych danych. Na tym założeniu oparte są m.in. modele ekonometrii finansowej, np. Blacka-Scholesa (Black, Scholes, 1973) służący do wyceny opcji, model Markowitza budowy portfela, model CAPM (*Capital Asset Pricing Model*) zaproponowany przez Sharpe'a (1964) czy też metody analizy technicznej. Jednak w późniejszych pracach, m.in. Mandelbrot (1963) oraz Fama (1965), wykazali, że założenie o normalności rozkładu ekonomicznych szeregów czasowych jest często nieprawdziwe.

Rozwój techniki komputerowej pozwalający na szybką analizę danych przyczynił się do wykazania, że dane mające postać szeregu czasowego odznaczają się większą kurtozą niż to jest w przypadku rozkładu normalnego, a ponadto są

asymetryczne. Wykazano też, że prawdopodobieństwo występowania wartości ekstremalnych jest dużo większe niż w rozkładzie gaussowskim (tzw. „grube ogony"). Odrzucenie normalności rozkładu wielu finansowych szeregów czasowych skłoniło badaczy do poszukiwania nowych rozkładów teoretycznych, które dobrze opisywałyby zachowanie się danych pochodzących z rynków finansowych.

Pomimo dość obszernej literatury dotyczącej tego tematu (zwłaszcza obcojęzycznej), badania opisywane w pracach naukowych zazwyczaj dotyczyły głównie indeksów lub spółek notowanych przede wszystkim na różnych giełdach światowych. Ponadto w większości przypadków autorzy poprzestawali na stwierdzeniu braku normalności tychże rozkładów lub na zbadaniu ich zgodności tylko z pewną grupą wspomnianych rozkładów.

W Polsce nie zostały dotąd przeprowadzone badania własności statystycznych wielkości wypłat z bankomatów. Autorom nie są również znane prace na ten temat w literaturze zagranicznej.

STRUKTURA DANYCH I ICH WŁASNOŚCI STATYSTYCZNE

W badaniach wykorzystano dane z 293 bankomatów firmy Euronet. Jest to największy spośród 28 działających operatorów sieci bankomatów w Polsce. Firma obsługuje prawie 25% bankomatów w Polsce, a w jej posiadaniu jest ok. 3 tys. bankomatów.

Do analizy uzyskano dane w postaci sumy dziennych wypłat ze wszystkich bankomatów znajdujących się w województwach małopolskim i podkarpackim. W przypadku bankomatów, które funkcjonowały przed 2008 r. wykorzystano dane od 1 stycznia 2008 r. Z kolei dla pozostałych bankomatów uwzględniono w badaniu dane od początku funkcjonowania bankomatu. Ostatnia obserwacja dotycząca wszystkich bankomatów pochodzi z 30 listopada 2011 r. Długość analizowanych szeregów czasowych kształtuje się od 280 do 1429, w zależności od tego, jak długo dany bankomat funkcjonował.

Analizowane dane są dobowymi sumami wypłat z danej maszyny, przy czym usunięto dane o wartości 0, czyli takie, które oznaczają, że bankomat był niedostępny (zepsuty lub w miejscu niedostępnym w danym dniu, np. nieczynny hipermarket).

Każdy bankomat opisano trzema cechami: województwo, liczba mieszkańców miejscowości (kategoria miejscowości) oraz typ lokalizacji. Jako typ lokalizacji Euronet wprowadza 15 kategorii: lotnisko, oddział bankowy, dworzec autobusowy, centrum rozrywki, zakład przemysłowy, hipermarket, hotel, stacja benzynowa, dworzec kolejowy, osiedle, restauracja, sklep, centrum handlowe, węzeł komunikacyjny, pozostałe. Miało też miejsce łączenie grup lokalizacyjnych w tych przypadkach, gdy liczebność bankomatów była zbyt mała, aby miało sens porównywanie z innymi lokalizacjami. Bankomaty, które znajdują się na lotnisku, dworcach kolejowym oraz autobusowym, a także w węźle komunika-

cyjnym zgrupowano w kategorię „transport osób". Niska liczba bankomatów zlokalizowanych w zakładach przemysłowych oraz na osiedlach była przyczyną zaliczenia tych bankomatów do kategorii „pozostałe". Dodatkowo każda z miejscowości, w której znajdują się bankomaty, została przydzielona do jednej z kategorii w zależności od liczby mieszkańców miejscowości, w której bankomat się znajdował:

- 0— 20 tys. — wsie oraz małe miasteczka — kategoria I,
- 21— 50 tys. — miasta średnie — kategoria II,
- 51—100 tys. — miasta duże — kategoria III,
- 101—200 tys. — miasta bardzo duże (Rzeszów, Tarnów) — kategoria IV,
- powyżej 200 tys. — metropolia Kraków — kategoria V.

Analizowane cechy pozwalają na zweryfikowanie hipotez związanych z tym, że wielkości wypłat z bankomatów mogą zależeć od ogólnie pojętego umiejscowienia bankomatu (województwo, miejscowość, lokalizacja).

TABL. 1. LICZBA BADANYCH BANKOMATÓW WEDŁUG LOKALIZACJI ORAZ WOJEWÓDZTW (stan na 1 marca 2011 r.)

Typ lokalizacji	Małopolskie		Podkarpackie		Razem	
	liczba bankomatów	w % ogólnej liczby bankomatów w województwie	liczba bankomatów	w % ogólnej liczby bankomatów w województwie	liczba bankomatów	w % ogólnej liczby bankomatów badanych województw
R a z e m	**232**	**79,18**	**61**	**20,82**	**293**	**100,00**
Oddział bankowy	57	24,57	27	44,26	84	28,67
Centrum rozrywki	6	2,59	—	—	6	2,05
Hipermarket	32	13,79	8	13,11	40	13,65
Hotel	6	2,59	—	—	6	2,05
Stacja paliw	23	9,91	2	3,28	25	8,53
Sklep	41	17,67	8	13,11	49	16,72
Centrum handlowe	30	12,93	12	19,67	42	14,33
Transport osób	4	1,72	—	—	4	1,37
Pozostałe	33	14,22	4	6,56	37	12,63

Ź r ó d ł o: opracowanie własne.

Z tabl. 1 wynika, że 232 bankomaty zlokalizowano w Małopolsce, natomiast 61 w woj. podkarpackim. Rozkład ze względu na typ lokalizacji również nie jest równomierny. Ponad 28% bankomatów było w oddziałach bankowych, przy czym w badanych województwach występuje zróżnicowanie w tym rozkładzie. W woj. podkarpackim ponad 44% analizowanych bankomatów było usytuowanych w oddziale bankowym, a w woj. małopolskim prawie co czwarty bankomat znajdował się w banku. Na kolejnych miejscach — jeśli chodzi o liczbę bankomatów — uplasowały się sklepy, centra handlowe, hipermarkety. Jednak w przypadku tych lokalizacji nie ma aż tak dużych dysproporcji między województwami.

Podobnego podziału dokonano ze względu na liczbę mieszkańców (tabl. 2).

TABL. 2. LICZBA BADANYCH BANKOMATÓW WEDŁUG LOKALIZACJI ORAZ KATEGORII MIEJSCOWOŚCI (stan na 1 marca 2011 r.)

Typ lokalizacji	Kategoria miejscowości				
	I	II	III	IV	V
	Liczba bankomatów				
R a z e m	**15**	**47**	**25**	**39**	**167**
Oddział bankowy	2	17	13	14	38
Centrum rozrywki	1	2	1	—	2
Hipermarket	3	2	5	5	25
Hotel	—	1	—	—	5
Stacja paliw	2	4	1	1	17
Sklep	5	10	—	4	30
Centrum handlowe	1	6	5	10	20
Transport osób	—	1	—	1	2
Pozostałe	1	4	—	4	28
	W % ogólnej liczby bankomatów w miejscowości				
R a z e m	**5,12**	**16,04**	**8,53**	**13,31**	**57,00**
Oddział bankowy	13,33	36,17	52,00	35,90	22,75
Centrum rozrywki	6,67	4,26	4,00	—	1,20
Hipermarket	20,00	4,26	20,00	12,82	14,97
Hotel	—	2,13	—	—	2,99
Stacja paliw	13,33	8,51	4,00	2,56	10,18
Sklep	33,33	21,28	—	10,26	17,96
Centrum handlowe	6,67	12,77	20,00	25,64	11,98
Transport osób	—	2,13	—	2,56	1,20
Pozostałe	6,67	8,51	—	10,26	16,77

Ź r ó d ł o: jak przy tabl. 1.

Można zauważyć, że większość badanych bankomatów znajdowała się w Krakowie (kategoria V), natomiast najmniejszy odsetek bankomatów odnotowano w miejscowościach o liczbie mieszkańców poniżej 20 tys. Również podział bankomatów w poszczególnych grupach kategorii miejscowości był różny. W najmniejszych miejscowościach przeważały bankomaty w sklepach, natomiast w pozostałych miejscowościach najwięcej bankomatów znajdowało się w oddziałach bankowych.

W celu lepszego poznania struktury danych przeanalizowano podstawową statystykę dotyczącą wszystkich bankomatów łącznie, a następnie podobnej analizy dokonano dla bankomatów należących do tych samych grup, porównując jednocześnie wyniki w poszczególnych grupach.

ZESTAWIENIE (1) STATYSTYKI ŚREDNICH DZIENNYCH WYPŁAT ZE WSZYSTKICH BANKOMATÓW

Średnia	Odchylenie standardowe	Współczynnik zmienności w %	Minimum	Maksimum
46857,7	27673,9	59,06	2755	161600

Ź r ó d ł o: jak przy tabl. 1.

W zestawieniu (1) przedstawiono podstawową statystykę dotyczącą średnich dziennych wypłat ze wszystkich bankomatów. Szczególną uwagę zwraca umiarkowana zmienność średnich. Współczynnik zmienności na poziomie ok. 60% świadczy o tym, że zróżnicowanie średnich wypłat z bankomatów jest dosyć duże.

W zestawieniu (2) przedstawiono analogiczną analizę współczynników zmienności ze wszystkich bankomatów wyrażonych w ułamkach dziesiętnych.

ZESTAWIENIE (2) STATYSTYKI WSPÓŁCZYNNIKA ZMIENNOŚCI WYPŁAT ZE WSZYSTKICH BANKOMATÓW

Średnia w %	Odchylenie standardowe	Współczynnik zmienności	Minimum	Maksimum
		w %		
201,91	0,471456	23,35	76	335

Ź r ó d ł o: jak przy tabl. 1.

Wyniki tej analizy wskazują na dużą zmienność w strukturach danych poszczególnych bankomatów. Średni współczynnik zmienności kształtował się na poziomie 200%. Jednak jego odchylenie standardowe jest nieduże, co świadczy o tym, że generalnie zmienność w przypadku wszystkich bankomatów jest zbliżona.

W analizie skośności poszczególnych bankomatów istotna jest liczba bankomatów o asymetrii lewo- i prawostronnej. W przypadku kurtozy podano liczbę bankomatów, dla których kurtoza jest wyższa (niższa) od tej dla rozkładu normalnego.

ZESTAWIENIE (3) WARTOŚCI WSPÓŁCZYNNIKÓW ASYMETRII ORAZ KURTOZY WYPŁAT ZE WSZYSTKICH BANKOMATÓW

Asymetria		Kurtoza	
lewostronna	prawostronna	ujemna	dodatnia
31 (10,6%)	262 (89,4%)	112 (38,2%)	181 (61,8%)

Ź r ó d ł o: jak przy tabl. 1.

W przypadku większości bankomatów uzyskane wartości współczynnika asymetrii wskazują na umiarkowaną lub dużą siłę asymetrii w badanych szeregach, co stanowi podstawy do testowania zgodności z rozkładami asymetrycznymi. Również wartości kurtozy prawie wszystkich bankomatów istotnie odbiegają od wartości charakterystycznej dla rozkładu normalnego, co również skłania do zastosowania rozkładów teoretycznych, których funkcja gęstości ma inny kształt niż w przypadku rozkładu gaussowskiego. Wyniki analizy przedstawiono w zestawieniu (3) wskazując, że znaczna większość danych bankomatowych charakteryzuje się prawostronną asymetrią. Większa część bankomatów charakteryzuje się wyższą kurtozą niż w przypadku danych z rozkładu normalnego.

W kolejności przedstawimy analogiczne analizy z uwzględnieniem odpowiednich grup związanych z lokalizacją.

W tabl. 3 zamieszczono podstawowe dane dotyczące przeciętnych dziennych wypłat. Wyniki wskazują, że średnia przeciętnych dziennych wypłat z bankomatów w woj. małopolskim była znacznie wyższa (o ponad 12 tys. zł) niż na Podkarpaciu. Jednocześnie zmienność tych średnich jest praktycznie identyczna. Analiza wariancji (*p-value* = 0,005 dla testu *F*) wykazała istotność w zróżnicowaniu średniej z przeciętnych wypłat w tych województwach.

TABL. 3. STATYSTYKA ŚREDNICH WYPŁAT Z BANKOMATÓW WEDŁUG WOJEWÓDZTW

Województwa	Średnia	Odchylenie standardowe	Współczynnik zmienności w %	Minimum	Maksimum
Małopolskie	49370,9	28615,3	57,96	2755	161600
Podkarpackie	37299,1	21384,0	57,33	6083	104000

Ź r ó d ł o: jak przy tabl. 1.

Informacje zawarte w tabl. 4 wskazują, że rodzaj miejscowości, w jakiej znajduje się bankomat, określony przez liczbę mieszkańców, ma związek ze średnią wielkością wypłat z bankomatów (*p-value* = 0 dla testu *F*). Nie jest to jednak zależność proporcjonalna. Można bowiem zauważyć, że najwyższa średnia wypłat była w najmniejszych miejscowościach. Nieco mniejszą średnią uzyskano w Krakowie, a najmniejszą w miejscowościach o liczbie mieszkańców od 50 tys. do 100 tys. Aby wyniki te można było poprawnie zinterpretować, należałoby zbadać dodatkowo liczbę bankomatów przypadających na mieszkańca w poszczególnych miejscowościach. Niestety nie jest to możliwe z powodu zbyt dużej zmienności liczby bankomatów na danym terenie.

TABL. 4. STATYSTYKA ŚREDNICH WYPŁAT Z BANKOMATÓW WEDŁUG KATEGORII MIEJSCOWOŚCI

Kategoria miejscowości	Średnia	Odchylenie standardowe	Współczynnik zmienności w %	Minimum	Maksimum
I	56480,0	19174,5	33,95	15260	87390
II	39671,3	22249,0	56,08	9623	124600
III	34225,6	17726,2	51,79	7011	89800
IV	42640,6	22711,6	53,26	6083	104000
V	50891,7	30787,8	60,50	2755	161600

Ź r ó d ł o: jak przy tabl. 1.

Na podstawie danych zawartych w tabl. 8 można stwierdzić, że najwięcej wypłacono z bankomatów znajdujących się w hipermarketach, centrach handlowych oraz na stacjach paliw. Przeciętnie najmniejsze wypłaty odnotowano w hotelach oraz oddziałach bankowych. Również w tym przypadku wpływ lokalizacji na wielkość dziennych wypłat z bankomatów jest istotny (*p-value* = 0,0023 dla testu *F*).

TABL. 5. STATYSTYKA ŚREDNICH WYPŁAT Z BANKOMATÓW WEDŁUG TYPU LOKALIZACJI

Typ lokalizacji	Średnia	Odchylenie standardowe	Współczynnik zmienności w %	Minimum	Maksimum
Oddział bankowy	30085,8	13460,7	44,74	2755	62100
Centrum rozrywki	47528,3	26451,9	55,66	10550	74730
Hipermarket	69484,8	32145,0	46,26	13570	155900
Hotel	26792,5	25503,8	95,19	5855	71060
Stacja paliw	60170,4	33343,4	55,41	18940	161600
Sklep	44952,4	19649,8	43,71	19310	100300
Centrum handlowe	61748,6	32066,9	51,93	7011	142100
Transport osób	38072,5	18725,0	49,18	17080	54790
Pozostałe	41192,2	17611,9	42,76	14990	90970

Źródło: jak przy tabl. 1.

Podsumowując tę część analizy można wskazać, że średnia wielkość wypłat jest istotnie uzależniona od tego, w którym województwie, w jakiej miejscowości oraz w jakiej lokalizacji znajduje się bankomat.

Przeprowadzono analogiczną analizę w zakresie współczynników zmienności, skośności oraz asymetrii. Wyniki wskazują, że również one przyjmują zróżnicowane wartości w zależności od zlokalizowania bankomatu.

Uzyskane wyniki mogą stać się podstawą do postawienia hipotezy, że własności statystyczne sumy dziennych wypłat z bankomatów są uzależnione od lokalizacji bankomatu. Istnieją powody, aby przypuszczać, że rodzaj rozkładu, który można dopasować do danych, również może być uzależniony od tej lokalizacji.

PRZEGLĄD ROZKŁADÓW ORAZ METODOLOGIA BADAŃ

W tej części artykułu scharakteryzujemy rozkłady teoretyczne, które zostały użyte w celu dopasowania do danych. W szczególności przedstawimy funkcje gęstości rozkładów wraz z opisem ich parametrów.

Rozkład logistyczny. Ten rozkład teoretyczny zaproponował w 1981 r. Smith (1981). Testowali go też m.in. Gray i French (1990), Peiró (1994), Aparicio i Estrada (2001) oraz Suder i Wójtowicz (2005). Jest to rozkład podobny do rozkładu normalnego, a główną różnicę stanowią „grube ogony". Funkcja gęstości rozkładu logistycznego dana jest wzorem:

$$f(x) = \frac{\exp\left(-\frac{x-\mu}{\alpha}\right)}{\alpha \cdot \left[1+\exp\left(-\frac{x-\mu}{\alpha}\right)\right]^2}$$

gdzie:

μ — średnia,
α — parametr skali,

przy czym dla wariancji prawdziwa jest formuła $\sigma^2 = \frac{\pi^2 \alpha^2}{3}$.

Rozkład skalowany *t*-Studenta. W pracach (Praetz, 1972; Blattberg, Gonedes, 1974; Peiró, 1994; Aparicio, Estrada, 2001; Broca, 2002; Suder, Wójtowicz, 2005) pokazano, że ten rozkład dobrze aproksymuje np. rozkłady empiryczne dziennych stóp zwrotu:

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sigma \cdot \Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}} \cdot \left[1 + \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{v}\right]^{-\frac{v+1}{2}}$$

gdzie:

$\Gamma(\cdot)$ — funkcja gammy Eulera,
$\mu \in (-\infty, \infty)$ — parametr położenia,
$\sigma^2 \neq 0$ — parametr rozproszenia,
$v > 0$ — parametr odpowiedzialny za kształt.

Jeśli zmienna losowa X ma rozkład skalowany *t*-Studenta o parametrach μ, σ oraz v, wówczas zmienna $\frac{x-\mu}{\sigma}$ ma rozkład *t*-Studenta z liczbą stopni swobody wynoszącą v.

Rozkład potęgowo-wykładniczy. Z uwagi na „grube ogony" i występujące w skalowanym rozkładzie *t*-Studenta wysokie prawdopodobieństwo małych zmian, Hsu (1982), Gray i French (1990) oraz Broca (2002) zaproponowali stosowanie w badaniu rozkładów stóp zwrotu rozkładu potęgowo-wykładniczego. Funkcja gęstości rozkładu potęgowo-wykładniczego rzędu β dana jest wzorem:

$$f(x) = \frac{1}{2\beta^{\frac{1}{\beta}} \cdot \sigma \cdot \Gamma\left(1 + \frac{1}{\beta}\right)} \exp\left(-\left|\frac{x-\mu}{\beta\sigma^{\beta}}\right|^{\beta}\right)$$

gdzie $\mu \in (-\infty, \infty)$, $\sigma > 0$ oraz $\beta \in (0, \infty)$ są parametrami odpowiednio: położenia, rozrzutu i kształtu.

Jeśli $\beta < 2,$ to rozkład posiada „grube ogony”, gdy $\beta = 2,$ to rozkład jest normalny, a dla $\beta > 2$ mamy do czynienia z rozkładem o „wąskich ogonach”. W przypadku gdy $\beta = 1$ otrzymujemy rozkład Laplace'a, natomiast przy $\beta \to \infty$ otrzymujemy rozkład jednostajny.

Mieszanka dwóch rozkładów normalnych. Innym, używanym w literaturze rozkładem teoretycznym (Press, 1967; Kon, 1984; Peiró, 1994; Aparicio, Estrada, 2001) jest mieszanka rozkładów normalnych. Funkcja gęstości mieszanki dwóch rozkładów normalnych ma postać:

$$f(x) = \lambda \cdot \frac{1}{\sigma_1^2 \sqrt{2\pi}} \exp\left(\frac{-(x-\mu_1)}{2\sigma_1^2} \right) + (1-\lambda) \frac{1}{\sigma_2^2 \sqrt{2\pi}} \exp\left(\frac{-(x-\mu_2)}{2\sigma_2^2} \right)$$

Parametry $-\infty < \mu_i < \infty$ oraz $\sigma_i > 0$ (dla i=1, 2) opisują średnią i odchylenie standardowe każdego z dwóch rozkładów mieszanki. Parametry λ oraz $1-\lambda$ są wagami odpowiednich rozkładów w mieszance $(\lambda \in (0,1)$.

Mieszanka trzech rozkładów normalnych. Jest to rozkład znacznie rzadziej stosowany w badaniach niż mieszanka dwóch rozkładów normalnych. Jednak zastosowanie mieszanki więcej niż dwóch rozkładów można znaleźć m.in. w pracach: Baringhaus i Henze (1988), Henze (2002), Rencher (1995). Funkcja gęstości mieszanki trzech rozkładów normalnych ma postać:

$$f(x) = \lambda_1 \cdot \frac{1}{\sigma_1^2 \sqrt{2\pi}} \exp\left(\frac{-(x-\mu_1)}{2\sigma_1^2} \right) + \lambda_2 \frac{1}{\sigma_2^2 \sqrt{2\pi}} \exp\left(\frac{-(x-\mu_2)}{2\sigma_2^2} \right) +$$
$$+ (1-\lambda_1-\lambda_2) \frac{1}{\sigma_3^2 \sqrt{2\pi}} \exp\left(\frac{-(x-\mu_3)}{2\sigma_3^2} \right)$$

Parametry $-\infty < \mu_i < \infty$ oraz $\sigma_i > 0$ (dla i=1, 2, 3) opisują średnią i odchylenie standardowe każdego z trzech rozkładów mieszanki. Parametry λ_1, λ_2 oraz $1-\lambda_1-\lambda_2$ są wagami odpowiednich rozkładów w mieszance $(\lambda_1, \lambda_2 \in (0,1),\ \lambda_1 + \lambda_2 < 1)$.

Rozkład hiperboliczny. Funkcja gęstości tego rozkładu dana jest wzorem:

$$f(x) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha \cdot \delta \cdot K_1\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} \exp\left\{ -\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu) \right\}$$

gdzie:

$\alpha(\alpha > 0)$ oraz $\beta(0 \le |\beta| < \alpha)$ — parametry kształtu,
$\delta(\delta > 0)$ — parametry skali,
$\mu(\mu \in R)$ — odpowiada za położenie,
$K_1(\cdot)$ — zmodyfikowana funkcja Bessela trzeciego rodzaju z indeksem 1.

Rozkład hiperboliczny zawiera jako przypadki graniczne m.in. rozkłady gaussowski (gdy $\alpha \to \infty, \delta \to \infty$) oraz Laplace'a (dla $\delta = 0$).

Rozkład normalny odwrotny gaussowski (NIG). Funkcja gęstości tego rozkładu dana jest wzorem:

$$f(x) = \frac{\alpha \cdot \delta}{\pi} \exp\left\{\delta\sqrt{\alpha^2 - \beta^2} + \beta(x-\mu)\right\} \frac{K_1\left(\alpha \cdot \sqrt{\delta^2 + (x-\mu)^2}\right)}{\sqrt{\delta^2 + (x-\mu)^2}}$$

Parametry tego rozkładu mają analogiczną interpretację, jak parametry rozkładu hiperbolicznego. Rozkład hiperboliczny oraz rozkład normalny odwrotny gaussowski należą do grupy tzw. uogólnionych rozkładów hiperbolicznych. Ich zastosowanie pokazują m.in. Eberlein i Keller (1995), Küchler i in. (1999) oraz Suder i Wójtowicz (2005).

Rozkład α-stabilny. W przeciwieństwie do omówionych rozkładów jest on zdefiniowany za pomocą funkcji charakterystycznej danej wzorem:

$$\Phi(t) = \begin{cases} \exp\left\{-\gamma^{\alpha}|t|^{\alpha}\left(1 - i\beta\,\mathrm{sgn}(t)\,tg\frac{\pi\alpha}{2}\right) + i\delta t\right\} & \alpha \neq 1 \\ \exp\left\{-\gamma|t|\left(1 + i\beta\frac{2}{\pi}\mathrm{sgn}(t)\ln|t|\right) + i\delta t\right\} & \alpha = 1 \end{cases}$$

Rozkład ma cztery parametry, które są związane z czterema pierwszymi momentami rozkładu. Parametr α ($0<\alpha\le 2$) jest nazywany indeksem stabilności rozkładu i określa „grubość ogona". Jeśli $\alpha=2$, to zmienna ma rozkład normalny, gdy $\alpha<2$, to rozkład ma „ogony grubsze" niż występujące w rozkładzie normalnym. $\beta(0\le|\beta|\le 1)$ jest parametrem skośności, natomiast $\delta(-\infty<\delta<\infty)$ i $\gamma(\gamma>0)$ są parametrami odpowiednio położenia i skali. Dla $\alpha = 2$ i $\beta = 0$ rozkład α-stabilny staje się rozkładem normalnym, dla $\alpha = 1$ i $\beta = 0$ jest rozkładem Cauchy'ego, natomiast przy $\alpha = \frac{1}{2}$ i $\beta = 0$ równy jest rozkładowi Levy'ego.

W celu zbadania zgodności rozkładu empirycznego sumy dziennych wypłat z poszczególnych bankomatów z analizowanymi rozkładami teoretycznymi za-

stosowano najbardziej popularne testy zgodności, tj. chi-kwadrat, Kołmogorowa-Smirnowa oraz Andersona-Darlinga.

Dla testu chi-kwadrat dane zostały podzielone na k grup. Wykorzystano statystykę:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

gdzie:

O_i — częstość grupy o indeksie i,
E_i — oczekiwana wartość teoretyczna częstości obliczana na podstawie dystrybuanty empirycznej.

Tak wyznaczona statystyka ma rozkład χ^2 z $k–p–1$ stopniami swobody (p określa liczbę estymowanych parametrów rozkładu). Test chi-kwadrat jest wrażliwy na sposób grupowania danych przy obliczaniu statystyki.

Wartość statystyki testowej testu chi-kwadrat jest uzależniona od liczności próby. Identyczne proporcje w tabeli w przypadku większej próby skutkują wyższą wartością statystyki testowej. Dlatego test ten nie może być wykorzystywany do porównania siły związku między zmiennymi w próbach różniących się licznością. Zależność wartości krytycznych testu od liczby stopni swobody nie pozwala na porównanie wyników uzyskanych w tabelach różniących się liczbą kolumn lub wierszy. Aby takie porównania były zasadne, trzeba przeliczyć wartość chi-kwadrat na odpowiedni do rodzaju danych współczynnik siły związku. Często jest to współczynnik V-Cramera. Wielkości te są miarami korelacji między zmiennymi nominalnymi i pozwalają wyrobić sobie zdanie na temat możliwości prognozowania jednej zmiennej ze względu na drugą zmienną. Im są one bliższe zera, tym możliwość ta jest mniejsza.

Test Kołmogorowa-Smirnowa sprawdza, czy rozkład analizowanych danych dla pewnej zmiennej losowej różni się od rozkładu teoretycznego przy danej skończonej liczbie obserwacji tej zmiennej. Test wykorzystuje statystykę postaci:

$$D_n = \sup_{1 \le i \le n} |F(x_i) - F_n(x_i)|$$

gdzie:

F — zadana dystrybuanta teoretyczna,
F_n — dystrybuanta empiryczna,
x_i — znane wartości zmiennej losowej.

Test Kołmogorowa-Smirnowa jest dobry do oceny zgodności rozkładu w otoczeniu średniej.

Jednym z głównych symptomów braku normalności danych jest występowanie „grubych ogonów". Do badania zgodności rozkładu wykorzystuje się tu statystykę Andersona-Darlinga, która jest zdefiniowana następująco:

$$A^2 = -n - S$$

gdzie:

$$S = \sum_{i=1}^{n} \frac{2i-1}{n} [\ln F(x_i) + \ln(1 - F(x_{n+1-i}))],$$

F — zadana dystrybuanta teoretyczna,
n — liczność próbki.

Statystyka Andersona-Darlinga jest pewną modyfikacją testu Craméra-von Misesa pozwalającą na poprawę jego czułości w „ogonach" testowanego rozkładu. Ze względu na wysoką czułość w „ogonach", czyli dla wartości ekstremalnych, w wielu pracach, także w naszej, statystyka ta jest rozstrzygającym kryterium w dopasowywaniu rozkładów.

Warto zwrócić uwagę na niedokładności związane z zastosowaniem opisanych testów. Najważniejszą z nich jest fakt, że testuje się zgodność z rozkładami, których parametry zostały wcześniej wyestymowane metodą największej wiarygodności. Pociąga to za sobą konieczność ostrożnej interpretacji uzyskanych wyników, w szczególności w przypadku, gdy hipoteza o zgodności z badanym rozkładem zostaje odrzucona.

WYNIKI DOPASOWANIA ROZKŁADÓW

Zwykle w statystyce odrzucenie hipotezy zerowej przez jeden z testów jest powodem ostatecznego odrzucenia tej hipotezy. Zdając sobie sprawę z różnej wrażliwości wymienionych trzech testów autorzy oszacowali „stopień niedopasowania" rozkładów, biorąc pod uwagę liczbę testów odrzucających hipotezę zerową.

Współczynniki rozważanych rozkładów wyestymowano metodą największej wiarygodności, z użyciem oprogramowania komputerowego R oraz Matlab, dla sumy dziennych wypłat każdego z bankomatów. Następnie, stosując testy: chi-kwadrat, Kołmogorowa-Smirnowa i Andersona-Darlinga, zweryfikowano hipotezy o dopasowaniu danych do wybranych rozkładów teoretycznych. W przeprowadzanych testach poziom istotności wynosił 0,05.

W dalszej części pracy przedstawiono wyniki analizy związanej z dopasowaniem każdego z rozkładów całej grupy bankomatów. Zgodność dopasowania ocenimy w skali 0—3, przy czym wartość ta określa, w przypadku ilu testów

hipoteza zerowa nie została odrzucona, np. 0 oznacza, że wszystkie testy odrzuciły hipotezę o zgodności danych z rozkładem teoretycznym, natomiast 3 oznacza, że żaden z testów nie odrzucił badanego rozkładu.

TABL. 6. SIŁA DOPASOWANIA POSZCZEGÓLNYCH ROZKŁADÓW TEORETYCZNYCH DO DANYCH EMPIRYCZNYCH

Typ rozkładu	0		1		2		3	
	liczba	w %	liczba	w %	liczba	w %	liczba	w %
Normalny	180	61,43	28	9,56	71	24,23	14	4,78
Logistyczny	213	72,70	37	12,63	38	12,97	5	1,71
Skalowany *t*-Studenta	162	55,29	30	10,24	87	29,69	14	4,78
Potęgowo-wykładniczy	160	54,61	26	8,87	89	30,38	18	6,14
Mieszanka dwóch rozkładów normalnych ...	46	15,70	28	9,56	178	60,75	41	13,99
Mieszanka trzech rozkładów normalnych ...	1	0,34	1	0,34	190	64,85	101	34,47
NIG	120	40,96	20	6,83	110	37,54	43	14,68
Hiperboliczny	129	44,03	19	6,48	103	35,15	42	14,33
α-stabilny	204	69,62	11	3,75	60	20,48	18	6,14

Ź r ó d ł o: jak przy tabl. 1.

W tabl. 6 zamieszczono wyniki dopasowania wybranych rozkładów według wszystkich danych bez ich kategoryzowania. Łatwo zauważyć, że rozkład normalny został odrzucony przez wszystkie trzy testy dotyczące dziennych wypłat z ponad 60% badanych bankomatów. Jeśli chodzi o dopasowanie, to gorzej procentowo wypadają rozkłady logistyczny oraz α-stabilny. Również dane z ponad połowy bankomatów nie zostały dopasowane do rozkładów skalowanego *t*-Studenta oraz potęgowo-wykładniczego. O niewiele lepszym dopasowaniu można mówić w przypadku rozkładów hiperbolicznego oraz odwróconego gaussowskiego. Dla danych z ponad 40% bankomatów wszystkie testy odrzuciły zgodność z tymi rozkładami. Ewidentnymi liderami, jeśli chodzi o zgodność z analizowanymi danymi empirycznymi są mieszanki rozkładów normalnych, zwłaszcza trzech rozkładów normalnych. Okazało się, że tylko w przypadku jednego bankomatu wszystkie testy odrzuciły mieszankę trzech rozkładów jako rozkład opisujący dzienne wypłaty z bankomatów. Dla 291 z 293 bankomatów przynajmniej dwa z trzech testów nie odrzuciły mieszanki trzech rozkładów normalnych.

Wyniki wskazują więc na mieszankę trzech rozkładów normalnych jako ten rozkład, który powinien zostać wzięty pod uwagę w dalszych badaniach nad tworzeniem modelu prognostycznego.

Z tabl. 6 można odczytać, jak każdy z rozkładów dopasowuje się do analizowanych danych empirycznych. Jednak nie ma tu możliwości porównania dopasowania poszczególnych rozkładów do konkretnych bankomatów. W tym celu dla każdego bankomatu wybrano rozkład, który najlepiej charakteryzuje sumę

dziennych wypłat. Za najlepiej dopasowany rozkład uznano ten, który nie został odrzucony przez największą liczbę testów. W przypadku gdy dwa rozkłady miały identyczny wynik wybierano ten, dla którego statystyka Andersona-Darlinga była największa. Jak już podkreślono, test ten charakteryzuje się wysoką czułością, bada zgodność danych empirycznych z rozkładem teoretycznym w „ogonach" rozkładu. Znajomość rozkładu w „ogonach" jest ważna ze względu na wybór odpowiedniego modelu prognostycznego. Wyniki opisanej analizy zamieszczono w tabl. 7.

TABL. 7. PORÓWNANIE WYNIKÓW DOPASOWANIA POSZCZEGÓLNYCH ROZKŁADÓW

Typ rozkładu	Liczba	W %
Brak	1	0,34
Potęgowo-wykładniczy	1	0,34
Skalowany *t*-Studenta	1	0,34
Mieszanka dwóch rozkładów normalnych	23	7,85
Mieszanka trzech rozkładów normalnych	238	81,23
NIG	13	4,44
Hiperboliczny	16	5,46

Ź r ó d ł o: jak przy tabl. 1.

Wyniki potwierdzają wcześniejsze analizy, bowiem w przypadku aż 238 bankomatów, czyli w ponad 80%, najlepszym dopasowaniem charakteryzowała się mieszanka trzech rozkładów normalnych. W tym rankingu kolejne miejsca zajęły: mieszanka dwóch rozkładów normalnych (7,85%), rozkład hiperboliczny (5,46%) i odwrócony gaussowski (4,44%). Tylko dla jednego bankomatu żaden rozkład nie został dopasowany.

Wyniki potwierdzają, że mieszanka trzech rozkładów normalnych najlepiej opisuje strukturę dziennych wypłat z bankomatów. Z tabl. 7 można dodatkowo odczytać, że w przypadku prawie 90% bankomatów mieszanki rozkładów normalnych najlepiej aproksymują rozkłady empiryczne analizowanych danych.

W kolejnej części pracy zbadano, czy istnieje zależność pomiędzy dopasowaniem rozkładu do danych a miejscem, w jakim dany bankomat się znajduje. Aby weryfikacja miała sens, pod uwagę wzięto tylko mieszanki rozkładów normalnych oraz rozkłady NIG i hiperboliczny.

TABL. 8. WYNIKI PORÓWNANIA DOPASOWANIA ROZKŁADÓW WEDŁUG WOJEWÓDZTW

Województwa	Hiperboliczny		Mieszanka dwóch rozkładów normalnych		Mieszanka trzech rozkładów normalnych		NIG	
	liczba	w %	liczba	w %	liczba	w %	liczba	w %
Małopolskie	11	4,80	19	8,30	188	82,10	11	4,80
Podkarpackie	5	8,20	4	6,56	50	81,97	2	3,28

Ź r ó d ł o: jak przy tabl. 1.

Wyniki zamieszczone w tabl. 8 wskazują na brak zróżnicowania w rozkładzie zgodności poszczególnych rozkładów w badanych województwach. Potwierdza to wynik testu niezależności, w którym *p-value*=0,6964.

TABL. 9. WYNIKI PORÓWNANIA DOPASOWANIA ROZKŁADÓW WEDŁUG KATEGORII MIEJSCOWOŚCI

Kategoria miejscowości	Hiperboliczny		Mieszanka dwóch rozkładów normalnych		Mieszanka trzech rozkładów normalnych		NIG	
	liczba	w %	liczba	w %	liczba	w %	liczba	w %
I	—	—	—	—	14	93,33	1	6,67
II	4	8,70	1	2,17	41	89,13	—	—
III	3	12,00	1	4,00	20	80,00	1	4,00
IV	1	2,56	4	10,26	34	87,18	—	—
V	8	4,85	17	10,30	129	78,18	11	6,67

Ź r ó d ł o: jak przy tabl. 1.

Dopasowanie rozkładów do danych dotyczących bankomatów znajdujących się w miejscowościach o różnym zaludnieniu wykazało lekkie dysproporcje (tabl. 9). Można zauważyć, że dla małych miejscowości mieszanka trzech rozkładów normalnych dopasowuje się najlepiej w ponad 90% przypadków, jednocześnie dla Krakowa (kategoria V) udział ten był poniżej 80%. Również wyniki pozostałych rozkładów wskazują na niewielkie zróżnicowanie zgodności dopasowania w zależności od liczby mieszkańców. Jednak przeprowadzony test niezależności nie wskazuje na istotność tego zróżnicowania. Wartość *p-value* w tym przypadku wynosi 0,1889.

TABL. 10. WYNIKI PORÓWNANIA DOPASOWANIA ROZKŁADÓW WEDŁUG TYPÓW LOKALIZACJI

Typ lokalizacji	Hiperboliczny		Mieszanka dwóch rozkładów normalnych		Mieszanka trzech rozkładów normalnych		NIG	
	liczba	w %	liczba	w %	liczba	w %	liczba	w %
Oddział bankowy	3	3,61	5	6,02	73	87,95	2	2,41
Centrum rozrywki	1	16,67	—	—	5	83,33	—	—
Hipermarket	3	7,89	6	15,79	27	71,05	2	5,26
Hotel	—	—	—	—	3	50,00	3	50,00
Stacja paliw	1	4,00	1	4,00	23	92,00	—	—
Sklep	1	2,04	4	8,16	43	87,76	1	2,04
Centrum handlowe	5	11,90	3	7,14	31	73,81	3	7,14
Pozostałe	1	2,70	3	8,11	33	89,19	—	—

Ź r ó d ł o: jak przy tabl. 1.

Wyniki zamieszczone w tabl. 10 wskazują, że rodzaj lokalizacji bankomatu najbardziej różnicuje typy rozkładów, które najlepiej opisują strukturę danych.

Według wszystkich typów lokalizacji mieszanka trzech rozkładów normalnych najtrafniej przybliża rozkład empiryczny danych, z wyjątkiem bankomatów w hotelach, gdzie równie dobre dopasowanie uzyskuje rozkład NIG. W przypadku bankomatów zlokalizowanych na stacjach paliw, w oddziałach bankowych i w sklepach mieszanka trzech rozkładów daje najlepsze dopasowanie w blisko 90% przypadków. Jednak w przypadku bankomatów w hipermarketach i centrach handlowych jest to niewiele ponad 70%. Jeśli chodzi o bankomaty zlokalizowane w hipermarketach, dobry wynik dopasowania uzyskuje również mieszanka dwóch rozkładów. W centrach handlowych dla niektórych bankomatów pozostałe rozkłady dają lepsze dopasowanie. Opisane zróżnicowanie zgodności rozkładów w zależności od typu lokalizacji okazuje się istotne statystycznie, gdyż wartość *p-value* dla testu niezależności wynosi 0.

Wnioski

W artykule autorzy podjęli próbę zbadania empirycznych własności dziennych wypłat z bankomatów ze szczególnym uwzględnieniem dopasowania rozkładów teoretycznych.

Uzyskane wyniki wskazują, że badane szeregi cechują się niejednorodnością. W znaczniej większości przypadków rozkładem, który najlepiej aproksymuje dane empiryczne jest mieszanka trzech rozkładów normalnych.

Dodatkowo wykazano, że największy wpływ na rodzaj rozkładu ma typ lokalizacji, natomiast rodzaj dopasowania rozkładu nie jest powiązany z województwem oraz miejscowością, w których znajduje się bankomat.

Uzyskane wyniki mogą mieć duże znaczenie przy wyborze metody prognozowania oraz modelu prognostycznego. Dobre dopasowanie mieszanki trzech rozkładów normalnych może wskazywać na możliwość zastosowania modelu przełącznikowego do opisu podszeregów dających się wyróżnić z danego szeregu czasowego wypłat w bankomatach na podstawie odpowiednio zdefiniowanych kryteriów.

Chodzi o to, że współczynniki kombinacji wypukłej (proporcje mieszanki) trzech rozkładów normalnych zapewne zmieniają się na skutek m.in. tzw. efektu kalendarza, obserwowanego w przypadku wypłat z bankomatów. Uwzględnienie tego faktu może nastąpić właśnie poprzez zastosowanie modelu przełącznikowego.

W związku z zaobserwowanym zahamowaniem rozwoju sieci bankomatowych w krajach wysoko rozwiniętych jest rzeczą ważną z praktycznego punktu widzenia określenie perspektywy rozwoju sieci bankomatów w tych krajach, a także w Polsce. W szczególności chodzi o rozpoznanie, w jaki sposób wpłynie to na posługiwanie się gotówką, bowiem obecnie przewiduje się raczej spadek popytu na gotówkę.

Chociaż jest to prawdopodobny scenariusz, to jednak rosnące koszty transakcji mogą sprzyjać pobieraniu gotówki. Otwierają się zatem perspektywy badań w zakresie związków pomiędzy koncentracją w sektorze bankowym i podażą bankomatów z jednej strony oraz popytem na gotówkę i dostępnością usług gotówkowych w postaci głównie bankomatów, ale także alternatywnych sposobów płatności, z drugiej strony.

prof. dr hab. Henryk Gurgul, mgr Marcin Suder — *AGH w Krakowie*

LITERATURA

Amromin E., Chakravorti S. (2007), *Debit card and cash usage: a cross-country analysis*, Technical report, Federal Reserve Bank of Chicago

Aparicio F. M., Estrada J. (2001), *Empirical Distribution of Stock Returns: European Securities Markets 1990—1995*, „The European Journal of Finance", No. 7

Armenise R., Birtolo C., Sangianantoni E., Troiano L. (2010), *A generative solution for ATM cash management*, International Conference of Soft Computing and Pattern Recognition 7—10 December

Baringhaus L., Henze N. (1988), *A consistent test for multivariate normality based on the empirical characteristic function*, „Metrika", No. 35

Black F., Scholes M. (1973), *The pricing of options and corporate liabilities*, „Journal of Political Economy", No. 81

Blattberg R., Gonedes N. (1974), *A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices*, „Journal of Business", No. 47

Boeschoten W. C. (1998), *Cash management, payment patterns and the demand for money*, „De Economist", No. 146

Broca D. S. (2002), *The distribution of Indian Stock Returns: A Tale of Two Tails*, „Decision", No. 29

Eberlein E., Keller U. (1995), *Hyperbolic distributions in finance*, „Bernoulli", No. 1

Fama E. F. (1965), *The behavior of stock market prices*, „Journal of Business", No. 38

Gray J. B., French D. W. (1990), *Empirical comparisons of distributional model for stock index prices*, „Journal of Business Finance and Accounting", No. 17

Hand D. J., Blunt G. (2001), *Prospecting for gems in credit card data*, „IMA Journal of Management Mathematics", No. 12

Henze N. (2002), *Invariant tests for multivariate normality: a critical review*, „Statistical papers", No. 43

Hsu D. (1982), *A Bayesian robust detection of shift in the risk structure of stock market returns*, „Journal of the American Statistical Association", March

Kon S. J. (1984), *Models of stock returns: a comparison*, „Journal of Finance", No. 39

Küchler U., Neumann K., Sørensen M., Streller A. (1999), *Stock returns and hyperbolic distributions*, „Mathematical and Computer Modelling", No. 29

Mandelbrot B. (1963), *The variation of certain speculative prices*, „Journal of Business", No. 36

Peiró A. (1994), *International evidence on the distribution of stock return*, „Applied Financial Economics", No. 4

Praetz P. D. (1972), *The distribution of share price changes*, „Journal of Business", No. 45

Press S. J. (1967), *A compound events model for security prices*, „Journal of Business", No. 45

Rencher A. C. (1995), *Methods of multivariate analysis*, Wiley, New York

Sharpe W. F. (1964), *Capital asset price: a theory of market equlibrium under conditions of risk*, „Journal of Finance", No. 19

Simutis R., Dilijonas D., Bastina L., Friman J., Drobinov P. (2007), *Optimization of cash management for ATM network*, „Information Technology and Control", No. 36

Smith J. B. (1981), *The probability distribution of market returns: a logistic hypothesis*, PhD dissertation, University of Utah

Snellman H., Viren M. (2006), *ATM networks and cash usage. Technical Report*, „Bank of Finland Research Discussion Papers", No. 21

Suder M., Wójtowicz T. (2005), *Rozkłady dziennych stop zwrotu wybranych indeksów GPW w Warszawie*, Zeszyty Naukowe Wyższej Szkoły Ekonomii i Informatyki w Krakowie, nr 3

SUMMARY

ATMs require a continuous connection with the performance management of payments, and distributing cash acceptance, attitude appropriate use of appropriate devices and software, as well as protection against fraud. This paper tries to fit the theoretical distributions to the empirical distribution of withdrawals by ATM installed in Małopolskie and Podkarpackie voivodships. Probability distribution may be useful as a signal to take a risk protection strategy or as a basis for filling ATMs with cash. Knowledge of the probability distribution is essential that the data from the past can be used to draw money management forecasts.

РЕЗЮМЕ

Банкоматы требуют непрерывного управления связанного с проведением выплат, приемом и распределением денег, обеспечением соответствующего оборудования и использованием программного обеспечения, а также защитой от мошенничества. Статья представляет методы приспособливания теоретических и эмпирических распределений выплат денег через банкоматы установленные в малопольском и подкарпатском воеводствах. Распределение вероятностей может быть полезным как сигнал для предпринимания защиты от рисков или же в качестве основы стратегии заполнения банкоматов наличными деньгами. Знание распределения вероятностей является необходимым для использования данных из прошлого в составлении прогнозов управления денежными средствами.

Tomasz GÓRECKI, Ewa ŁAŹNIEWSKA

Funkcjonalna analiza składowych głównych PKB

Coraz bardziej złożone zagadnienie analiz regionalnych (związane z koniecznością uwzględniania większych terytoriów oraz większej liczby jednostek statystycznych) przyczynia się do poszukiwania nowszych metod analizy, a także wizualizacji zjawisk — również takich, które pokazują ich dynamikę. Zaprezentowana w artykule metoda może być pomocna w określaniu konkurencyjności regionalnej, polegającej na porównywaniu charakterystyki uwzględniającej cechy brane pod uwagę w *benchmarkingu* regionów.

Postęp technologiczny, zwłaszcza w zakresie techniki komputerowej, znacznie ułatwił analizę danych wysokowymiarowych lub danych z powtarzanymi pomiarami. Jeśli informacje takie zbierane są dostatecznie często w jakimś przedziale czasu I, np. przez urządzenie pomiarowe, nazywa się je danymi funkcyjnymi, gdzie dla każdego przypadku mamy jedną krzywą. Jest to wykorzystywane w przypadku, gdy zjawisko obserwujemy z błędem, ponieważ wygładzanie może znacznie zredukować efekt występowania szumu. W takich przypadkach możemy całą krzywą traktować jako wykres funkcji $X_i(t)$, którą traktujemy jak funkcję ciągłą, pomimo że czas jest dyskretny. Statystyczna analiza n takich krzywych jest powszechnie nazywana **funkcjonalną analizą danych (FDA[1])**[2]. FDA jest zatem metodą analizy, która pozwala odpowiedzieć na pytania typu: *jak różnią się dwie krzywe między sobą?* Unikalną cechą tego typu analizy jest możliwość wykorzystania informacji o szybkości zmian funkcji i jej pochodnych. Możemy wykorzystać wszystkie informacje, z których otrzymujemy gładkie funkcje, a których nie mamy używając jedynie dyskretnych danych. Tego typu analiza, której głównym celem jest dostarczenie narzędzi służących do opisu i modelowania zbioru krzywych jest obecnie celem coraz większego zainteresowania społeczności statystycznej.

W książkach Ramsaya i Silvermana (2002, 2005) możemy znaleźć opis procedur radzenia sobie z danymi funkcyjnymi. Podane tam metody znalazły szerokie zastosowanie m.in. w chemometrii, klimatologii, biologii oraz ekonomii. Statystycy w pierwszym kroku zazwyczaj próbują tak zredukować dane, aby zachować maksymalnie dużo informacji przy minimalnej liczbie zmiennych. Optymalne byłoby uzyskanie dwóch lub trzech zmiennych, które można w prosty sposób wizualizować.

Z kolei za pomocą **funkcjonalnej analizy składowych głównych (FPCA[3])** możemy dokonać rzutu wysokowymiarowych danych na przestrzeń o dużo

[1] Ang. *Functional Data Analysis*.
[2] Ramsay J. O., Dalzell C. J. (1991), s. 539—572.
[3] Ang. *Functional Principal Components Analysis*.

mniejszym wymiarze, jednocześnie zachowując maksymalnie dużo informacji (w tym przypadku zmienności danych). Główna idea metody polega na znalezieniu takich funkcji, których iloczyn skalarny[4] z danymi pozwala otrzymać maksymalną zmienność. Pierwsza składowa wyjaśnia najwięcej zmienności, druga jest prostopadła do pierwszej i wyjaśnia maksymalnie dużo z tego, co zostało itd.

Wkład każdej kolejnej składowej w ogólną zmienność początkowych zmiennych jest nie tylko znany, ale i coraz mniejszy, od pewnego momentu składowe będą wyjaśniać znikomy zakres zmienności. Z punktu widzenia badacza nie będą już zatem wnosiły istotnych informacji i mogą być pominięte w dalszej analizie. Rozpatrywane dane można w ten sposób zredukować, przy jednoczesnej pełnej kontroli badacza nad stopniem utraty informacji. Większość zmienności zachowana jest wtedy przez kilka pierwszych funkcjonalnych składowych głównych. Metoda konstrukcji takich składowych opisana przez Ramsaya i Silvermana (2005) znalazła wiele zastosowań. Ze względu na sposób konstrukcji funkcjonalnych składowych głównych są one wzajemnie nieskorelowane, a zatem można je również wykorzystać w dalszej analizie z użyciem tych metod, w których wymagane jest założenie o braku współliniowości (jak np. analiza dyskryminacyjna, analiza regresji wielorakiej). Jednym z takich zastosowań jest estymacja losowych trajektorii z dostępnych danych[5].

METODA KONSTRUKCJI FUNKCJONALNYCH SKŁADOWYCH GŁÓWNYCH

Załóżmy, że obserwujemy proces losowy $X(t)_{t\in I} \in L^2(I)$, gdzie $L^2(I)$ jest przestrzenią Hilberta[6] funkcji całkowalnych z kwadratem w przedziale czasu I, wyposażoną w iloczyn skalarny. Dodatkowo załóżmy, że $\forall_{t\in I} EX(t) = 0$ (tzn. proces jest scentrowany) oraz ma skończoną wariancję (zmienność).

Możemy założyć, że taki proces losowy $X(t)$ jest reprezentowany przez skończoną liczbę funkcji ortonormalnych. Wyrazimy to wzorem[7]:

$$X(t) = \sum_{k=0}^{K} c_k \phi_k(t) = c'\phi(t) \quad t \in I \tag{1}$$

[4] W przypadku przestrzeni $L^2(I)$ iloczyn skalarny funkcji f oraz g wyraża się wzorem:

$$\langle f, g \rangle = \int_I f(t)\overline{g(t)}\, dt$$

[5] Ingrassia S., Constanzo G. D. (2005), s. 351—358.

[6] Zauważmy, że jeśli skupimy się na funkcjach mających ograniczone normy, to wystarcza założenie, aby $L^2(I)$ była jedynie przestrzenią Hilberta. Jednakże w wielu przypadkach jest to niewystarczające. W takiej sytuacji potrzebna jest nam przestrzeń Hilberta z jądrem reprodukującym (Wahba, 1990).

[7] Pierwszy znak w równości jest formalnie niepoprawny — w praktyce aproksymujemy $X(t)$. Jednakże dla zachowania prostoty zapisu będziemy pomijać różnicę pomiędzy prawdziwym $X(t)$ i jego przybliżeniem.

gdzie:

$\{\phi_k\}$ — $(K+1)$ pierwsze elementy bazy ortonormalnej w przestrzeni $L^2(I)$,

$\{c_k\}$ — zmienne losowe z wartością oczekiwaną równą zero ($E(c)=0$) oraz skończoną, niezerową wariancją ($\mathrm{Var}(c)=\mathbf{\Sigma}$).

Macierz $\mathbf{\Sigma}$ jest nieznana, dlatego użyjemy estymatora $\hat{\mathbf{\Sigma}}$ opartego na N niezależnych realizacjach wektora losowego c:

$$\hat{C} = \begin{bmatrix} \hat{c}_{10} & \cdots & \hat{c}_{1K} \\ \vdots & \ddots & \vdots \\ \hat{c}_{N0} & \cdots & \hat{c}_{NK} \end{bmatrix} \tag{2}$$

gdzie $\hat{c}_{ik}$ — oszacowania parametrów c_{ik} uzyskane metodą najmniejszych kwadratów w reprezentacji:

$$x_i(t) = \sum_{k=0}^{K} c_{ik}\phi_k(t) \quad t \in I, \quad i = 1, 2 ..., N \tag{3}$$

procesu losowego $X(t)$.

Wygładzenie liniowe otrzymujemy, jeśli wyznaczymy współczynniki rozwinięcia c_k poprzez minimalizację kryterium najmniejszych kwadratów. W taki sposób otrzymujemy następujące rozwiązanie (Ramsay, Silverman, 2005):

$$\hat{c} = (\phi'\phi)^{-1}\phi' x \tag{4}$$

Estymator $\hat{\Sigma}$ nieznanej macierzy $\mathbf{\Sigma}$ ma postać:

$$\hat{\Sigma} = \frac{1}{N}\hat{C}'\hat{C} \tag{5}$$

Teraz możemy już wyznaczyć wartości własne $\hat{\lambda}_k$ (wszystkie są niezerowe i różne) oraz odpowiadające im wektory własne $\hat{u}_k$ macierzy $\hat{\Sigma}$. Następnie możemy znaleźć funkcje wagowe:

$$\hat{u}_k(t) = \hat{u}_k'\phi(t) \quad t \in I \tag{6}$$

Ostatecznie współrzędna rzutu i-tej realizacji procesu $X(t)$ na j-tą funkcjonalną składową główną ma postać:

$$\hat{U}_{ij} = \sum_{k=0}^{K} \hat{c}_{ik}\hat{u}_{jk} = \hat{c}_i'\hat{u}_j \quad i = 1, 2, ..., N, \quad j = 1, 2, ... \tag{7}$$

W tej formule wykorzystujemy klasyczny iloczyn skalarny wektorów. Poziom (siła) wygładzenia zależy od K, stąd małe (duże) wartości K dają bardziej (mniej) wygładzone krzywe. Optymalna liczba elementów bazy K jest wyznaczana za pomocą bayesowskiego kryterium informacyjnego[8] (BIC) dla każdej funkcji $x_i(t)$ osobno, następnie najczęściej występująca wartość (dominanta) wyznacza optymalną wartość K dla całego zbioru danych. Przyjęto, że K nie może przekraczać wartości 50[9]. Jako baza wybrana została ortonormalna baza wielomianów Legendre'a[10] w przestrzeni $L^2([-1,1])$, która ma postać:

$$\widetilde{P}_k(x) = \sqrt{\frac{2k+1}{2}} P_k(x) \tag{8}$$

gdzie:

$$P_{k+1}(x) = \frac{1}{k+1}[(2k+1)xP_k(x) - kP_{k-1}(x)], \quad k \geq 1,$$
$$P_0(x)=1,$$
$$P_1(x)=x.$$

Najczęściej rozpatrywane obiekty przedstawiane są w układzie dwóch pierwszych funkcjonalnych składowych głównych. W tym przypadku kryterium jakości skonstruowanych funkcjonalnych składowych głównych jest wyrażenie (Hastie i in., 2009):

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{\sum \hat{\lambda}_i} 100\% \tag{9}$$

gdzie $\hat{\lambda}_i \geq \hat{\lambda}_{i+1}, \quad i = 1, 2, \ldots$ — niezerowe wartości własne macierzy $\hat{\Sigma}$.

Im wyrażenie to jest większe, tym większą zmienność wykazują dwie pierwsze funkcjonalne składowe główne, a co za tym idzie, jakość reprezentacji jest lepsza.

---

[8] Jest to kryterium wyboru modelu spośród skończonego zbioru modeli, który oparty jest na funkcji największej wiarygodności i jest ściśle związane z kryterium informacyjnym Akaike (AIC). Podczas dopasowywania modeli można podnieść wartość funkcji największej wiarygodności poprzez zwiększenie liczby parametrów, powodując często przeuczenie modelu (model jest idealnie dopasowany do danych uczących, ale nie ma własności generalizacji, zatem nie sprawdza się na innych danych tego samego typu). Kryteria informacyjne rozwiązują ten problem poprzez wprowadzenie kary zależnej od liczby użytych w modelu parametrów (tzw. kary za złożoność), przy czym w kryterium BIC ta kara jest wyższa (McQuarrie, Tsai, 1998).

[9] Liczba funkcjonalnych składowych głównych jest mniejsza od N (lub N–1, gdy dane zostały scentrowane). Jeśli jednak $K<N$, to liczba funkcjonalnych składowych głównych jest równa K.

[10] Wybrano bazę wielomianów Legendre'a ponieważ uzyskuje się dla niej najlepsze wyniki (Górecki, Krzyśko, 2012, s. 71—87) — porównywane były bazy ortonormalne: Fouriera, cosinusów, sinusów oraz baza wielomianów Legendre'a.

ZASTOSOWANIE FPCA W ANALIZACH REGIONALNYCH

Rozpatrywane były dane dotyczące PKB *per capita* w latach 1999—2008 w Polsce. Dane te pokazano (dla skrajnych lat) na wykr. 1 i 2. Kody na wykresach oznaczają regiony Polski, rozpoczynają się dwuliterowym kodem naszego kraju, kolejne poziomy podziału oznaczane są cyframi. Jeśli istnieje więcej niż 9 jednostek numeracja jest kontynuowana z użyciem wielkich liter poczynając od A.

W takiej sytuacji mamy $N = 66$ szeregów czasowych o długości $T = 10$. Na wykr. 3 przedstawiono trajektorie tych danych, czerwoną pogrubioną linią zaznaczono trajektorię średniej w badanym okresie zaprezentowanego zagadnienia. Możemy zatem zaobserwować, jak zachowuje się badany wskaźnik w ciągu dekady. Niestety dość trudne jest określenie, które regiony zmieniały się podobnie w rozważanym okresie pod względem PKB *per capita*.

Kolejny wykres (4), to wykres piargowy[11] (osypisko) dla FPCA[12]. Widzimy, że pierwsza funkcjonalna składowa główna wyjaśnia 93,31% zmienności, podczas gdy druga 5,33%. W sumie dwie pierwsze funkcjonalne składowe główne wyjaśniają aż 98,64% zmienności[13]. W takiej sytuacji rzut na płaszczyznę skonstruowaną za pomocą dwóch pierwszych funkcjonalnych składowych głównych wydaje się być bardzo dobrym rozwiązaniem. Algorytm wybrał K=11 elementów ortonormalnej bazy wielomianów Legendre'a w przestrzeni $L^2([-1,1])$.

Wykres 5 przedstawia dwie pierwsze funkcje wagowe $u_k(t)$, k=1,2, dla K=11 elementów ortonormalnej bazy wielomianów Legendre'a w przestrzeni $L^2([-1,1])$. Pokazują one, w jaki sposób taki zestaw danych funkcjonalnych oscyluje wokół średniej, czyli w zasadzie potwierdzają zaobserwowane wcześniej trendy. Zauważmy jednak, że obie funkcje (linie ciągłe) przyjmują wartości ujemne (i są malejące), natomiast wyjściowe dane przyjmowały jedynie wartości dodatnie (i były rosnące). Wyjaśnienie tego fenomenu leży w pewnej, niewielkiej, niejednoznaczności wyznaczonych składowych głównych. Mianowicie, jak już wspomniano, składowe wyznaczone są jednoznacznie z dokładnością do znaku. Gdybyśmy zmienili znaki wyznaczonych składowych, otrzymalibyśmy funkcje (linie przerywane) cały czas dodatnie (i rosnące), które dobrze wpasowują się w trend pokazany na wykr. 3. Widzimy więc, że pierwsza funkcja wagowa odpowiada za trend, natomiast druga za odchylenia okresowe od średniej.

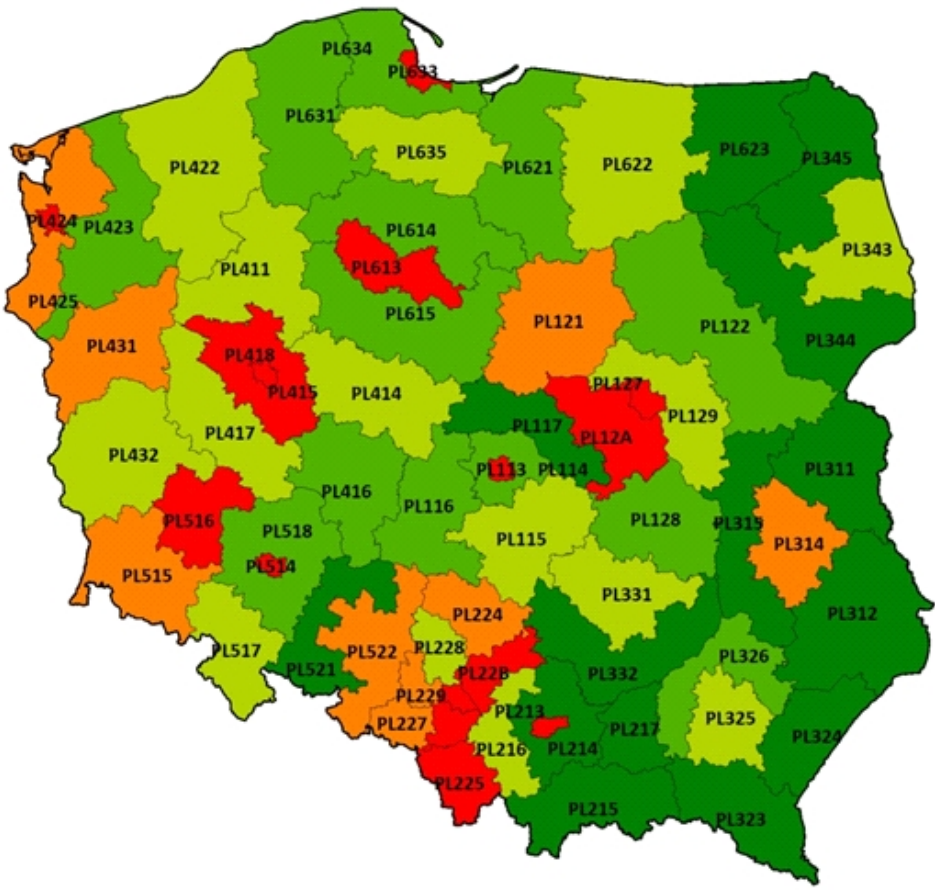
---

[11] Wykres pokazujący kolejne wartości własne. Kolejne funkcjonalne składowe główne wyodrębniają coraz mniejszą część zmienności, to i spadki wartości na wykresie są coraz mniejsze. Test osypiska polega na znalezieniu miejsca, na prawo od którego regularny spadek wartości własnych staje się wolniejszy. Nie należy wyodrębniać więcej czynników niż znajduje się po lewej stronie tego punktu.

[12] Całość obliczeń została wykonana w programie Matlab R2011a (Ramsay, Matlab, 2005).

[13] W przypadku klasycznej analizy składowych głównych otrzymujemy, że pierwsza składowa główna wyjaśnia 77,34% zmienności, natomiast druga — 8,21%. W sumie składowe te wyjaśniają 85,55% zmienności. Co jest wynikiem znacznie gorszym od metody FPCA.

Wykr. 1. PKB *PER CAPITA* W 1999 R.

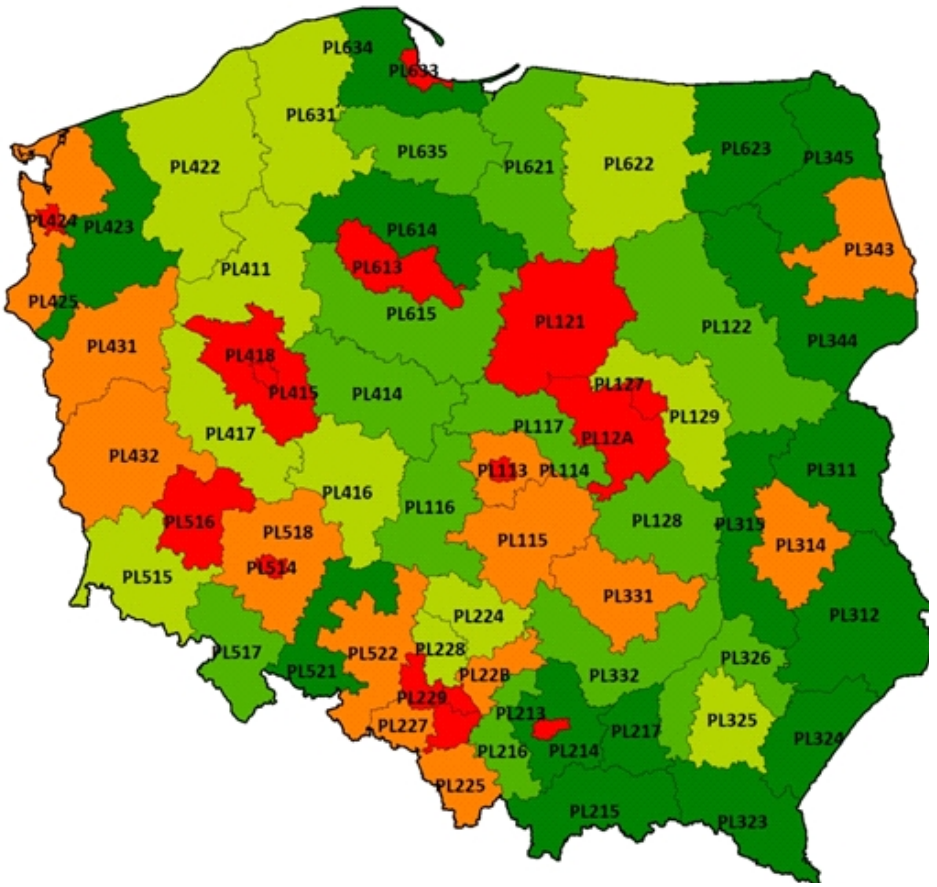


PKB *per capita*

od 9000 do 25400 (15)
od 7700 do 9000 (9)
od 6800 do 7700 (15)
od 6100 do 6800 (13)
od 5000 do 6100 (14)

Ź r ó d ł o: opracowanie własne.

Wykr. 2. PKB *PER CAPITA* W 2008 R.



PKB *per capita*

od 14900 do 41600 (15)

od 12000 do 14900 (13)

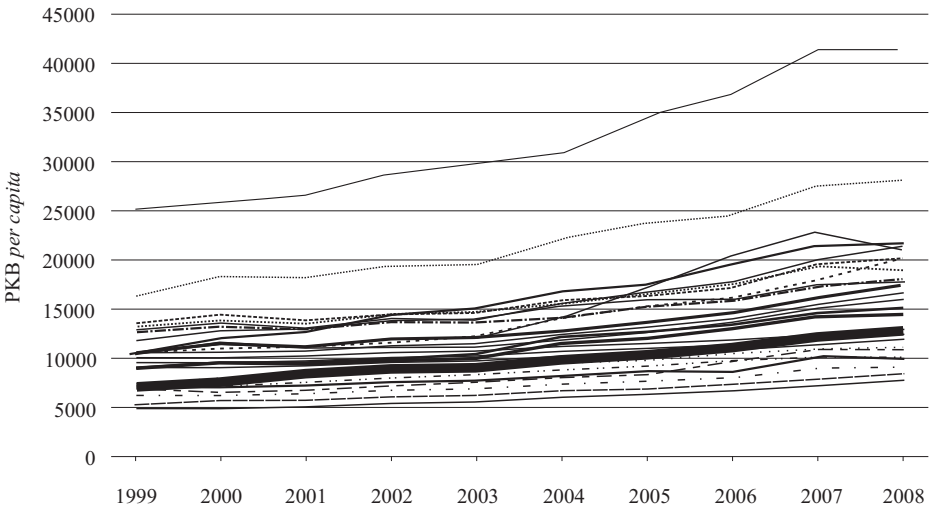
od 11000 do 12000 (11)

od 9900 do 11000 (12)

od 8100 do 9900 (15)

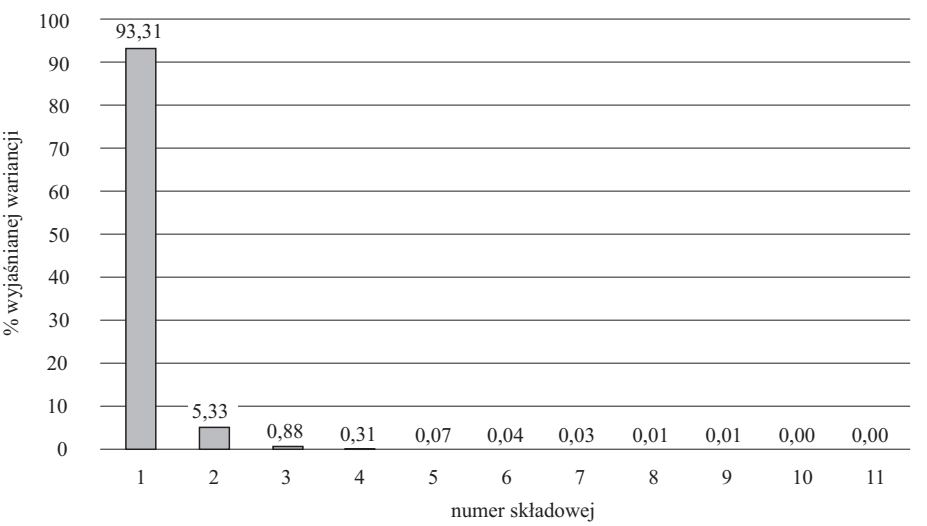
Źródło: jak przy wykr. 1.

Wykr. 3. TRAJEKTORIE PKB *PER CAPITA*



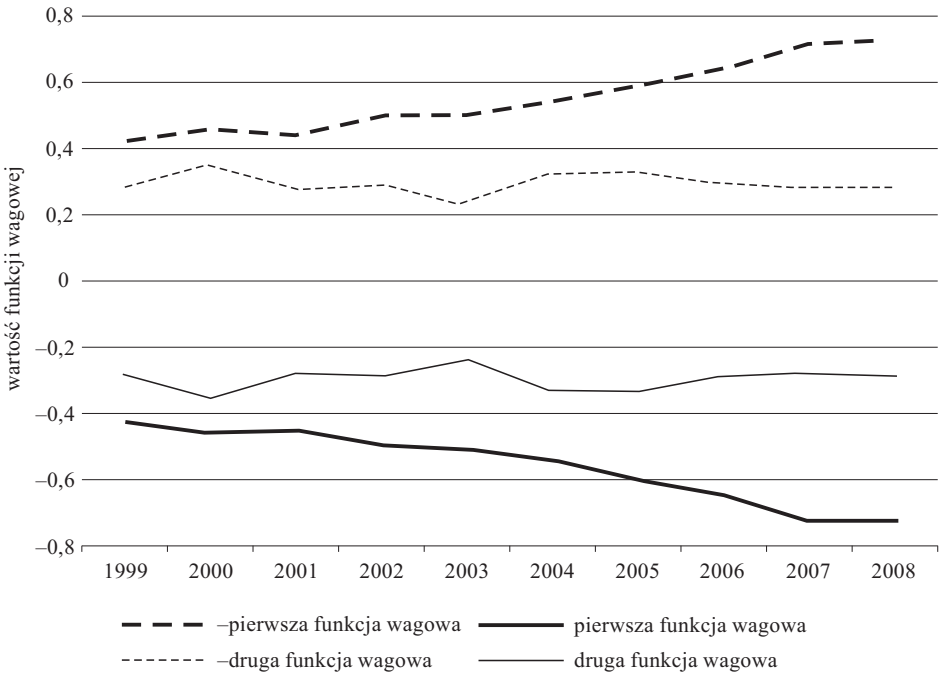
Ź r ó d ł o: jak przy wykr. 1.

Wykr. 4. WYKRES PIARGOWY (osypiska)



Ź r ó d ł o: jak przy wykr. 1.

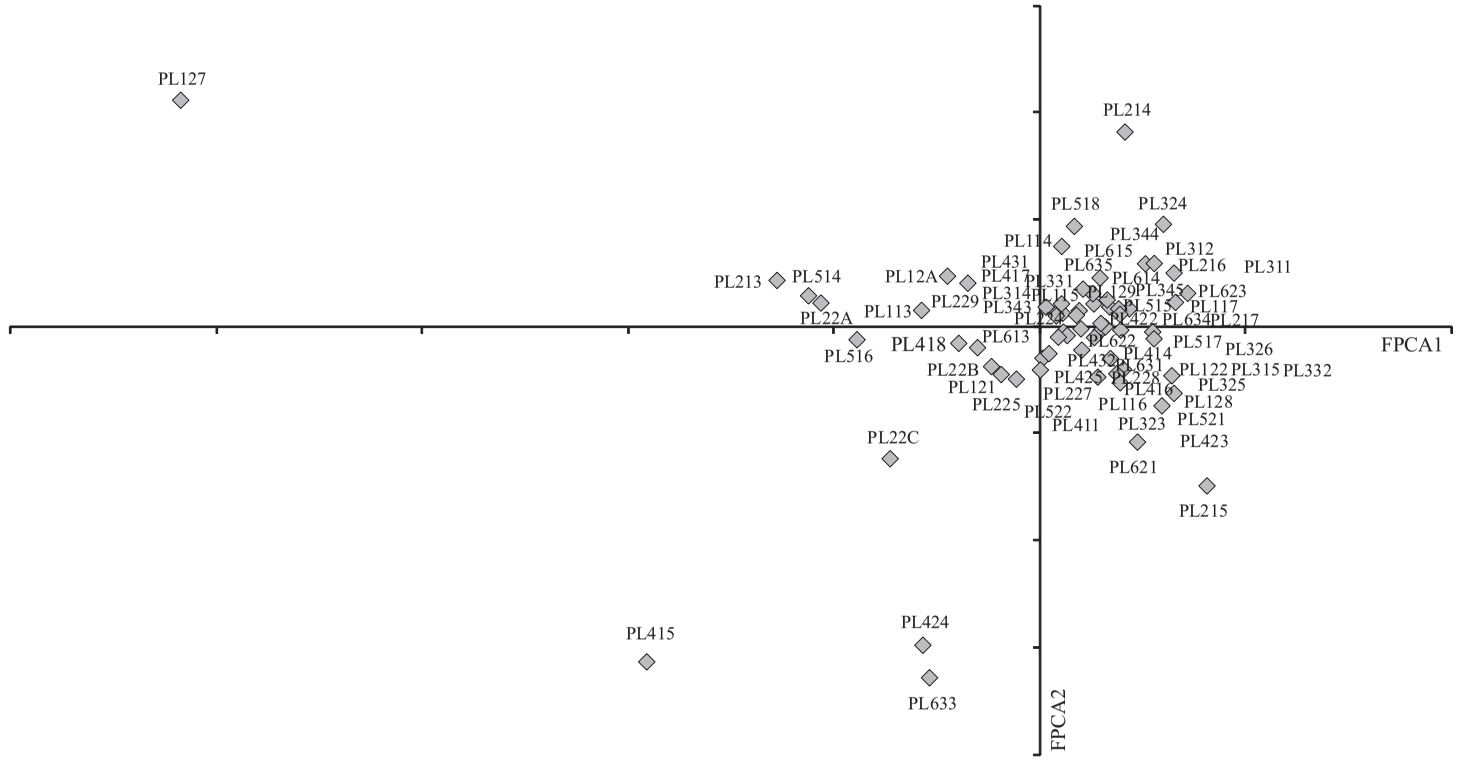
Wykr. 5. DWIE PIERWSZE FUNKCJE WAGOWE



Ź r ó d ł o: jak przy wykr. 1.

Na koniec pozostaje najważniejsza wizualizacja. Mianowicie, rzut danych na przestrzeń dwóch pierwszych funkcjonalnych składowych głównych, który ma nam pokazać, które regiony zachowywały się podobnie w rozpatrywanym okresie pod względem zmian oraz poziomu PKB *per capita*. Jeśli punkty leżą blisko siebie oznacza to, że regiony zachowywały się pod tym względem podobnie. Na wykr. 6 widzimy, że np. Szczecin (PL424) oraz podregion trójmiejski (PL633) zachowywały się w badanym okresie pod względem poziomu oraz zmian PKB *per capita* podobnie, natomiast stolica Polski (PL127) wyróżniała się zdecydowanie.

Wykr. 6. RZUT DANYCH NA UKŁAD DWÓCH PIERWSZYCH FUNKCJONALNYCH SKŁADOWYCH GŁÓWNYCH



Ź r ó d ł o: jak przy wykr. 1.

Wnioski końcowe

Przydatność tej metody do analiz ekonomicznych wydaje się być niezmiernie ciekawa i ważna ze statystycznego punktu widzenia. Trudno wskazać wszystkie możliwe zastosowania, ale zauważmy, że jednym z nich może być analiza stóp zwrotu walut czy spółek giełdowych. Można stwierdzić, że użyteczność tej metody w analizach regionalnych jest istotna, ponieważ na jej podstawie dokonujemy porównań w zachowaniu poszczególnych obiektów analizy z perspektywy analizowanego okresu. W naszym przypadku była to analiza podregionów w okresie 1999—2008. Z obserwacji tej wynika, że regiony o podobnym poziomie zjawiska wykazywały podobieństwo w utrzymaniu lub zmianie poziomu PKB *per capita*, a regiony o wyższym PKB *per capita*, np. podregion stołeczny, zachowywały się zupełnie inaczej. Takie wnioski mogą okazać się przydatne do formułowania zaleceń w polityce regionalnej.

Przydatność tej analizy znajduje także odzwierciedlenie w badaniu konkurencyjności regionalnej. Naszym zdaniem, uzyskane wyniki pokazują użyteczność opisanej metodologii w wizualizacji oraz wnioskowaniu na temat ekonomicznych szeregów czasowych z nieco innej perspektywy niż było to czynione do tej pory przy użyciu powszechnie wykorzystywanych metod, jak chociażby klasyczna analiza składowych głównych czy analiza czynnikowa. Klasyczne metody, pomimo że prostsze, nie pokazują jednak pełnego obrazu sytuacji. Z jednej strony mogą one pomijać fakt, że dane mają postać szeregu czasowego i z tego powodu istotna jest kolejność zmiennych (np. klasyczna analiza składowych głównych). Z drugiej strony mogą nie mieć statystycznych własności (np. szereg ciekawych wizualizacji), które daje nam zaproponowana metoda.

---

dr Tomasz Górecki — *Uniwersytet im. Adama Mickiewicza w Poznaniu*, **dr hab. Ewa Łaźniewska** — profesor *Uniwersytetu Ekonomicznego w Poznaniu*

LITERATURA

Górecki T., Krzyśko M. (2012), *Functional Principal Components Analysis*, [w:] *Data analysis methods and its applications*, red. J. Pociecha, R. Decker, Wydawnictwo C. H. Beck

Hastie T., Tibshirani R., Friedman J. (2009), *The Elements of Statistical Learning*, Springer

Ingrassia S., Constanzo G. D. (2005), *Functional principal components analysis of financial time series*, [w:] *New Developments in Classification and Data Analysis*, eds. Vichi M., Monari P., Mignani S., Montanari A., Springer, Berlin, s. 351—358

McQuarrie A. D. R., Tsai C. L. (1998), *Regression and Time Series Model Selection*, World Scientific

Ramsay J., Matlab R. (2005), *S-Plus Functions for Functional Data Analysis*, McGill University, ftp://ego.psych.mcgill.ca/pub/FDAfuns.pdf

Ramsay J. O., Dalzell C. J. (1991), *Some tools for functional data analysis,* „Journal of the Royal Statistical Society", Series B (Statistical Methodology), No. 53

Ramsay J. O., Silverman B. W. (2002), *Applied Functional Data Analysis: methods and case studies*, Springer

Ramsay J. O., Silverman B. W. (2005), *Functional Data Analysis*, Springer

Wahba G. (1990), *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia

SUMMARY

The article describes the use of functional GDP principal components analysis from the exploratory point of view. This analysis is designed to show the variation in the entire sample, not only discrete observations. This is a technique that is often used as an introduction (e.g., dimensionality reduction and data visualization) for further analysis. This approach emphasizes the significant characteristics of associated statistically data sets and provides useful tools for statistical analysis of economic data sets. The study concerns the GDP per capita in Poland. The results showed that the two main functional components explain 98,64% of the variation. This means that the presented visualization and interpretation of the results is reliable.

РЕЗЮМЕ

В статье характеризуется использование функционального анализа основных компонентов ВВП с исследовательской точки зрения. Этот анализ должен показать изменения во всей выборке, а не только дискретные наблюдения. Эту технику часто используют в качестве введения (напр. сокращение размерности или визуализация данных) для дальнейшего анализа. Такой подход подчеркивает важные статистические характеристики связанных наборов данных, а также предоставляет полезные инструменты для статистического анализа экономических наборов данных. Исследование касается ВВП на душу населения в Польше. Полученные результаты показали, что два основных функциональных компонента объясняют 98,64% вариаций, то есть представленная визуализация и интерпретация результатов является надежной.

Dominik ŚLIWICKI, Marek RĘKLEWSKI

Ekonometryczna analiza czynników aktywności zawodowej kobiet i mężczyzn

Celem badania była próba identyfikacji czynników wpływających na zakwalifikowanie do grupy pracujących osób aktywnych zawodowo. Według metodologii Badania Aktywności Ekonomicznej Ludności (BAEL) (*Aktywność ekonomiczna*…, 2010) ludność aktywna zawodowo określana jest jako zasoby pracy, podaż pracy lub siła robocza obejmująca wszystkie osoby w wieku 15 lat i więcej uznane za pracujące lub bezrobotne. Do kategorii pracujących zalicza się osoby, które w okresie badanego tygodnia:

a) wykonywały przez co najmniej 1 godzinę pracę przynoszącą zarobek lub dochód;
b) miały pracę, ale jej nie wykonywały z powodu choroby, urlopu macierzyńskiego lub wypoczynkowego lub z innych powodów, przy czym długość przerwy w pracy wynosiła do 3 miesięcy albo powyżej 3 miesięcy, ale w drugim przypadku osoby te były pracownikami najemnymi i w tym czasie otrzymywały co najmniej 50% dotychczasowego wynagrodzenia;
c) uczniowie, z którymi zakłady pracy lub osoby fizyczne zawarły umowę o naukę zawodu lub przyuczenie do określonej pracy, jeżeli otrzymywali wynagrodzenie.

Do bezrobotnych natomiast zaliczono osoby w wieku 15—74 lata, które:

a) w okresie badanego tygodnia nie były osobami pracującymi, aktywnie poszukiwały pracy, były gotowe podjąć pracę w ciągu dwóch tygodni następujących po tygodniu badanym;
b) nie poszukiwały pracy, ponieważ miały pracę załatwioną i oczekiwały na jej rozpoczęcie przez okres nie dłuższy niż 3 miesiące oraz były gotowe tę pracę podjąć.

W analizach rynku pracy bardzo często wykorzystuje się modele zmiennych jakościowych, do których należą m.in. modele logitowe. Na potrzeby tego badania przyjęto, że modele te opisywać będą prawdopodobieństwo zakwalifikowania danej osoby aktywnej zawodowo do kategorii pracujących.

TEORETYCZNE PODSTAWY MODELI LOGITOWYCH

Modele logitowe należą do grupy modeli jakościowych, tzn. takich, w których zmienna objaśniana Y jest zmienną jakościową postaci:

$$y_i = \begin{cases} 1 & \text{dla osoby pracującej} \\ 0 & \text{dla osoby bezrobotnej} \end{cases}$$

Model logitowy przyjmuje postać (Gruszczyński, 2010):

$$y_i^* = \ln\frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i$$

gdzie:

y_i^* — zmienna nieobserwowalna,

$\ln\frac{p_i}{1-p_i}$ — logit,

β_j — parametr strukturalny modelu, j=1, 2, …, k,

u_i — składnik losowy,

x_{ji} — wartości zmiennych objaśniających modelu,

p_i — prawdopodobieństwo przyjęcia przez zmienną zależną Y wartości 1, wyznaczone na podstawie funkcji gęstości rozkładu logistycznego:

$$p_i = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} = \frac{1}{1+\exp(-x_i'\beta)} = \frac{1}{1+e^{-y_i}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+...+\beta_k x_{ki})}}$$

Zmienną nieobserwowalną y_i^* nazywa się zmienną ukrytą. Tym, co obserwujemy jest zmienna zero-jedynkowa y_i postaci:

$$y_i = \begin{cases} 1 \text{ dla } & y_i^* > 0 \\ 0 \text{ dla } & y_i^* \le 0 \end{cases}$$

Logit to logarytm ilorazu szans przyjęcia lub nieprzyjęcia wartości 1 przez zmienną y_i. Jeśli szanse są jednakowe (p_i=0,5), to logit równa się zeru. Dla p_i<0,5 logit jest ujemny, a dla p_i>0,5 jest dodatni. Logitowa transformacja prawdopodobieństwa pozwala zastąpić wartość p_i przez liczbę z przedziału $(-\infty,+\infty)$. Jeśli oznaczymy (Gruszczyński, 2010):

$$\frac{p_i}{1-p_i} = \exp(x_i'\beta) = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki})$$

to oznacza to, że przyrost wartości X_{ji} o jednostkę wiąże się, *ceteris paribus*, z exp(β_j)-krotną zmianą ilorazu szans. W przypadku exp(β_j)>1 mamy wzrost, a w przypadku exp(β_j)<1 obserwujemy spadek ilorazu szans $\frac{p_i}{1-p_i}$.

Efekty krańcowe w modelu logitowym nie są stałe i zależą od zmiennych objaśniających:

$$\frac{\partial p_i}{\partial X_{ji}} = \beta_j \frac{\exp(x_i'\beta)}{[1+\exp(x_i'\beta)]^2} = \beta_j p_i (1-p_i)$$

Interpretacja otrzymanych parametrów strukturalnych modelu jest podobna do interpretacji parametrów modelu liniowego. Odczytywane są one jako przyrost prawdopodobieństwa zdarzenia Y=1 związany z jednostkowym przyrostem wartości zmiennej X_j. W przypadku dodatniego β_j, wzrost X_j wiąże się ze wzrostem szans na to, że Y=1, natomiast spadkowi X_j towarzyszy spadek szans na to, że Y=1. Dla ujemnego β_j wzrost X_j wiąże się ze spadkiem szans na to, że Y=1, natomiast spadkowi X_j towarzyszy wzrost szans na to, że Y=1. Efekty krańcowe można wyznaczyć dla średnich albo ustalonych wartości zmiennych objaśniających.

Istotność całego modelu weryfikuje się za pomocą testu ilorazu wiarygodności, w którym układ hipotez ma postać:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$$
$$H_1 : \exists_{1 \le j \le k} \beta_j \neq 0$$

Hipoteza zerowa głosi, że wszystkie parametry przy zmiennych objaśniających są zerami, to znaczy prawdziwy jest jedynie model z wyrazem wolnym. Statystyka testu przyjmuje wtedy postać:

$$LR = 2(\ln L_p - \ln L_{ww})$$

gdzie:

L_p — wartość funkcji wiarygodności dla pełnego modelu,
L_{ww} — wartość funkcji wiarygodności dla modelu zawierającego tylko wyraz wolny.

Jakość dopasowania modelu zmiennej zero-jedynkowej można ocenić na podstawie R^2, tzw. pseudo-R^2. Wartości tego współczynnika mieszczą się w przedziale [0; 1], a jego wyższe wartości mają świadczyć o lepszym dopasowaniu modelu. W badaniu wykorzystano wskaźnik pseudostatystyki R^2 McFaddena. Opiera się on na porównaniu modelu pełnego z modelem zredukowanym tylko do wyrazu wolnego. Oblicza się go według wzoru (Kośko i in., 2007):

$$McFaddenR^2 = 1 - \frac{\ln L_p}{\ln L_{ww}}$$

gdzie:

$\ln L_p$ — logarytm funkcji wiarygodności modelu pełnego,
$\ln L_{ww}$ — iloraz funkcji wiarygodności modelu, w którym występuje tylko wyraz wolny.

W praktyce wartości R^2 McFaddena są niewielkie, bliższe 0 niż 1, z uwagi na fakt, że do estymacji modeli logitowych wykorzystuje się mikrodane, których prognozowanie stwarza znaczne trudności.

R^2 można wyznaczyć na podstawie współczynnika korelacji między y oraz $\hat{y}$. Miara ta oparta jest na resztowej sumie kwadratów (Maddala, 2008):

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \right]$$

W przypadku zależnej zmiennej binarnej mamy:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = n_1 - n\left(\frac{n_1}{n_0}\right)^2 = \frac{n_1 n_0}{n}$$

Z tego wzoru otrzymujemy:

$$R^2 = 1 - \frac{n}{n_1 n_0} \sum_{i=1}^{n} (y_i - \hat{y})^2$$

gdzie:

n_1 — liczba obserwacji, dla których zmienna binarna przyjmuje wartość 1 w n-elementowej próbie,
n_0 — liczba obserwacji tej zmiennej równych 0.

Dodatkowym sposobem na zbadanie jakości dopasowania modelu jest przedstawienie wyników prognozy na podstawie modelu. Prognoza oparta jest na oszacowanym prawdopodobieństwie $\hat{p}_i$, które jest funkcją $F(x_i'\beta)$. Zazwyczaj przyjmuje się, że jeśli $F(x_i'\beta) \geq 0{,}5$, to prognoza jest równa $\hat{y}_i = 1$. Jeśli

$F(x_i'\beta) < 0{,}5,$ to prognoza z modelu równa się $\hat{y}_i = 0$. Na tej podstawie generowana jest tablica trafności w następującej postaci:

Faktyczne	Przewidywane		Razem
	$\hat{Y}=0$	$\hat{Y}=1$	
Y=0	n_{00}	n_{01}	N_{p0}
Y=1	n_{10}	n_{11}	N_{p1}
Razem	N_{f0}	N_{f1}	N

gdzie liczba przypadków, dla których:

n_{00} — wartość rzeczywista i przewidywana są równe 0,
n_{01} — wartość rzeczywista wynosi 0, a przewidywana 1,
n_{10} — wartość rzeczywista wynosi 1, a przewidywana 0,
n_{11} — wartość rzeczywista i przewidywana są równe 1.

Zatem procentową trafność prognoz oblicza się następująco (Kufel, 2011):

ogółem: $$Traf\ \mathrm{Prog} = \frac{n_{00} + n_{11}}{N} \cdot 100$$

dla $Y=1$: $$Traf\ \mathrm{Prog}_1 = \frac{n_{11}}{N_{p1}} \cdot 100$$

dla $Y=0$: $$Traf\ \mathrm{Prog}_0 = \frac{n_{00}}{N_{p0}} \cdot 100$$

Ponadto trafność modeli zmiennych jakościowych można przedstawić za pomocą ilorazu szans według wzoru:

$$IRS = \frac{n_{11} \cdot n_{00}}{n_{01} \cdot n_{10}}$$

DANE WYKORZYSTANE DO ESTYMACJI MODELI LOGITOWYCH

Podstawę oszacowania modeli prawdopodobieństwa zakwalifikowania do grupy pracujących stanowiły dane jednostkowe uzyskane w ramach reprezentacyjnego BAEL w IV kwartale 2010 r. Zbiór zmiennych objaśniających liczył 54 zmienne binarne (zestawienie) opisujące społeczno-ekonomiczną sytuację ankietowanych na rynku pracy. Zaliczono do nich: poziom wykształcenia (*WYKSZ*), miejsce zamieszkania z podziałem na województwa (*WOJ*), a także miasto/wieś

(*MIASTO*), stopień pokrewieństwa z głową rodziny (*SP*), wcześniejszą sytuację na rynku pracy (*ROK_WCZ*), stan cywilny (*STCYW*) oraz grupy wiekowe (*GW*).

ZESTAWIENIE ZMIENNYCH OBJAŚNIAJĄCYCH

Zmienne	Opis zmiennych
	Grupa wiekowa
*GW_DO*24	poniżej 25 lat
*GW*_25*DO*34	25—34 lata
*GW*_35*DO*44	35—44 lata
*GW*_45*DO*54	45—54 lata
*GW*_55*DO*64	55—64 lata
*GW_OD*65	powyżej 65 lat
	Stopień pokrewieństwa z głową rodziny
*SP*_01	głowa gospodarstwa domowego
*SP*_02	mąż/żona
*SP*_03	partner/partnerka
*SP*_04	syn/córka
*SP*_05	zięć/synowa
*SP*_06	ojciec/matka/teść/teściowa
*SP*_07	dziadek/babcia/wnuk/wnuczka/prawnuk/prawnuczka
*SP*_08	brat/siostra
*SP*_09	wujek/ciotka/dalszy krewny
*SP*_10	niespokrewniony członek gospodarstwa domowego
	Województwo zamieszkania
*WOJ*_02	dolnośląskie
*WOJ*_04	kujawsko-pomorskie
*WOJ*_06	lubelskie
*WOJ*_08	lubuskie
*WOJ*_10	łódzkie
*WOJ*_12	małopolskie
*WOJ*_14	mazowieckie
*WOJ*_16	opolskie
*WOJ*_18	podkarpackie
*WOJ*_20	podlaskie
*WOJ*_22	pomorskie
*WOJ*_24	śląskie
*WOJ*_26	świętokrzyskie
*WOJ*_28	warmińsko-mazurskie
*WOJ*_30	wielkopolskie
*WOJ*_32	zachodniopomorskie
	Stan cywilny
*STCYW*_01	kawaler/panna
*STCYW*_02	żonaty/zamężna
*STCYW*_03	wdowiec/wdowa
*STCYW*_04	rozwiedziony(a)/w separacji
	Sytuacja na rynku pracy rok wcześniej
*ROK_WCZ*_1	praca
*ROK_WCZ*_2	bezrobocie
*ROK_WCZ*_3	nauka/szkolenie
*ROK_WCZ*_4	emerytura, wcześniejsza emerytura
*ROK_WCZ*_5	niepełnosprawność
*ROK_WCZ*_6	zasadnicza służba wojskowa
*ROK_WCZ*_7	obowiązki rodzinne
*ROK_WCZ*_8	inna forma bierności zawodowej

ZESTAWIENIE ZMIENNYCH OBJAŚNIAJĄCYCH (dok.)

Zmienne	Opis zmiennych
	Wykształcenie
WYKSZ_10	wyższe ze stopniem naukowym (co najmniej doktora), tytułem magistra lub równorzędnym, tytułem licencjata lub inżyniera, dyplom ukończenia kolegium policealne
WYKSZ_20	średnie zawodowe
WYKSZ_30	średnie ogólnokształcące
WYKSZ_40	zasadnicze zawodowe
WYKSZ_50	gimnazjalne
WYKSZ_60	podstawowe
WYKSZ_70	niepełne podstawowe
WYKSZ_80	bez wykształcenia szkolnego
WYKSZ_90	
	Miejsce zamieszkania (miasto/wieś)
MIASTO (*miasto*=1)	
WIEŚ (*wieś*=1)	

Ź r ó d ł o: opracowanie własne na podstawie formularzy statystycznych o symbolu ZD, ZG.

WYNIKI ESTYMACJI MODELI

W celach porównawczych oszacowano modele logitowe opisujące prawdopodobieństwo zakwalifikowania do grupy pracujących dla mężczyzn oraz dla kobiet. Wyniki oszacowań modeli zaprezentowano w tabl. 1—5.

TABL. 1. WYNIKI OSZACOWANIA MODELU LOGITOWEGO DLA KOBIET

Zmienne	Współczynnik	Błąd standardowy	*z*	Wartość *p*	Efekt krańcowy	Iloraz szans
const	1,8101	0,4904	3,6910	0,0002***	x	x
*GW_DO*24	–1,8799	0,4972	–3,7810	0,0002***	–0,1607	0,1526
*GW_*25*DO*34	–1,5199	0,4906	–3,0980	0,0020***	–0,0930	0,2187
*GW_*35*DO*44	–1,4944	0,4899	–3,0500	0,0023***	–0,0890	0,2244
*GW_*45*DO*54	–1,7590	0,4883	–3,6020	0,0003***	–0,1089	0,1722
*GW_*55*DO*64	–1,4088	0,4838	–2,9120	0,0036***	–0,0983	0,2444
*SP_*02	–0,3468	0,0906	–3,8260	0,0001***	–0,0143	0,7069
*SP_*03	–0,6848	0,1848	–3,7050	0,0002***	–0,0379	0,5042
*SP_*04	–0,7637	0,0967	–7,8960	0,0000***	–0,0398	0,4659
*SP_*08	–0,5328	0,2880	–1,8500	0,0644*	–0,0277	0,5870
*WOJ_*12	–0,2220	0,1146	–1,9380	0,0527*	–0,0099	0,8009
*WOJ_*18	0,3286	0,1199	2,7400	0,0061***	0,0118	1,3890
*WOJ_*26	–0,2739	0,1230	–2,2270	0,0259**	–0,0124	0,7604
*STCYW_*2	0,5094	0,0944	5,3970	0,0000***	0,0228	1,6643
MIASTO	–0,3749	0,0634	–5,9120	0,0000***	–0,0149	0,6874
*ROK_WCZ_*1	3,0370	0,0954	31,8400	0,0000***	0,3336	20,8418
*ROK_WCZ_*2	–0,8110	0,0964	–8,4090	0,0000***	–0,0456	0,4444
*ROK_WCZ_*3	0,7728	0,1374	5,6250	0,0000***	0,0230	2,1658
*ROK_WCZ_*4	1,2044	0,2599	4,6350	0,0000***	0,0297	3,3349
*WYKSZ_*10	0,9899	0,1244	7,9580	0,0000***	0,0348	2,6910
*WYKSZ_*20	0,3735	0,1619	2,3070	0,0211**	0,0131	1,4528
*WYKSZ_*30	0,4835	0,1162	4,1620	0,0000***	0,0176	1,6218
*WYKSZ_*40	0,4137	0,1287	3,2150	0,0013***	0,0146	1,5124
*WYKSZ_*50	0,2941	0,1144	2,5700	0,0102**	0,0111	1,3419
*WYKSZ_*60	0,9198	0,2490	3,6940	0,0002***	0,0252	2,5088

U w a g a. Poziom istotności parametrów: *** α=0,01, ** α=0,05, * α=0,10.

Ź r ó d ł o: obliczenia własne w programie Gretl.

TABL. 2. TABLICA TRAFNOŚCI DLA KOBIET

Faktyczne	Przewidywane		Razem
	$\hat{Y}=0$	$\hat{Y}=1$	
Y=0	963	1067	2030
Y=1	564	17759	18323
Razem	1527	18826	20353

Ź r ó d ł o: obliczenia własne.

TABL. 3. WYNIKI OSZACOWANIA MODELU LOGITOWEGO DLA MĘŻCZYZN

Zmienne	Współczynnik	Błąd standardowy	*z*	Wartość *p*	Efekt krańcowy	Iloraz szans
const	1,0428	0,4920	2,1200	0,0340**	x	x
*GW_DO*24	–1,5760	0,4963	–3,1750	0,0015***	–0,1041	0,2068
*GW_25DO*34	–1,6776	0,4914	–3,4140	0,0006***	–0,0949	0,1868
*GW_35DO*44	–1,8136	0,4907	–3,6960	0,0002***	–0,1105	0,1631
*GW_45DO*54	–1,9616	0,4891	–4,0100	0,0001***	–0,1223	0,1406
*GW_55DO*64	–1,8985	0,4869	–3,8990	0,0001***	–0,1350	0,1498
*SP_*01	0,9304	0,0874	10,6500	0,0000***	0,0376	2,5354
*SP_*02	–0,5086	0,1125	–4,5190	0,0000***	–0,0221	0,6013
*WOJ_*26	–0,1814	0,1080	–1,6790	0,0932*	–0,0070	0,8341
*WOJ_*32	–0,2601	0,1197	–2,1730	0,0298**	–0,0104	0,7710
*STCYW_*2	0,6608	0,0836	7,9070	0,0000***	0,0268	1,9364
MIASTO	–0,4298	0,0592	–7,2620	0,0000***	–0,0153	0,6506
*ROK_WCZ_*1	3,3065	0,0636	51,9700	0,0000***	0,3759	27,2897
*ROK_WCZ_*3	0,9475	0,1040	9,1100	0,0000***	0,0231	2,5793
*ROK_WCZ_*4	1,5591	0,2598	6,0010	0,0000***	0,0296	4,7545
*ROK_WCZ_*5	0,8678	0,1981	4,3810	0,0000***	0,0214	2,3816
*WYKSZ_*10	0,8813	0,1266	6,9600	0,0000***	0,0250	2,4140
*WYKSZ_*20	0,5035	0,2148	2,3440	0,0191**	0,0145	1,6546
*WYKSZ_*30	0,3512	0,1057	3,3230	0,0009***	0,0116	1,4208
*WYKSZ_*40	0,3432	0,1334	2,5720	0,0101**	0,0108	1,4094
*WYKSZ_*50	0,2135	0,0970	2,2000	0,0278**	0,0075	1,2380
*WYKSZ_*60	0,4128	0,1687	2,4460	0,0144**	0,0124	1,5110

U w a g a. Jak przy tabl. 1.
Ź r ó d ł o: jak przy tabl. 1.

TABL. 4. TABLICA TRAFNOŚCI DLA MĘŻCZYZN

Faktyczne	Przewidywane		Razem
	$\hat{Y}=0$	$\hat{Y}=1$	
Y=0	988	1169	2157
Y=1	592	21153	21745
Razem	1580	22322	23902

Ź r ó d ł o: jak przy tabl. 2.

W tabl. 3 i 5 przedstawiono liczbę trafień wartości prognozowanych 0 i 1 względem wartości rzeczywistych. Wartość prognozowaną zdefiniowano jako zmienną zero-jedynkową $\hat{Y}$ taką, że: gdy $\tilde{y}_i > 0{,}5$, to $\hat{y}_i = 1$, a gdy $\tilde{y}_i < 0{,}5$, to $\hat{y}_i = 0$, przy czym $\tilde{y}_i = F(x_i'\beta)$ jest oszacowaniem wyznaczonym z modelu.

TABL. 5. STATYSTYKA DOPASOWANIA MODELI LOGITOWYCH

Wyszczególnienie	Dla kobiet	Dla mężczyzn
R^2 McFaddena	0,3769	0,3645
Skorygowany R^2	0,3732	0,3614
Test ilorazu wiarygodności	4979,3	5280,8
Logarytm wiarygodności	–4115,1	–4604,3
Trafność prognozy w %: ogółem	92,0	92,6
dla Y=1	96,9	97,3
dla Y=0	47,4	45,8
Iloraz szans	28,4	30,2

U w a g a. Skorygowany R^2 jest to skorygowany stopniami swobody R^2 McFaddena.
Ź r ó d ł o: jak przy tabl. 2.

Wyniki testu ilorazu wiarygodności (wartość p=0,000<0,10) oznaczają istotność oszacowanych modeli. Ogólna procentowa trafność wartości prognozowanych przez modele była bardzo wysoka i wynosiła: model 1 — 92,0%, model 2 — 92,6%. Na uwagę zasługuje również fakt, że trafność prognoz dla Y=1 jest ponad dwukrotnie wyższa od trafności prognoz dla Y=0. Świadczy to, że model zidentyfikował mechanizm oddziaływania poszczególnych zmiennych na prawdopodobieństwo zakwalifikowania danej osoby do kategorii pracujących.

Nieistotne statystycznie zmienne modelu eliminowano metodą *a posteriori*. Procedura ta polega na porównaniu wartości p z przyjętym poziomem istotności. Jeżeli p było większe od α, zmienną eliminowano z modelu i dokonywano reestymacji. Działanie to powtarzano aż do uzyskania modelu z wartościami p mniejszymi lub równymi α. Wartość α przyjęto na poziomie 10%.

Podsumowanie

Z punktu widzenia walorów poznawczych badania największą wartość dają efekty krańcowe oraz ilorazy szans. Efekty krańcowe (obliczane dla średnich wartości zmiennych objaśniających) interpretuje się jako wpływ zmiennej objaśniającej na prawdopodobieństwo sukcesu (czyli przyjęcia wartości 1 przez zmienną objaśnianą). Ilorazy szans natomiast interpretuje się jako procentowy wpływ jednostkowej zmiany wartości zmiennej objaśniającej na iloraz szans, liczony jako stosunek prawdopodobieństwa sukcesu do prawdopodobieństwa porażki.

Na podstawie przedstawionych w tabl. 1 i 3 modeli logitowych można określić kierunek i siłę wpływu poszczególnych cech społeczno-demograficznych na prawdopodobieństwo zakwalifikowania poszczególnych osób do grupy pracujących. Wyznaczenie odrębnych modeli dla kobiet i mężczyzn pozwala na przeprowadzenie analizy porównawczej.

Pierwszym czynnikiem istotnie wpływającym na prawdopodobieństwo zakwalifikowania do grupy pracujących jest wiek, reprezentowany w modelu przez zmienne binarne opisujące poszczególne grupy wiekowe. Zarówno dla

kobiet, jak i dla mężczyzn przynależność do którejkolwiek kategorii wiekowej wiąże się ze spadkiem prawdopodobieństwa znalezienia się w grupie pracujących. Wiek do 24 lat zmniejsza prawdopodobieństwo bycia pracującym o 10,41 p.proc. w przypadku mężczyzn oraz o 16,07 p.proc. w przypadku kobiet. Przynależność do pozostałych grup wiekowych w przypadku mężczyzn silniej wpływa na prawdopodobieństwo zakwalifikowania do grupy pracujących aniżeli w przypadku kobiet.

Kolejnym czynnikiem istotnie wpływającym na prawdopodobieństwo przynależności do zbiorowości pracujących jest stopień pokrewieństwa z głową rodziny. W przypadku mężczyzn bycie głową gospodarstwa domowego powoduje wzrost owego prawdopodobieństwa o 3,76 p.proc., a szansa na to, że dany mężczyzna będzie osobą pracującą jest o 153,54% wyższa w porównaniu z innymi stopniami pokrewieństwa. W przypadku kobiet natomiast zmienna oznaczająca głowę gospodarstwa domowego okazała się statystycznie nieistotna. Gdy mężczyzna jest mężem głowy gospodarstwa domowego, to prawdopodobieństwo, że jest on osobą pracującą jest mniejsze o 2,21 p.proc., a szansa jest niższa o 39,87% w porównaniu z innymi stopniami pokrewieństwa. W przypadku kobiet wszystkie istotne zmienne opisujące stopień pokrewieństwa z głową gospodarstwa domowego wywierają ujemny wpływ na prawdopodobieństwo bycia pracującą. Największy spadek występuje wtedy, gdy kobieta jest córką głowy gospodarstwa domowego — wynosi on 3,98 p.proc., natomiast najmniejszy jest w sytuacji, gdy kobieta jest żoną głowy gospodarstwa domowego — wówczas wynosi on 1,43 p.proc.

Spośród zmiennych opisujących województwo zamieszkania, w modelu logitowym dla kobiet istotne okazały się te, które odnoszą się do województw małopolskiego, podkarpackiego oraz świętokrzyskiego. Spośród nich tylko zamieszkiwanie w woj. podkarpackim powoduje wzrost prawdopodobieństwa zakwalifikowania do grupy pracujących o 1,18 p.proc., a szansa na bycie pracującym wzrasta o 38,9%. Zamieszkiwanie w pozostałych województwach wiąże się z ujemnym wpływem na prawdopodobieństwo bycia pracującym. W przypadku mężczyzn istotne okazały się tylko dwie zmienne dotyczące województw świętokrzyskiego oraz zachodniopomorskiego. W obu przypadkach zamieszkiwanie w ich granicach powoduje spadek prawdopodobieństwa zaliczenia do pracujących średnio o odpowiednio 0,7 p.proc. oraz 1,04 p.proc.

W obu modelach spośród zmiennych opisujących stan cywilny jako istotna pozostała zmienna określająca osobę żonatą lub zamężną. Zarówno w przypadku kobiet, jak i w przypadku mężczyzn wywiera ona dodatni wpływ na prawdopodobieństwo przynależności do grupy pracujących, przy czym wpływ ten jest większy w przypadku mężczyzn i wynosi 2,68 p.proc., a dla kobiet — 2,28 p.proc. Szansa na to, że żonaty mężczyzna będzie pracującym jest o 93,64% większa w stosunku do pozostałych kategorii stanu cywilnego. W przypadku kobiet natomiast szansa na to, że zamężna kobieta będzie osobą pracującą jest większa o 66,43% w stosunku do pozostałych kategorii stanu cywilnego.

Następnym istotnym czynnikiem wpływającym na prawdopodobieństwo przynależności do pracujących jest miejsce zamieszkania. Zarówno w przypadku kobiet, jak i w przypadku mężczyzn zamieszkiwanie w mieście powoduje zmniejszenie prawdopodobieństwa bycia pracującym odpowiednio o 1,49 p.proc. oraz 1,53 p.proc. Kobieta mieszkająca w mieście ma mniejszą o 31,26% szansę na bycie pracującą w porównaniu z mieszkanką wsi. Mężczyzna mieszkający w mieście ma o 34,94% mniejszą szansę na bycie pracującym w porównaniu z mieszkańcem wsi.

Istotnym czynnikiem wpływu na prawdopodobieństwo bycia pracującym ma również status na rynku pracy sprzed roku. Zmiennymi, które wystąpiły w obu modelach są oznaczające: pracę, naukę/szkolenia oraz emeryturę lub wcześniejszą emeryturę. Każda z tych zmiennych powoduje wzrost prawdopodobieństwa zakwalifikowania danej osoby jako pracującej, przy czym największy wzrost powoduje posiadanie pracy rok wcześniej — w przypadku kobiet o 33,36 p.proc., a w przypadku mężczyzn o 37,59 p.proc. Ponadto w modelu dla kobiet jako istotna wystąpiła również zmienna oznaczająca bezrobocie. Zmienna ta wywiera ujemny wpływ na prawdopodobieństwo bycia pracującą średnio o 4,56 p.proc. W przypadku mężczyzn istotna okazała się zmienna opisująca niepełnosprawność. Bycie niepełnosprawnym rok wcześniej powoduje wzrost prawdopodobieństwa bycia pracującym średnio o 2,14 p.proc.

Ostatnia grupa zmiennych uwzględnionych w modelach dotyczyła wykształcenia. Zarówno w przypadku kobiet, jak i mężczyzn największy wzrost prawdopodobieństwa zakwalifikowania do grupy pracujących miały osoby legitymujące się wykształceniem wyższym — w odniesieniu do kobiet wzrost ten wynosił 3,48 p.proc., a w odniesieniu do mężczyzn — 2,14 p.proc. Najmniejszy, ale dodatni wpływ wywiera posiadanie zarówno przez kobiety, jak i mężczyzn wykształcenia zasadniczego zawodowego — wzrost odpowiednio o 1,11 p.proc. i 0,75 p.proc.

dr Dominik Śliwicki — *Urząd Statystyczny w Bydgoszczy, Wyższa Szkoła Gospodarki w Bydgoszczy*
mgr Marek Ręklewski — *Urząd Statystyczny w Bydgoszczy*

LITERATURA

Aktywność Ekonomiczna Ludności Polski IV kwartał 2010 (2011), GUS, Warszawa

Gruszczyński M. (2010), *Mikroekonometria. Modele i metody analizy danych indywidualnych*, Wolters Kluwer Polska, Warszawa

Kośko M., Osińska M., Stempińska J. (2007), *Ekonometria współczesna*, TNOiK, Toruń

Kufel T. (2011), *Ekonometria — rozwiązywanie problemów z wykorzystaniem programu Gretl*, PWN, Warszawa

Maddala G. S. (2008), *Ekonometria*, PWN, Warszawa

SUMMARY

The aging population poses to labor market policy challenges that, i.a. apply to support active aging people and entering the labour market for young people. According to the statistics in these two age groups is relatively a lot of unemployed people who are economically active but not working. The article presents the logit econometric models describing the likely of qualifying a group of persons classified as employed, according to the methodology of the Labour Force Survey, for the economically active category. Estimated separately for men and women models show the factors which have a significant impact on the classification them for the working group. The analysis identified differences in the direction and strength of the impact of specific factors on the likelihood of a person's credit to the category of working.

РЕЗЮМЕ

Старение населения ставит перед политикой рынка труда новые вызовы, которые касаются, в частности, поддержки профессиональной активности пожилых людей и введения на рынок труда молодых людей. По статистическим данным в этих двух возрастных группах относительно много безработных, которые являются профессионально активными, но они не работают. В статье представляются логитные эконометрические модели, характеризующие вероятность зачисления в группу занятых лиц определенных по методолгии Обследования экономической активности населения к категории активных с профессиональной точки зрения. Оценены отдельно для женщин и мужчин модели показывают факторы существенно влияющие на квалификацию их к группе занятых. В результате анализа были определены различия в направлении и силе воздействия конкретных факторов на вероятность зачисления данного лица к категории занятых.

Małgorzata JUCHNIEWICZ, Urszula TOMCZYK

Kapitał intelektualny sektora przedsiębiorstw na tle produktu krajowego brutto według województw

Niematerialne zasoby przedsiębiorstw budują kluczowe czynniki przewagi konkurencyjnej we współczesnej gospodarce. Największe szanse powodzenia w „grze" rynkowej mają podmioty, których konkurencyjność oparta jest na informacjach, wiedzy oraz szybkości generowania innowacyjnych pomysłów w zakresie produktów i usług, procesów technologicznych, rozwiązań organizacyjnych, a także na budowaniu relacji z otoczeniem (Limański, 2008).

W praktyce kapitał intelektualny określają definicje, które podali: Edvinsson i Malon, Steward oraz G. Roosa i J. Roosa. Edvinson i Malon na podstawie badania zasobów niematerialnych szwedzkiej firmy „Skandia" zdefiniowali kapitał intelektualny jako zbiór wartości kreowany przez kapitał ludzki i strukturalny, czyli wiedzę pracowników oraz wiedzę jednostki organizacyjnej (Kotarba, 2006). W ujęciu Stewarda kapitał intelektualny to ważny potencjał konkurencyjności przedsiębiorstwa, który może podwyższyć wycenę jego wartości finansowej (Haffer, 2006). Z kolei definicja kapitału intelektualnego podana przez G. Roosa i J. Roosa zakłada, że niematerialne zasoby przedsiębiorstw tworzone są przez kapitał ludzki, wiedzę, umiejętności, talenty oraz doświadczenie pracowników, a kapitał strukturalny tworzony jest przez własność intelektualną, procesy i kulturę organizacyjną. Kapitał strukturalny przedsiębiorstwa zdaniem autorów tworzą zatem: relacje, organizacja, odnowa oraz zdolność rozwoju.

Konkurencyjność w warunkach gospodarki globalnej stawia przed przedsiębiorstwami wymagania nieustannego dostosowywania działań do zmieniających się uwarunkowań budowania przewagi na rynku.

Jednym z najczęściej stosowanych mierników oceny rozwoju społeczno-gospodarczego jest PKB, który obliczany w cenach rynkowych równa się sumie wartości dodanej brutto wytworzonej przez wszystkie krajowe jednostki instytucjonalne powiększonej o podatki od produktów oraz pomniejszonej o dotacje do produktów (*Produkt…*, 2010). Jednym z najczęściej stosowanych mierników zamożności kraju, regionu lub województwa jest PKB *per capita*, czyli jego wartość przypadająca na osobę zamieszkującą dane terytorium. Pomimo że miernik ten nie uwzględnia różnic cen w analizowanych obszarach oraz produkcji gospodarstw domowych na własne potrzeby (np. praca gospodyń domowych), jest jednak jednym z najczęściej stosowanych mierników oceny rozwoju gospodarczego, poziomu zamożności i konkurencyjności regionów.

Problematyka konkurencyjności oraz pomiaru kapitału intelektualnego podejmowana jest w wielu badaniach i opracowaniach teoretycznych.

Ocena wartości zasobów jest zmienna i zależy m.in. od kontekstu działania przedsiębiorstw (Sopińska, 2010). Zdaniem autorki, z określonego profilu zasobów osiągnąć można różne rezultaty działalności oraz różny poziom konkurencyjności. Istotne znaczenie mają tu również uwarunkowania budowania przewagi konkurencyjnej regionu, które nie mają charakteru uniwersalnego, lecz tworzone są przez zróżnicowanie posiadanego potencjału rozwoju (Gawlikowska-Hueckel, 2008).

Celem artykułu jest ocena zróżnicowania kapitału intelektualnego sektora przedsiębiorstw w województwach na tle PKB *per capita*. Przedstawione w opracowaniu wyniki są etapem badania, którego głównym celem jest ocena poziomu kapitału intelektualnego sektora przedsiębiorstw w województwach w latach 2009—2011. W ramach badania dokonano systematyzacji koncepcji kapitału intelektualnego przedsiębiorstw w zakresie identyfikacji zasobów niematerialnych, narzędzi analitycznych służących jego pomiarowi oraz weryfikacja dostępności danych charakteryzujących kapitał intelektualny przedsiębiorstw w statystyce publicznej. Następnie opracowano model kapitału intelektualnego przedsiębiorstw pozwalający dokonać wyboru cech diagnostycznych, pomiaru jego poziomu oraz oceny na podstawie parametrów miernika taksonomicznego. Do obecnego opracowania wykorzystano dane za 2009 r., natomiast dane dotyczące PKB w okresie realizacji badania dostępne były za 2008 r. Kolejne etapy badania obejmować będą pomiar kapitału intelektualnego w latach 2010—2011 oraz ocenę zmiany jego poziomu w latach 2009—2011.

Praktyczny wymiar badania związany jest przede wszystkim z opracowaniem narzędzia służącego monitorowaniu poziomu rozwoju sektora przedsiębiorstw na podstawie zasobów niematerialnych, które umożliwi jego ocenę oraz analizę w ujęciu terytorialnym. Model pomiaru kapitału intelektualnego może służyć jednostkom samorządu terytorialnego oraz organizacjom i zrzeszeniom przedsiębiorców jako narzędzie identyfikacji zagadnienia rozwoju sektora przedsiębiorstw.

Zaproponowany przez autorów opracowania model pomiaru kapitału intelektualnego przedsiębiorstw stanowi rezultat prac badawczych opartych na analizie materiałów wtórnych oraz danych gromadzonych przez statystykę publiczną. Wyniki przeprowadzonych badań poddano analizie w ujęciu zależności pomiędzy poziomem kapitału intelektualnego w województwach i wartością PKB *per capita* tworzoną przez sektor przedsiębiorstw.

MODEL BADAWCZY POMIARU KAPITAŁU INTELEKTUALNEGO SEKTORA PRZEDSIĘBIORSTW

Podstawą konstrukcji modelu badawczego było zebranie danych dotyczących m.in. struktury podmiotowej gospodarki narodowej, jakości kapitału ludzkiego, działalności inwestycyjnej, innowacyjnej oraz relacji finansowych.

Utworzony zbiór cech diagnostycznych opracowano z zastosowaniem kluczowych kryteriów selekcji zmiennych w następujących etapach (Zeliaś, 2002): określenie celu i zakresu badań, gromadzenie danych, przyjęcie sposobu wyboru cech diagnostycznych oraz normalizacja i sumowanie zmiennych.

Korzystając z zasobów Banku Danych Lokalnych GUS oraz danych zamieszczonych na stronach WWW 16 urzędów statystycznych w województwach, utworzono zbiór 119 zmiennych. Pogrupowano je w 3 kategorie: kapitał ludzki, kapitał strukturalny koordynacji oraz kapitał strukturalny otoczenia.

Zgromadzony zbiór danych statystycznych poddano procedurze doboru cech diagnostycznych. W pierwszym kroku na podstawie współczynnika zmienności (V_i) dokonano redukcji zmiennych, których miara zróżnicowania nie dywersyfikowała województw (Strzała, Przechlewski, 2006). Biorąc pod uwagę wartość krytyczną współczynnika zmienności ($V_i \leq 0{,}15$) odrzucono 23 cechy. Pomimo osiągnięcia wartości niższej od krytycznego poziomu współczynnika zmienności nie odrzucono trzech zmiennych: aktywni zawodowo pracujący — wykształcenie policealne oraz średnie zawodowe; aktywni zawodowo pracujący — wykształcenie zasadnicze zawodowe; przedsiębiorstwa innowacyjne przemysłowe według rodzajów wprowadzonych innowacji. Zmienne te zaliczono do zbioru cech diagnostycznych ze względu na stosunkowo niewielkie przekroczenie wartości krytycznej współczynnika zmienności, a przede wszystkim z uwagi na istotne znaczenie w ocenie kapitału ludzkiego.

Kolejnym krokiem doboru cech diagnostycznych było określenie relacji zachodzących pomiędzy wartościami 96 zmiennych. Ocenę powiązań pomiędzy cechami, a następnie redukcję tych, które były nośnikami podobnych informacji przeprowadzono na podstawie macierzy współczynników korelacji liniowej przybierającej postać (Śmiłowska, 1997):

$$\mathbf{R} = \begin{pmatrix} 1 & r_{L2} & \cdots & r_{1L} \\ r_{2L} & 1 & \cdots & r_{2L} \\ \cdots & \cdots & \cdots & \cdots \\ r_{L1} & r_{L2} & \cdots & 1 \end{pmatrix}$$

co można zapisać w następujący sposób:

$$\mathbf{R} = (r_{pq})$$

gdzie $r_{pq} = r(x_{p,} x_q)\,(p, q = 1, 2, 3, ..., L; \quad p \neq q)$ oznacza współczynnik korelacji liniowej między wskaźnikami x_p i $x_{q.}$

Następnie na podstawie parametrycznej metody doboru cech diagnostycznych, zaproponowanej przez Hellwiga, dokonano redukcji macierzy korelacji,

której podstawą był parametr r^*, określany często jako krytyczna wartość współczynnika korelacji ($0 < r^* \leq 0{,}5$). Dyskryminacja cech została przeprowadzona za pomocą następujących czynności (Młodak, 2006):

1) wskazanie sumy elementów każdej kolumny (każdego wiersza) macierzy **R**:

$$\mathbf{R}_q = \sum_{p=1}^{L} |r_{jk}|$$

2) wskazanie kolumny o numerze *s*, posiadającej największą sumę elementów:

$$\mathbf{R}_k = \max_k \{\mathbf{R}_k\}$$

3) w kolumnie zostały wyróżnione elementy spełniające nierówność $|r_{jk}| > r^*$ oraz odpowiadające tym elementom wiersze. Cechę odpowiadającą wyróżnionej kolumnie uznano za pierwszą cechę centralną, natomiast cechy odpowiadające wyróżnionym wierszom za cechy satelitarne;
4) skreślenie wyróżnionych kolumn i wierszy z macierzy **R** i otrzymanie nowej, tzw. zredukowanej macierzy korelacji;
5) kontynuowanie postępowania opisanego w punktach 1—4 aż do momentu wyczerpania zbioru cech, pozostawienie tylko cech centralnych i izolowanych (czyli niemających cech satelitarnych).

Na podstawie opisanej procedury w zbiorze pozostało 41 cech diagnostycznych.

Ze względu na odmienne zakresy zmian cech diagnostycznych i przestrzenie uniemożliwiające ich porównanie, przeprowadzono standaryzację badanych parametrów (Nowak, 1990):

$$t_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}$$

gdzie:

t_{ij} — wartość zestandaryzowana cechy *j* dla województwa *i*,

x_{ij} — wartość cechy *xj* w województwie *i*,

$\bar{x}_j$ — średnia arytmetyczna cechy *j*,

S_j — odchylenie standardowe cechy *xj*.

Kolejny etap wyłonienia zmiennych diagnostycznych polegał na wyróżnieniu symulant oraz destymulant poziomu kapitału intelektualnego przedsiębiorstw. Za stymulanty uznano te zmienne, których wyższe wartości decydują o lepszym

poziomie rozpatrywanego zjawiska. Z kolei wysokie wartości cech: aktywni zawodowo pracujący — wykształcenie gimnazjalne, podstawowe i niższe oraz zwolnieni w drodze wypowiedzenia przez pracownika uznano za destymulanty, gdyż ich niskie wartości świadczą o pozytywnym rozwoju badanego zjawiska (Feltynowski, Nowakowska, 2009). W przypadku kapitału intelektualnego przedsiębiorstw, niskie wartości tych cech podwyższają zdolność kreowania kapitału intelektualnego przedsiębiorstw. Standaryzacja ma tu postać:

$$t_{ij} = -\frac{x_{ij} - \bar{x}_j}{S_j}$$

gdzie:

t_{ij} — wartość zestandaryzowana cechy *xj* dla województwa *i*,

x_{ij} — wartość cechy *xj* w województwie *i*,

$\bar{x}_j$ — średnia arytmetyczna cechy *xj*,

S_j — odchylenie standardowe cechy *xj*.

Kolejnym krokiem budowy wskaźnika syntetycznego kapitału intelektualnego przedsiębiorstw było zastosowanie metody sum standaryzowanych Perkala (W_i) według wzoru (Nowak, 1990):

$$W_i = \frac{1}{m}\sum_{j=1}^{n} t_{ij}$$

gdzie:

W_i — wskaźnik względnego poziomu kapitału intelektualnego,

t_{ij} — wartość zestandaryzowana cechy *j* dla województwa *i*,

m — liczba cech wchodząca w skład wskaźnika.

Ocenę wielkości kapitału intelektualnego województw przeprowadzono według sześciu grup typologicznych (tabl. 1).

Kapitał ludzki zdefiniowany został jako kapitał: podaży pracy, architektury zatrudnienia, postawy proedukacyjnej oraz zdolności komunikacyjnych. Zbiór zmiennych obejmował dane statystyczne dotyczące aktywnych zawodowo pracujących według: wykształcenia, sektorów ekonomicznych, wymiaru czasu pracy oraz wskaźniki świadczące o zdolności zatrudnionych do uczestnictwa w nowoczesnej wymianie informacji (przedsiębiorstwa otrzymujące i składające zamówienia na towary i usługi poprzez sieci komputerowe).

TABL. 1. KLASY WOJEWÓDZTW WEDŁUG MIARY OCENY POZIOMU KAPITAŁU INTELEKTUALNEGO PRZEDSIĘBIORSTW

Poziom kapitału intelektualnego województw	Miara oceny kapitału intelektualnego przedsiębiorstw
Wysoki	$W_i \geq \overline{W} + S_w$
Względnie wysoki	$\overline{W} + 0{,}5 \cdot S_w \leq W_i < \overline{W} + S_w$
Średni	$\overline{W} \leq W_i < \overline{W} + (0{,}5 \cdot S_w)$
Opóźniony	$\overline{W} - 0{,}5 \cdot S_w \leq W_i < \overline{W}$
Znacznie opóźniony	$\overline{W} - S_w \leq W_i < \overline{W} - 0{,}5 \cdot S_w$
Niski	$W_i \leq \overline{W} - S_w$

U w a g a. W_i — wskaźnik syntetyczny kapitału intelektualnego przedsiębiorstw, $\overline{W}$ — wartość przeciętna, S_w — odchylenie standardowe.

Ź r ó d ł o: na podstawie Strzała i Przechlewski (2006).

W przypadku kapitału strukturalnego koordynacji — składającego się z innowacji, inwestycji oraz wiedzy chronionej — zbiór cech diagnostycznych utworzyły cechy dotyczące odsetka przedsiębiorstw innowacyjnych, nakładów inwestycyjnych poniesionych przez przedsiębiorstwa na innowacje według rodzajów oraz wzorów użytkowych i wynalazków krajowych.

Z kolei kapitał strukturalny otoczenia utworzyły zmienne pogrupowane w cząstkowy kapitał intelektualny: lokalny, zagraniczny oraz relacji ekonomicznych. Zbiór zmiennych obejmował dane statystyczne dotyczące: zdolności przyciągania kapitału zagranicznego przez podmioty z kapitałem zagranicznym, środków uzyskanych z zagranicy na działalność innowacyjną w przedsiębiorstwach przemysłowych, środków przekazanych przedsiębiorstwom z budżetów gmin na działalność naukowo-badawczą oraz efektów działalności przedsiębiorstw w postaci rentowności obrotu netto.

Kapitały: ludzki, strukturalny koordynacji i strukturalny otoczenia stanowią składniki kapitału intelektualnego, dla których utworzono odrębne modele taksonomiczne, zgodnie z zaprezentowaną metodologią.

Zaproponowany model struktury kapitału intelektualnego przedsiębiorstw w województwach odpowiada filarom gospodarki opartej na wiedzy oraz takim systemom, jak: innowacyjność, edukacja, informacyjno-komunikacyjny otoczenia instytucjonalno-biznesowego i regionalnego w zarządzaniu wiedzą (Grudzewski, Hejduk, 2004).

WOJEWÓDZTWA POLSKI O WYSOKIM POZIOMIE KAPITAŁU INTELEKTUALNEGO SEKTORA PRZEDSIĘBIORSTW

Najwyższą miarą PKB *per capita* charakteryzowały się w 2009 r. województwa: mazowieckie, śląskie oraz dolnośląskie. Kapitał intelektualny przedsiębiorstw osiągnął tam wysoki poziom (tabl. 2). Sektor przedsiębiorstw w woj. mazowieckim konkurencyjność kreował dzięki najwyższej wartości w kraju

zasobów ludzkich, strukturalnej koordynacji oraz otoczenia. Mazowsze znajdowało się na pozycji lidera w rankingu województw ze względu na wartość PKB *per capita* oraz zasobów niematerialnych przedsiębiorstw. Woj. śląskie budowało konkurencyjną pozycję głównie na aktywach strukturalnych przedsiębiorstw: innowacjach, inwestycjach i wiedzy chronionej. Nieznacznie niższe wartości osiągnęły tam kapitały strukturalny otoczenia oraz ludzki. Z kolei wartość niematerialnych zasobów w woj. dolnośląskim kreowały w największym stopniu kapitały ludzki i strukturalny otoczenia. Nieznacznie mniejszy potencjał konkurencyjności przedsiębiorstw tworzyły aktywa strukturalne koordynacji.

W województwach: mazowieckim, śląskim i dolnośląskim wysokiej wartości kapitału intelektualnego przedsiębiorstw odpowiadały wysokie na tle Polski wartości PKB *per capita*. Sektor przedsiębiorstw miał tam duży potencjał konkurencyjności w postaci zasobów ludzkich, innowacji, inwestycji oraz wiedzy chronionej. Ponadto relacje z otoczeniem tworzyły wysokie wartości mierników, tj.: posiadanego kapitału zagranicznego, lokalnego czynnika sprawczego konkurencyjności oraz relacji z otoczeniem wskazującej na zdolność finansowania działalności i rozwoju przedsiębiorstw.

W przypadku wspomnianych województw zaobserwowano wysoki i względnie zbliżony poziom cząstkowych kapitałów zasobów intelektualnych sektora przedsiębiorstw, który dysponował potencjałem pozwalającym osiągnąć konkurencyjną pozycję na rynku krajowym.

Odmienne uwarunkowania potencjału konkurencyjności charakteryzowały sektor przedsiębiorstw woj. małopolskiego. Był on wyposażony w wysoki kapitał intelektualny, natomiast cechował się stosunkowo niską wartością PKB *per capita*. Poziom kapitału intelektualnego przedsiębiorstw kreowały głównie aktywa kapitału strukturalnego koordynacji tworzonego przez innowacje, inwestycje oraz wiedzę chronioną. W nieznacznie mniejszym stopniu konkurencyjność niematerialnych zasobów sektora małopolskich przedsiębiorstw oparta była na relacjach z otoczeniem. Poziom kapitału intelektualnego przedsiębiorstw znacząco obniżył potencjał zasobów ludzkich w tych województwach, których wynik oceny był niższy od wartości krytycznej (wartość krytyczna 0).

TABL. 2. WOJEWÓDZTWA O WYSOKIM POZIOMIE KAPITAŁU INTELEKTUALNEGO SEKTORA PRZEDSIĘBIORSTW W 2009 R. WEDŁUG PKB *PER CAPITA* ORAZ STANDARYZOWANYCH WARTOŚCI KAPITAŁU PRZEDSIĘBIORSTW

Województwa	PKB *per capita*[a]	Kapitał ludzki	Kapitał strukturalny koordynacji	Kapitał strukturalny otoczenia	Syntetyczny wskaźnik kapitału intelektualnego przedsiębiorstw
Mazowieckie	52770	1,18826	1,18853	1,44584	3,82263
Śląskie	36126	0,39982	0,78971	0,45766	1,64718
Dolnośląskie	35989	0,48717	0,38377	0,43956	1,31049
Małopolskie	28878	–0,17418	0,43039	0,34968	0,60589

a Dane za 2008 r.
Ź r ó d ł o: opracowanie własne na podstawie danych GUS.

W większości województw o najwyższej randze kapitału intelektualnego przedsiębiorstw wartość PKB *per capita* osiągnęła również najwyższe wielkości, wyjątkiem było tu woj. małopolskie.

WOJEWÓDZTWA POLSKI O WZGLĘDNIE WYSOKIM POZIOMIE KAPITAŁU INTELEKTUALNEGO SEKTORA PRZEDSIĘBIORSTW

Wartość wskaźnika syntetycznego zadecydowała o zaklasyfikowaniu województw podkarpackiego i łódzkiego do obszarów o względnie wysokim poziomie kapitału intelektualnego (tabl. 3). Pomiędzy wartością syntetycznego wskaźnika kapitału intelektualnego a wartością PKB *per capita* danego województwa nie zaobserwowano istotnej zależności statystycznej.

TABL. 3. WOJEWÓDZTWA O WZGLĘDNIE WYSOKIM POZIOMIE KAPITAŁU INTELEKTUALNEGO SEKTORA PRZEDSIĘBIORSTW W 2009 R. WEDŁUG PKB *PER CAPITA* ORAZ STANDARYZOWANYCH WARTOŚCI KAPITAŁU PRZEDSIĘBIORSTW

Województwa	PKB *per capita*[a]	Kapitał ludzki	Kapitał strukturalny koordynacji	Kapitał strukturalny otoczenia	Syntetyczny wskaźnik kapitału intelektualnego przedsiębiorstw
Podkarpackie	23101	–0,08251	0,76880	–0,21973	0,46656
Łódzkie	31140	0,40266	–0,21014	0,15827	0,35080

a Jak przy tabl. 2.
Ź r ó d ł o: jak przy tabl. 2.

Sektor przedsiębiorstw w woj. podkarpackim dysponował wysokiej wartości zasobami strukturalnymi koordynacji, tworzonymi w największym stopniu przez innowacje oraz inwestycje. Z kolei ocenę kapitału strukturalnego otoczenia w największym stopniu obniżyła ocena zasobów lokalnych przedsiębiorstw, relacji z otoczeniem oraz kapitału zagranicznego. Również zasoby ludzkie, czyli podaż pracy oraz postawy proedukacyjne nie budowały przewagi konkurencyjnej przedsiębiorstw. Względnie wysoki poziom kapitału intelektualnego przedsiębiorstwa zawdzięczały wysokiej jakości zasobów strukturalnych koordynacji, jednak w rankingu według wartości PKB *per capita* woj. podkarpackie zajęło 16. miejsce, tuż za województwami o niskim poziomie kapitału intelektualnego.

W woj. łódzkim o względnie wysokim kapitale intelektualnym, sektor przedsiębiorstw kreował konkurencyjność dzięki stosunkowo wysokiej wartości zasobów ludzkich oraz strukturalnych otoczenia. Największą wartość tych zasobów tworzył kapitał podaży pracy, architektura zatrudnienia, relacje ekonomiczne oraz kapitał lokalny. Poziom aktywów niematerialnych przedsiębiorstw obniżyła ocena wartości kapitału strukturalnego koordynacji, kreowanego przez inwestycje i innowacje. Relatywnie wysokiej wartości zasobów ludzkich i otoczenia przedsiębiorstw w woj. łódzkim oraz niskiemu poziomowi zasobów koordynacji odpowiadała stosunkowo wysoka na tle kraju wartość rezultatów działalności podmiotów gospodarki narodowej. Wartość PKB *per capita* pozwoliła uplasować omawiane województwo w rankingu na 6. miejscu.

W omawianych tu województwach źródłem konkurencyjności przedsiębiorstw był odmienny układ aktywów niematerialnych. Wysokiej wartości kapitału strukturalnego koordynacji przedsiębiorstw w woj. podkarpackim odpowiadały niskie wartości zasobów ludzkich oraz kapitału strukturalnego otoczenia. Z kolei w woj. łódzkim względnie wysoką ocenę uzyskał kapitał ludzki oraz strukturalny otoczenia, natomiast niską — zasoby strukturalne koordynacji.

WOJEWÓDZTWA POLSKI CHARAKTERYZUJĄCE SIĘ OPÓŹNIENIEM LUB ZNACZNYM OPÓŹNIENIEM POZIOMU KAPITAŁU INTELEKTUALNEGO SEKTORA PRZEDSIĘBIORSTW

Pomimo braku istotności statystycznej pomiędzy wartością kapitałów cząstkowych i wartością syntetycznego wskaźnika kapitału intelektualnego i wartością PKB *per capita*, podobne zależności zaobserwowano w województwach, w których poziom zasobów niematerialnych przedsiębiorstw świadczył o opóźnieniu lub znacznym opóźnieniu w porównaniu z przeciętnym poziomem w kraju.

Jak wynika z tabl. 4 kapitał intelektualny przedsiębiorstw woj. wielkopolskiego tworzyły względnie wysokiej jakości zasoby ludzkie i strukturalne otoczenia, natomiast wartość aktywów niematerialnych przedsiębiorstw obniżyła ocena aktywów strukturalnych koordynacji. Z kolei wysokiej ocenie wartości zasobów ludzkich i strukturalnych otoczenia oraz niskiej ocenie wartości strukturalnych koordynacji towarzyszyła stosunkowo wysoka na tle kraju wartość PKB *per capita*.

Odmienny układ oceny, tj. niski zasobów ludzkich i strukturalnych otoczenia oraz względnie wysoki kapitału strukturalnego koordynacji determinował znaczne opóźnienie poziomu kapitału intelektualnego przedsiębiorstw w woj. kujawsko-pomorskim. Sektor przedsiębiorstw tego województwa, budując konkurencyjną pozycję opartą głównie na zasobach strukturalnych, uplasował się na 8. miejscu, niższym niż woj. wielkopolskie, w rankingu rezultatów wyrażonych w PKB *per capita*.

TABL. 4. WOJEWÓDZTWA CHARAKTERYZUJĄCE SIĘ OPÓŹNIENIEM LUB ZNACZNYM OPÓŹNIENIEM POZIOMU KAPITAŁU INTELEKTUALNEGO PRZEDSIĘBIORSTW W 2009 R. WEDŁUG PKB *PER CAPITA* ORAZ STANDARYZOWANYCH WARTOŚCI KAPITAŁU PRZEDSIĘBIORSTW

Województwa	PKB *per capita*[a]	Kapitał ludzki	Kapitał strukturalny koordynacji	Kapitał strukturalny otoczenia	Syntetyczny wskaźnik kapitału intelektualnego przedsiębiorstw
Region charakteryzujący się opóźnieniem poziomu kapitału intelektualnego przedsiębiorstw					
Wielkopolskie	34934	0,06634	–0,16159	0,01283	–0,08242
Region charakteryzujący się znacznym opóźnieniem poziomu kapitału intelektualnego przedsiębiorstw					
Kujawsko-pomorskie	28926	–0,11539	0,09173	–0,16279	–0,18645

a Jak przy tabl. 2.
Ź r ó d ł o: jak przy tabl. 2.

WOJEWÓDZTWA POLSKI O NISKIM POZIOMIE KAPITAŁU INTELEKTUALNEGO SEKTORA PRZEDSIĘBIORSTW

Niski poziom zasobów kapitału intelektualnego przedsiębiorstw charakteryzował województwa: pomorskie, świętokrzyskie, lubelskie, opolskie, lubuskie, podlaskie, warmińsko-mazurskie i zachodniopomorskie. Zasoby niematerialne w wymienionych województwach osiągnęły wielkości niższe od krytycznej (wielkość krytyczna 0). Poziom wyższy od wielkości krytycznej osiągnęła jedynie ocena kapitału strukturalnego koordynacji w woj. pomorskim oraz strukturalnego otoczenia w woj. świętokrzyskim.

Miejsce w rankingu dotyczące rezultatów działalności podmiotów gospodarki narodowej wyrażonych w PKB *per capita* w niewielkim stopniu różnicowało województwa o niskim poziomie kapitału intelektualnego przedsiębiorstw.

Województwa z tą kategorią przedsiębiorstw zajęły najniższą pozycję w rankingu PKB *per capita* wyprzedzając jedynie woj. podkarpackie, reprezentujące klasę względnie wysokiego poziomu zasobów niematerialnych przedsiębiorstw. Wyjątek stanowią województwa pomorskie i zachodniopomorskie, w których sektor przedsiębiorstw osiągnął miejsce odpowiednio 4. i 7. w rankingu rezultatów wyrażonych w PKB *per capita*.

W przypadku woj. pomorskiego stosunkowo wysoka ocena kapitału strukturalnego koordynacji tworzona była głównie przez zasoby inwestycji i innowacji przedsiębiorstw oraz jeden z kapitałów cząstkowych, tj. relacje ekonomiczne, który kreował wartości zasobów budowania przewagi konkurencyjnej. Podobnie w przypadku kapitału cząstkowego — postawy proedukacyjne osób w wieku 25—64 lata — który stanowił cenny potencjał w zasobach przedsiębiorstw.

W woj. zachodniopomorskim zasoby niematerialne przedsiębiorstw oceniono poniżej wartości krytycznej. Wśród kapitałów cząstkowych tylko postawy proedukacyjne zasobów ludzkich wyróżniały województwo na tle innych.

TABL. 5. WOJEWÓDZTWA O NISKIM POZIOMIE KAPITAŁU INTELEKTUALNEGO SEKTORA PRZEDSIĘBIORSTW W 2009 R. WEDŁUG PKB *PER CAPITA* ORAZ STANDARYZOWANYCH WARTOŚCI KAPITAŁU PRZEDSIĘBIORSTW

Województwa	PKB *per capita*[a]	Kapitał ludzki	Kapitał strukturalny koordynacji	Kapitał strukturalny otoczenia	Syntetyczny wskaźnik kapitału intelektualnego przedsiębiorstw
Pomorskie	31754	–0,22336	0,12512	–0,09939	–0,19763
Świętokrzyskie	26763	–0,11223	–0,31376	0,04041	–0,38559
Lubelskie	23219	–0,04992	–0,15690	–0,71027	–0,91709
Opolskie	28379	–0,38893	–0,51819	–0,12925	–1,03637
Lubuskie	28709	–0,03864	–0,79940	–0,23779	–1,07583
Podlaskie	24434	–0,41847	–0,43825	–0,22316	–1,07989
Warmińsko-mazurskie	24814	–0,39349	–0,54778	–0,66955	–1,61083
Zachodniopomorskie	30357	–0,54713	–0,63201	–0,45232	–1,63146

a Jak przy tabl. 2.
Ź r ó d ł o: jak przy tabl. 2.

Wnioski

Przeprowadzone badanie wskazuje na brak istotnej statystycznie zależności między poziomem kapitału intelektualnego przedsiębiorstw i wielkością PKB *per capita*. Zaobserwowano pewne relacje pomiędzy tymi wskaźnikami, ale odnoszą się one jedynie do województw o wysokim poziomie kapitału intelektualnego, w których zasoby przedsiębiorstw osiągają wysoką i względnie zbliżoną wielkość poziomu kapitałów cząstkowych: ludzkiego, strukturalnego koordynacji i otoczenia oraz wysoką wielkość PKB *per capita*.

Niskiemu poziomowi regionalnemu kapitału intelektualnego przedsiębiorstw odpowiadała stosunkowo niska wielkość PKB *per capita*. Odstępstwem były tu województwa pomorskie i zachodniopomorskie charakteryzujące się niską wartością poziomu kapitału intelektualnego i jednocześnie stosunkowo wysokim PKB *per capita*.

Wysoka i względnie zbliżona wielkość zasobów tworzących kapitał intelektualny pozwala przedsiębiorstwom osiągać wysokie wartości PKB *per capita* w województwach. Zróżnicowana wielkość cząstkowych kapitałów zasobów niematerialnych sektora przedsiębiorstw powoduje, że otrzymują różne rezultaty oraz odmienny poziom konkurencyjności. Brak istotnego statystycznie powiązania pomiędzy zasobami niematerialnymi i rezultatami działalności przedsiębiorstw wskazuje, że uwarunkowania budowania przewagi konkurencyjnej nie mają charakteru uniwersalnego, lecz tworzone są przez zdolność wykorzystania zróżnicowanego potencjału rozwoju.

Zaproponowaną metodę analizy poziomu kapitału intelektualnego oparto na taksonomicznych metodach klasyfikacji umożliwiających grupowanie województw według wartości zasobów niematerialnych przedsiębiorstw. Przeprowadzona analiza pozwoliła określić, czy zachodzą relacje pomiędzy osiąganą wielkością poziomu kapitału intelektualnego przedsiębiorstw w województwach i wielkością PKB *per capita*.

Procedura doboru mierników diagnostycznych oraz model klasyfikacji województw stanowić może narzędzie systematycznej obserwacji dynamiki rozwoju kapitału intelektualnego i wyróżnienia tych mierników, które w sposób istotny wpływają na jego poziom w badanym okresie.

dr hab. Małgorzata Juchniewicz — *Uniwersytet Warmińsko-Mazurski w Olsztynie*,
mgr Urszula Tomczyk — *Instytut Badań i Analiz Grupa OSB w Olsztynie*

LITERATURA

Feltynowski M., Nowakowska A. (2009), *Ranking potencjału innowacyjnego polskich regionów z wykorzystaniem miar syntetycznych*, [w:] *Zdolności innowacyjne polskich regionów*, red. A. Nowakowska, Wydawnictwo Uniwersytetu Łódzkiego

Gawlikowska-Hueckel K. (2008), *Potencjał rozwojowy polskich województw w latach 1995—2005*, [w:] *Lokalizacja przemysłu a konkurencyjność polskich regionów (w kontekście integracji europejskiej)*, red. A. Zielińska-Głębocka, Wydawnictwo Uniwersytetu Gdańskiego

Grudzewski W. M., Hejduk I. K. (2004), *Zarządzanie wiedzą w przedsiębiorstwach*, Centrum Doradztwa i Informacji Difin, Warszawa

Haffer R. (2006), *Metody i narzędzia pomiaru zasobów wiedzy i oceny poziomu zarządzania wiedzą w przedsiębiorstwach*, [w:] *Zarządzanie wiedzą jako kluczowy czynnik międzynarodowej konkurencyjności przedsiębiorstwa*, red. M. J. Stankiewicz, Towarzystwo Naukowe Organizacji i Kierownictwa Stowarzyszenie Wyższej Użyteczności „Dom Organizatora", Toruń

Kotarba W. (2006) (red.), *Ochrona wiedzy a kapitał intelektualny organizacji*, PWE, Warszawa

Limański A. (2008), *Rola i znaczenie wiedzy w zarządzaniu współczesnym przedsiębiorstwem*, [w:] *Uwarunkowania rynkowe konkurencyjności przedsiębiorstw*, red. A. Birski, Uniwersytet Warmińsko-Mazurski w Olsztynie

Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej*, Centrum Doradztwa i Informacji Difin, Warszawa

Nowak E. (1990), *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*, PWE, Warszawa

Produkt Krajowy Brutto. Rachunki Regionalne w 2008 r. (2010), Urząd Statystyczny w Katowicach, GUS

Sopińska A. (2010), *Wiedza jako strategiczny zasób przedsiębiorstwa. Analiza i pomiar kapitału intelektualnego*, Wydanie II, Oficyna Wydawnicza SGH

Strzała K., Przechlewski T. (2006), *Ekonometria inaczej*, Wydawnictwo Uniwersytetu Gdańskiego

Śmiłowska T. (1997), *Statystyczna analiza poziomu życia ludności Polski w ujęciu przestrzennym*, „Studia i prace. Z prac Zakładu Badań Statystyczno-Ekonomicznych GUS i PAN", nr 247, GUS

Zeliaś A. (2002), *Some Notes on the Selection of Normalization of Diagnostic Variables*, „Statistics in Transition", vol. 5, No. 5

SUMMARY

This paper presents the results of a study on the relationship between the level of the intellectual capital of companies in the voivodships and the GDP per capita. Used for the analysis a taxonomic classification method, allowed voivodship grouping by intangible resources of companies. On this basis, the selection of diagnostic features, measurement and evaluation of the intellectual capital of the business sector in the provinces was made. Next, the values of correlation coefficients were defined, which showed no statistically significant correlation between the level of intellectual capital in companies and the GDP per capita. There has been, however, determined the relationship between these indicators. Voivodships, in which the intangible assets of enterprises have achieved a high and relatively similar size of the human capital level, structural and environmental conditions also have a high GDP per capita.

РЕЗЮМЕ

Статья представляет результаты анализа идентификации взаимосвязи между уровнем интеллектуального капитала предприятий в воеводствах и размером ВВП на душу населения. Для анализа был использован таксономический метод классификации, позволяющий группировать воеводства по нематериальным ресурсам предприятий. На этой основе был сделан выбор диагностических признаков, измерения и оценки интеллектуального капитала сектора предприятий в воеводствах. Затем были определены значения коэффициентов корреляции, которые не показали значимой со статистической точки зрения зависимости между уровнем интеллектуального капитала в предприятиях и размером ВВП на душу населения. Тем не менее были определены связи между этими показателями. Воеводства, в которых нематериальные активы предприятий добились высокого и относительно сближенного уровня человеческого, структурного капиталов, а также капитала окружающей среды имели также высокий ВВП на душу населения.

Mirosław GORCZYCA

Wydatki na mieszkanie w budżetach gospodarstw domowych[1]

Wydatki na mieszkanie stanowią drugą po wydatkach na żywność pozycję w budżetach polskich gospodarstw domowych. „Ekonomizacja" usług mieszkaniowych i komplementarnych sprawiła, że rósł ich poziom i udział. Na przykład udział wydatków na mieszkanie zwiększył się z 15% w 1991 r. do 19% w 2000 r. W pierwszej dekadzie XXI w. co piąty złoty w budżetach polskich rodzin wydawany był na mieszkanie *sensu stricto* (koszty utrzymania mieszkania i nośników energii). Ich wielkość determinuje wiele czynników. Są to przede wszystkim: standard mieszkań, charakter własności oraz zamożność użytkowników i racjonalność korzystania z nośników energii. Ważnym czynnikiem kreującym wysokość wydatków na mieszkanie jest wielkość gospodarstw domowych. Ponad 50% kosztów utrzymania mieszkania stanowią koszty stałe, związane z powierzchnią mieszkaniową, które są niezależne od liczby osób w gospodarstwach.

WYDATKI NA MIESZKANIE W RÓŻNYCH GRUPACH GOSPODARSTW

Przed przystąpieniem do analizy przedstawię relacje między dochodami rozporządzalnymi a wydatkami ogółem (tabl. 1). Konfrontacja ta pozwala na ukazanie rosnącej nadwyżki dochodów nad wydatkami.

Wydatki na mieszkanie w 2011 r. to ok. 210 zł miesięcznie na osobę i niespełna 21% ogółu wydatków w gospodarstwach domowych. Ich poziom i udział był wyraźnie zróżnicowany w poszczególnych grupach społeczno-ekonomicznych (tabl. 2).

Różnica wartości krańcowych wynosiła ok. 144 zł miesięcznie na osobę (od niespełna 133 zł w gospodarstwach domowych rolników do ok. 276 zł u emerytów). W ujęciu względnym udział wydatków na mieszkanie w ich ogólnym poziomie różnił się o od ok. 18% w pracowniczych nierobotniczych i rolniczych gospodarstwach domowych do blisko 27% u rencistów. Niski poziom wydatków na mieszkanie w gospodarstwach domowych rolników wynikał z tego, że ich mieszkania prawie w 100% należały do ich użytkowników oraz z niższego standardu instalacyjnego mieszkań. Jednak znaczne różnice między wysokością wynikały z faktu, że w gospodarstwach emerytów średnia liczba osób była ok. 2 razy niższa niż w gospodarstwach domowych rolników.

[1] W artykule korzystano z publikacji: *Budżety gospodarstw domowych* (edycje dla odpowiednich lat), GUS; Gorczyca M. (2008), *Stan i perspektywy rozwoju mieszkalnictwa w Polsce — na tle wybranych krajów,* Wyższa Szkoła Informatyki i Zarządzania w Rzeszowie (materiały powielane); Gorczyca M. (2011), *Wydatki na mieszkanie, zaległości w opłatach za ich użytkowanie oraz dodatki mieszkaniowe w 2011 r*., ZRSM RP, Warszawa (maszynopis); *Statistical Abstract of the United States U.S.* (edycje dla odpowiednich lat), Census Bureau, Washington; *Consumption expenditure of households on goods and services, EU-27*, www.eurostat.eu.

TABL. 1. PRZECIĘTNE MIESIĘCZNE DOCHODY I WYDATKI NA OSOBĘ W GOSPODARSTWACH DOMOWYCH

Wyszczególnienie	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
						W zł						
Dochód rozporządzalny	610,51	644,48	664,21	711,96	735,40	761,46	834,68	928,87	1045,52	1114,49	1192,82	1226,95
Wydatki	599,49	609,72	624,99	677,81	694,70	690,30	744,81	809,95	904,27	956,68	991,44	1015,12
Nadwyżka dochodów nad wydatkami	11,02	34,76	39,22	34,15	40,70	71,16	89,97	118,92	141,25	157,81	201,38	211,83
						W %						
Nadwyżka dochodów nad wydatkami	1,8	5,7	6,3	5,0	5,9	10,3	12,1	14,7	15,6	16,5	20,3	20,9

Ź r ó d ł o: obliczenia na podstawie *Budżety gospodarstw domowych* (edycje dla odpowiednich lat), GUS.

TABL. 2. MIESIĘCZNE WYDATKI NA OSOBĘ NA MIESZKANIE WEDŁUG GRUP SPOŁECZNO-EKONOMICZNYCH GOSPODARSTW DOMOWYCH W 2011 R.

Wydatki	Ogółem	W tym gospodarstwa domowe								
		pracowników			rolników	pracujących na własny rachunek	emerytów i rencistów		
		razem	na stanowiskach				razem	emerytów	rencistów
			robotniczych	nierobotniczych					
					W zł				
O g ó ł e m	**1015,12**	**1006,05**	**771,01**	**1297,62**	**736,02**	**1221,52**	**1080,14**	**1124,91**	**897,86**
W tym na użytkowanie mieszkania i nośniki energii	210,34	193,41	160,80	233,85	132,52	226,97	269,07	276,13	240,32
					W %				
W tym na użytkowanie mieszkania i nośniki energii	20,7	19,2	20,9	18,0	18,0	18,6	24,9	24,5	26,8

Ź r ó d ł o: obliczenia na podstawie *Budżety gospodarstw domowych w 2011 r.,* GUS, 2012.

WYDATKI NA MIESZKANIE W GOSPODARSTWACH DOMOWYCH WEDŁUG LICZBY OSÓB

Różnice w poziomie jednostkowych wydatków na mieszkanie egzemplifikują dane o ich wysokości w gospodarstwach domowych różnej wielkości (tabl. 3).

TABL. 3. MIESIĘCZNE WYDATKI NA OSOBĘ NA MIESZKANIE GOSPODARSTW DOMOWYCH WEDŁUG LICZBY OSÓB W 2011 R.

Wydatki	Ogółem	Gospodarstwa domowe o liczbie osób					
		1	2	3	4	5	6 i więcej
				W zł			
O g ó ł e m	**1015,12**	**1647,90**	**1370,54**	**1099,08**	**904,62**	**720,62**	**574,76**
W tym na użytkowanie mieszkania i nośniki energii	210,34	431,27	297,96	216,27	168,84	136,06	111,21
				W %			
W tym na użytkowanie mieszkania i nośniki energii	20,7	26,2	21,7	19,7	18,7	20,3	19,3

Ź r ó d ł o: obliczenia na podstawie *Budżety gospodarstw domowych w 2011 r.*, GUS, 2012.

Przeciętny poziom wydatków miesięcznych na mieszkanie w gospodarstwach jednoosobowych stanowił 388% wysokości wydatków w gospodarstwach składających się z co najmniej 6 osób. Tymczasem wydatki ogółem stanowiły mniej, bo 287%.

WYDATKI NA MIESZKANIE W MIASTACH I NA WSI ORAZ W GRUPACH DOCHODOWYCH GOSPODARSTW DOMOWYCH

Wykorzystując dane GUS o budżetach gospodarstw dotyczące 2007 r., w tabl. 4 i 5 przedstawiono dane o poziomie wydatków na mieszkanie w miastach różnej wielkości oraz grupach dochodowych gospodarstw domowych, czyli w miarę wzrostu wielkości miast rósł jednostkowy poziom wydatków (tabl. 4).

TABL. 4. MIESIĘCZNE WYDATKI NA OSOBĘ NA MIESZKANIE W MIASTACH I NA WSI W 2007 R.

Wydatki	Miasta						Wieś
	razem	o liczbie mieszkańców w tys.					
		do 20	20—100	100—200	200—500	powyżej 500	
				W zł			
O g ó ł e m	**907,75**	**747,27**	**816,76**	**866,17**	**964,42**	**1205,22**	**652,57**
W tym na utrzymanie mieszkania i nośniki energii	172,14	145,24	161,21	171,32	180,17	212,18	112,19

TABL. 4. MIESIĘCZNE WYDATKI NA OSOBĘ NA MIESZKANIE W MIASTACH I NA WSI W 2007 R. (dok.)

Wydatki	Miasta						Wieś
	razem	o liczbie mieszkańców w tys.					
		do 20	20—100	100—200	200—500	powyżej 500	
	W %						
W tym na utrzymanie mieszkania i nośniki energii	19,0	19,4	19,7	19,8	19,4	17,6	17,2

Ź r ó d ł o: obliczenia na podstawie *Budżety gospodarstw domowych w 2007 r.,* GUS, Warszawa 2008.

Jeśli za 100% przyjmiemy poziom wydatków na mieszkanie gospodarstw domowych na wsi, to widać, że w miarę wzrostu wielkości miast zwiększała się relacja wydatków miast do wsi, co pokazuje następująca tablica:

TABL. 5. RELACJA WYDATKÓW NA MIESZKANIE MIAST DO WSI W % WEDŁUG LICZBY MIESZKAŃCÓW

Wydatki	Liczba mieszkańców miast w tys.				
	do 20	20—100	100—200	200—500	powyżej 500
O g ó ł e m	**115**	**125**	**133**	**148**	**185**
W tym na utrzymanie mieszkania i nośniki energii	129	144	153	161	189

Poziom jednostkowych wydatków na mieszkanie w znaczącym stopniu zależy też od zamożności gospodarstw domowych. Na ogół te o lepszej kondycji ekonomicznej zajmują lokale o wyższym standardzie i stąd większe wydatki jednostkowe (tabl. 6).

TABL. 6. MIESIĘCZNE WYDATKI NA OSOBĘ NA MIESZKANIE WEDŁUG GRUP KWINTYLOWYCH DOCHODÓW GOSPODARSTW DOMOWYCH W 2007 R.

Wydatki	Ogółem	Grupy kwintylowe dochodów				
		I	II	III	IV	V
	W zł					
O g ó ł e m	**809,95**	**396,72**	**537,30**	**700,68**	**913,30**	**1504,67**
W tym na użytkowanie mieszkania i nośniki energii	149,14	73,32	107,69	140,51	178,40	246,18
	W %					
W tym na użytkowanie mieszkania i nośniki energii	18,4	18,5	20,0	20,5	19,5	16,4

Ź r ó d ł o: obliczenia na podstawie *Budżety gospodarstw domowych w 2007 r.,* GUS, 2008.

Jednostkowy poziom wydatków na mieszkanie w 1/5 najlepiej sytuowanych gospodarstw domowych objętych badaniami budżetów był o 236% wyższy niż w 1/5 najuboższych. Była to rozpiętość nieznacznie niższa od rozpiętości poziomu jednostkowych wydatków ogółem (279%).

ZMIANY POZIOMU I UDZIAŁU WYDATKÓW NA MIESZKANIE

Wydatki na mieszkanie na ogół rosły z roku na rok w nieco szybszym tempie aniżeli wydatki ogółem (tabl. 7).

TABL. 7. MIESIĘCZNE WYDATKI NA OSOBĘ NA MIESZKANIE W GOSPODARSTWACH DOMOWYCH W CENACH BIEŻĄCYCH

Wydatki	2000	2001	2002	2003	2004	2005
			W zł			
O g ó ł e m	**599,49**	**609,72**	**624,99**	**677,81**	**694,70**	**690,30**
W tym na użytkowanie mieszkania i nośniki energii	107,21	114,91	124,47	142,41	140,78	135,64
			W %			
W tym na użytkowanie mieszkania i nośniki energii	17,9	18,8	19,9	21,0	20,3	19,6

(dok.)

Wydatki	2006	2007	2008	2009	2010	2011
			W zł			
O g ó ł e m	**744,81**	**809,95**	**904,27**	**956,68**	**991,44**	**1015,12**
W tym na użytkowanie mieszkania i nośniki energii	146,34	149,14	170,80	188,03	199,88	210,34
			W %			
W tym na użytkowanie mieszkania i nośniki energii	19,7	18,4	18,9	19,7	20,2	20,7

Ź r ó d ł o: obliczenia na podstawie *Budżety gospodarstw domowych* (edycje dla odpowiednich lat), GUS.

Wydatki ogółem w 2011 r. były wyższe w porównaniu z 2000 r. o 69%, natomiast wydatki na mieszkanie o ponad 96%. Zwiększył się więc udział wydatków na mieszkanie w ogólnych wydatkach gospodarstw do 20,7% w 2011 r. Rozpatrując zmiany poziomu wydatków na mieszkanie należy pamiętać o tym, że ich wzrost wynikał z podwyżek cen nośników energii.

Należy sądzić, że wydatki na mieszkanie będą rosły, gdyż ceny usług mieszkaniowych zwiększyły się w latach 2010—2012 (rok do roku) odpowiednio o: 3,8%, 6,2%, 6,1%, a wskaźniki cen towarów i usług konsumpcyjnych kolejno o: 2,6%, 4,3%, 4,0%.

WYDATKI NA MIESZKANIE W POLSCE NA TLE INNYCH KRAJÓW

Wydatki na mieszkanie w naszym kraju stanowią niespełna 20% ogółu wydatków gospodarstw domowych. W krajach o wyższym standardzie mieszkaniowym stanowią one od lat najwyższy ich składnik, przewyższając znacząco udział wydatków na żywność. Najwyższy udział wydatków na mieszkanie jest w Stanach Zjednoczonych, w których stanowią one ok. 1/3 wszystkich wydatków i są ok. 3 razy wyższe od wydatków na żywność. Nieco niższy udział występuje w Kanadzie i Izraelu. Jak podaje Eurostat, wydatki na mieszkanie w 27 krajach Unii Europejskiej (UE) to 21,7% wydatków konsumpcyjnych, a dla całej Europy — 22,0% (w 2005 r.).

Wysoki udział wydatków na mieszkanie w budżetach polskich gospodarstw domowych na tle innych krajów Europy to konsekwencja niskich dochodów. A pamiętać przy tym musimy, że mamy jeden z najniższych w Europie standardów mieszkaniowych. U nas jest to ok. 24 m^2 powierzchni użytkowej *per capita*, wobec 70 m^2 w Stanach Zjednoczonych i 40—50 m^2 w krajach „dawnej" UE.

Patrząc na relatywnie wysoki udział wydatków na mieszkanie, niewiele niższy niż w krajach znacznie wyprzedzających nas w rankingu mieszkaniowym, musimy pamiętać o znacząco niższym poziomie PKB w Polsce (2—3 razy).

---

dr hab. Mirosław Gorczyca — profesor w *Wyższej Szkole Informatyki i Zarządzania w Rzeszowie*

SUMMARY

The article presents the level of expenditure on housing in households in 2011. The analysis of housing expenditures is conducted by socio-economic households and by the number of people in the household, income status, as well as on a regional basis. The dynamics of changes in the years 2000—2011 is presented, too.

РЕЗЮМЕ

Статья характеризует уровень расходов на жилье в домашних хозяйствах в 2011 г. Анализ расходов на жилье был представлен по социально-экономическим группам домашних хозяйств и по количеству лиц составляющих семью, по уровню доходов, а также в региональном распределении. Была представлена динамика их изменений в 2000—2011 гг.

Sylwia KOSZNIK-BIERNACKA

Analiza wyników matur z języka polskiego w wybranych województwach

Odkąd obowiązuje tzw. nowa matura[1], czyli od 2005 r., i wyniki egzaminów maturalnych są wyrażone w skali procentowej, można dokonać dokładniejszej analizy uzyskiwanych wyników przez abiturientów. Poprzednia sześciostopniowa skala oceny matur uniemożliwiała np. badania postaci rozkładu uzyskiwanych wyników, czego podjęto się właśnie w obecnym opracowaniu.

Na podstawie danych udostępnionych przez Okręgową Komisję Egzaminacyjną w Gdańsku dokonano analizy wyników pisemnych matur z języka polskiego na poziomie podstawowym w województwach pomorskim i kujawsko-pomorskim w 2011 r. Liczba osób zdających maturę wynosiła 38493. Wielkość tej próby podzielono według następujących kategorii, tj. na uzyskane wyniki:

- przez maturzystów ogółem,
- w poszczególnych województwach,
- z uwzględnieniem szkół publicznych i niepublicznych[2],
- z uwzględnieniem typu szkoły[3].

W artykule podano wartości podstawowych parametrów charakteryzujących otrzymane wyniki. Na podstawie tak licznej zbiorowości podjęto próbę dopasowania do danych empirycznych odpowiedniego rozkładu prawdopodobieństwa, opisującego wyniki matur. Z literatury (Jóźwiak, Podgórski, 2001; Plucińska, Pluciński, 2000) wiadomo, że jeżeli na dane zjawisko oddziałuje duża liczba niezależnych czynników, których wpływ (traktowany odrębnie) jest mało znaczący, to takie zmienne mają często rozkład zbliżony do normalnego, np.: waga oraz wzrost osobników jednorodnych populacji ludzkich lub zwierzęcych, losowe błędy pomiarów czy cechy elementów pochodzących z masowej produkcji. Zatem w opracowaniu postawiono hipotezę, że wyniki pisemnych matur z języ-

[1] Rozporządzenie Ministra Edukacji Narodowej z 30 kwietnia 2007 r. w sprawie warunków i sposobu oceniania, klasyfikowania i promowania uczniów i słuchaczy oraz przeprowadzania sprawdzianów i egzaminów w szkołach publicznych (Dz. U. z 11.05.2007 r., Nr 83, poz. 562) oraz rozporządzenie Ministra Edukacji Narodowej z 17 listopada 2010 r. zmieniające rozporządzenie w sprawie warunków i sposobu oceniania, klasyfikowania i promowania uczniów i słuchaczy oraz przeprowadzania sprawdzianów i egzaminów w szkołach publicznych (Dz. U. z 2.12.2010 r., Nr 228, poz. 1491).

[2] Szkoły niepubliczne z uprawnieniami publicznymi.

[3] Technikum, liceum ogólnokształcące, liceum profilowane, liceum uzupełniające, technikum uzupełniające.

ka polskiego na poziomie podstawowym można opisać rozkładem normalnym. W celu weryfikacji postawionej hipotezy przeprowadzono stosowne analizy i odpowiednie testy zgodności. Ostatecznie okazało się, że nie ma podstaw do odrzucenia postawionej hipotezy dla wszystkich rozpatrywanych kategorii danych. Przeprowadzono także szczegółową analizę wariancji (Jóźwiak, Podgórski, 2001; Stanisz, 2007), aby porównać średnie wyniki osiągnięte przez maturzystów w zakresie poszczególnych kategorii.

WYNIKI MATUR OGÓŁEM Z JĘZYKA POLSKIEGO W WOJEWÓDZTWACH POMORSKIM I KUJAWSKO-POMORSKIM

Wyznaczono wartości podstawowych parametrów statystycznych.

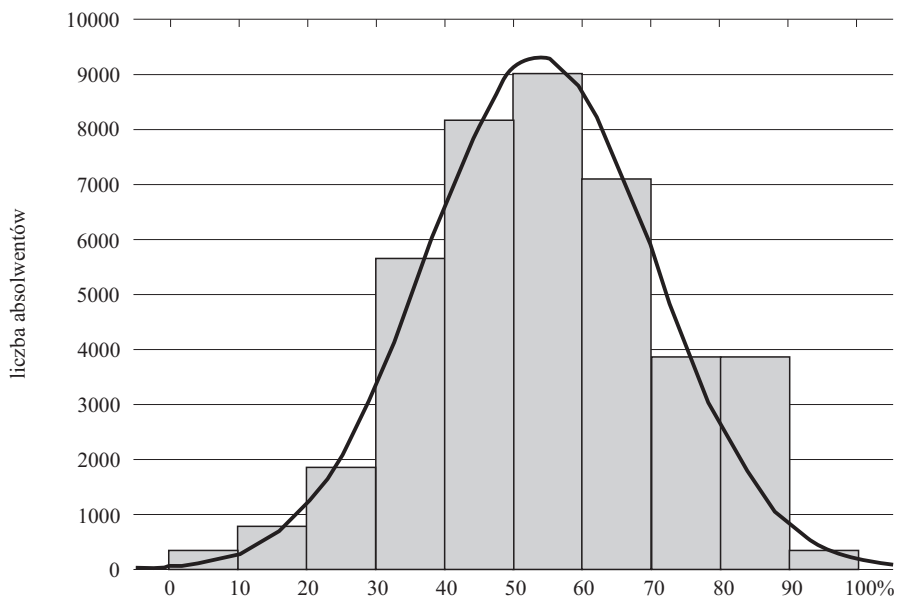
ZESTAWIENIE PODSTAWOWYCH WARTOŚCI PARAMETRÓW STATYSTYCZNYCH WYNIKÓW MATUR OGÓŁEM Z JĘZYKA POLSKIEGO W BADANYCH WOJEWÓDZTWACH

N ważnych	Średnia	Mediana	Moda	Wariancja	Odchylenie standardowe	Skośność	Kurtoza
38493	53,38	53,00	54,00	271,95	16,49	–0,10	0,02

Ź r ó d ł o: opracowanie własne.

Otrzymane wyniki matury z języka polskiego przedstawiono też na histogramie z dopasowanym wykresem rozkładu normalnego.

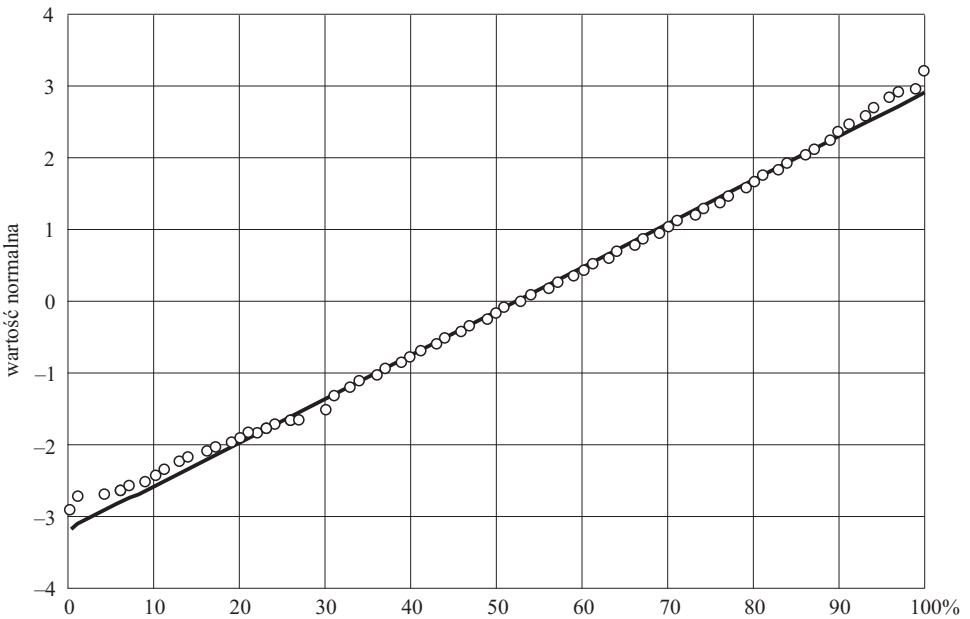
Wykr. 1. HISTOGRAM WYNIKÓW MATUR Z JĘZYKA POLSKIEGO OGÓŁEM Z DOPASOWANYM WYKRESEM ROZKŁADU NORMALNEGO



Ź r ó d ł o: opracowanie własne.

Sporządzono również wykres normalności.

Wykr. 2. WYKRES NORMALNOŚCI WYNIKÓW MATUR Z JĘZYKA POLSKIEGO OGÓŁEM



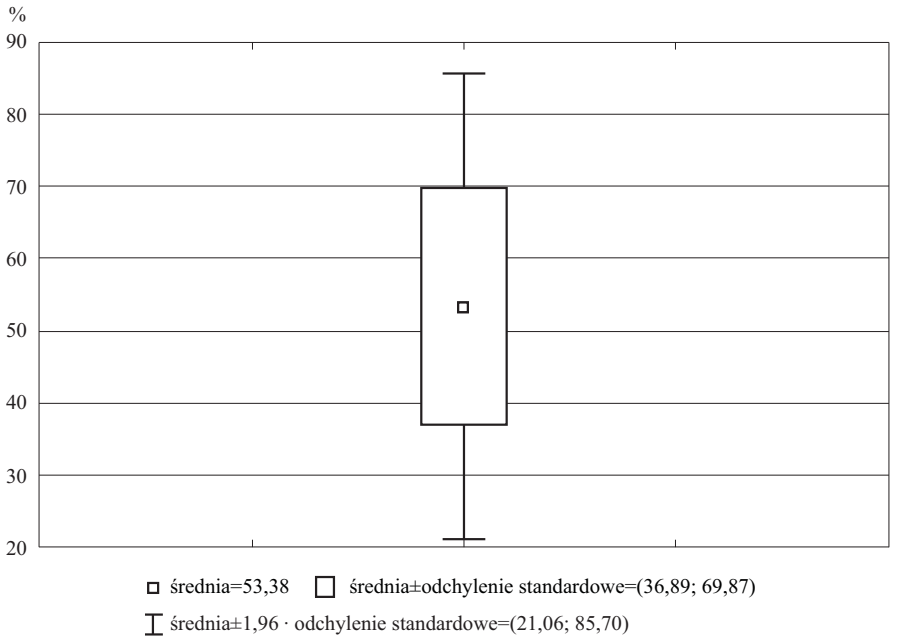
Ź r ó d ł o: jak przy wykr. 1.

Na podstawie wykr. 1 i 2 oraz obliczonych wartości skośności i kurtozy bliskich zeru (zestawienie) można postawić hipotezę, że rozkład normalny *N*(53,38; 16,49) będzie dobrze opisywał badane zjawisko (Drapella, 2004).

Ponieważ wybrana zbiorowość jest dość liczna, a dane mają wiele wariantów, wydaje się, że test zgodności Mizesa-Smirnowa jest najodpowiedniejszy do zweryfikowania postawionej hipotezy. Obliczona wartość statystyki testowej w tym teście wynosi 0,006697. Porównując ją na poziomie istotności 0,05 z wartościami krytycznymi testu Mizesa-Smirnowa, którego implementację komputerową można znaleźć w pracy A. Drapelli (2000) okazuje się, że nie ma podstaw do odrzucenia postawionej hipotezy, czyli wyniki matur z języka polskiego na poziomie podstawowym mogą być opisane rozkładem normalnym *N*(53,38; 16,49).

Drugim aspektem opracowania było badanie średniej w poszczególnych, wyodrębnionych kategoriach wyników matur. Dla danych ogółem została ona przedstawiona na wykr. 3 wraz z uwzględnieniem typowej zmienności, o której można mówić w przypadku rozkładów symetrycznych zbliżonych do rozkładu normalnego.

Wykr. 3. WARTOŚĆ ŚREDNIA WYNIKÓW MATURALNYCH Z JĘZYKA POLSKIEGO WOJEWÓDZTWA KUJAWSKO-POMORSKIE I POMORSKIE OGÓŁEM



Ź r ó d ł o: jak przy wykr. 1.

Jak prezentuje wykr. 3, ok. 66% maturzystów zdało maturę z języka polskiego na poziomie 37—70%. Niestety więcej niż 5% zdających nie uzyskało wymaganego progu 30%.

WYNIKI MATUR Z JĘZYKA POLSKIEGO W BADANYCH WOJEWÓDZTWACH

Obliczono podstawowe parametry statystyczne badanych danych.

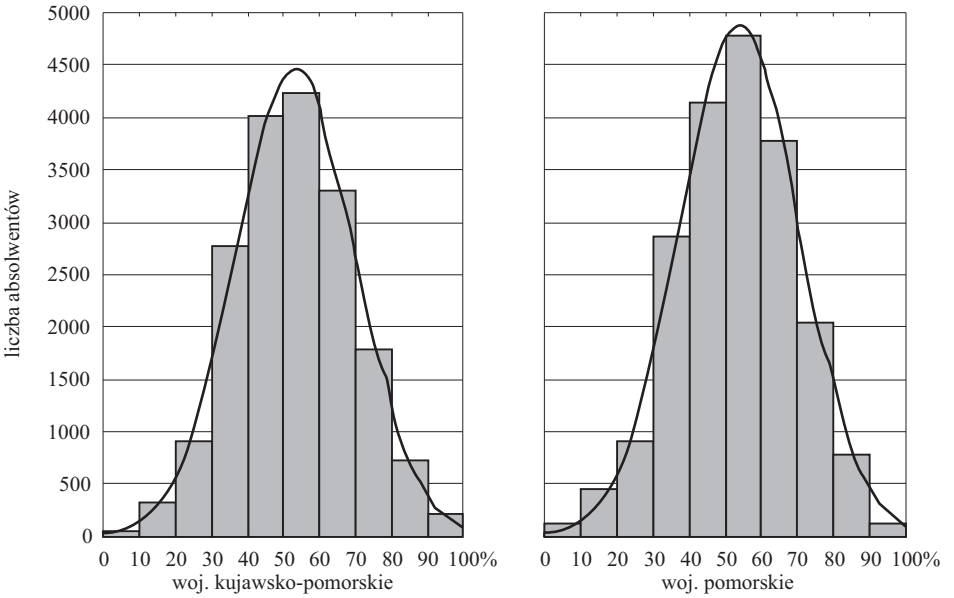
TABL. 1. PODSTAWOWE WARTOŚCI PARAMETRÓW STATYSTYCZNYCH WYNIKÓW MATUR Z JĘZYKA POLSKIEGO WEDŁUG WOJEWÓDZTW

Województwa	*N* ważnych	Średnia	Mediana	Moda	Wariancja	Odchylenie standardowe	Skośność	Kurtoza
Kujawsko-pomorskie	18434	53,34057	53	50	272,5063	16,50776	–0,014934	0,018193
Pomorskie	20059	53,41996	54	54	271,4421	16,47550	–0,174153	0,024298

Ź r ó d ł o: opracowanie własne.

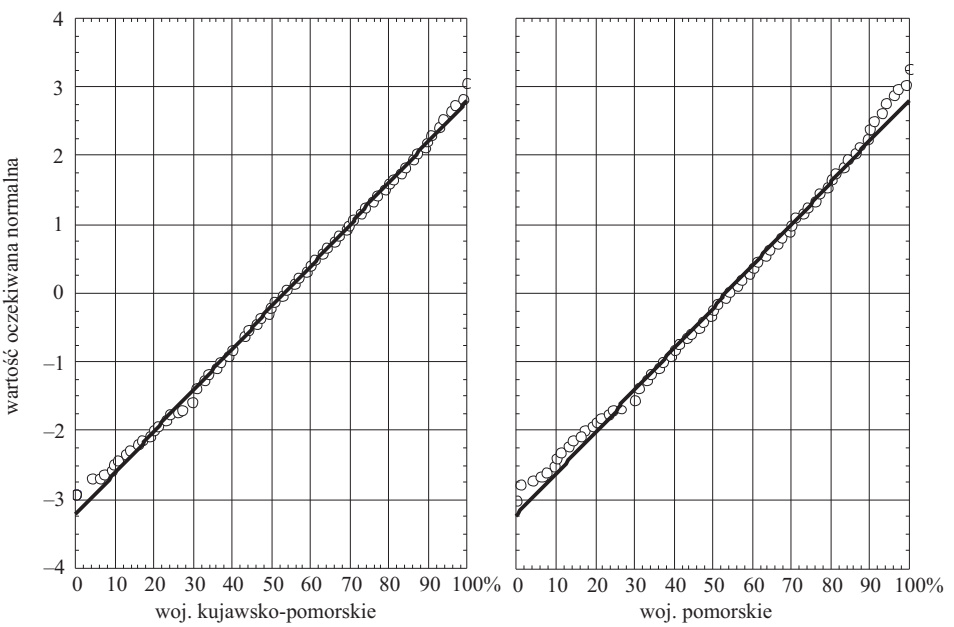
Wyniki egzaminu z języka polskiego uzyskane przez maturzystów w województwach pomorskim i kujawsko-pomorskim przedstawiono też na histogramie z dopasowanym wykresem rozkładu normalnego.

Wykr. 4. HISTOGRAM WYNIKÓW MATUR Z JĘZYKA POLSKIEGO WEDŁUG WOJEWÓDZTW Z DOPASOWANYM WYKRESEM ROZKŁADU NORMALNEGO



Ź r ó d ł o: jak przy wykr. 1.

Wykr. 5. WYKRES NORMALNOŚCI WYNIKÓW MATUR Z JĘZYKA POLSKIEGO WEDŁUG WOJEWÓDZTW

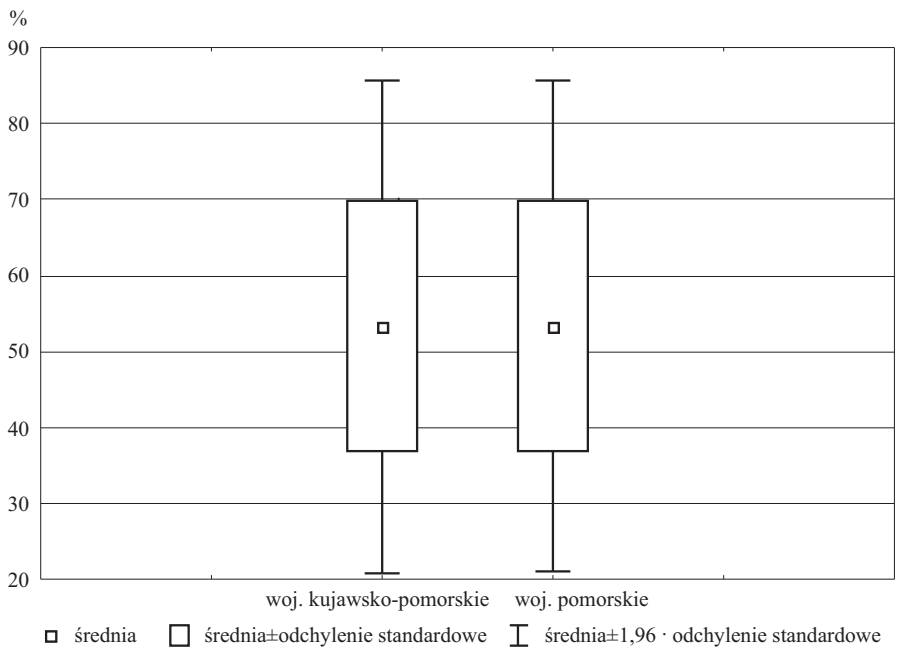


Ź r ó d ł o: jak przy wykr. 1.

Na podstawie wykresów 4 i 5 oraz bliskich zeru wartości skośności i kurtozy (tabl. 1) postawiono hipotezę, że wyniki matur z języka polskiego na poziomie podstawowym w województwach kujawsko-pomorskim i pomorskim można opisać rozkładem normalnym odpowiednio z parametrami $N(53{,}34;\ 16{,}51)$ i $N(53{,}42;\ 16{,}48)$. Hipotezę tę zweryfikowano przeprowadzając dla obu przypadków test zgodności Mizesa-Smirnowa. Otrzymane wartości statystyki testowej to odpowiednio 0,009272 i 0,005536. Porównując je z wartościami krytycznymi w obu przypadkach nie ma podstaw do odrzucenia hipotezy zerowej, że wyniki matur można opisać rozkładem normalnym odpowiednio $N(53{,}34;\ 16{,}51)$ i $N(53{,}42;\ 16{,}48)$.

Rozważono także wartości średnie wyników maturalnych z języka polskiego w badanych województwach, które porównano na wykr. 6.

Wykr. 6. ŚREDNIE WYNIKI MATUR Z JĘZYKA POLSKIEGO WEDŁUG WOJEWÓDZTW Z UWZGLĘDNIENIEM TYPOWEJ ZMIENNOŚCI



Ź r ó d ł o: jak przy wykr. 1.

Na podstawie wykr. 6 można sądzić, że otrzymane wyniki średnie w badanych województwach są zbliżone. Aby sprawdzić, czy rzeczywiście tak jest przeprowadzono analizę wariancji. Można tego dokonać, bo jak wykazano wcześniej, wyniki z obu województw mogą być opisane rozkładem normalnym, więc podstawowe założenie klasycznej analizy wariancji jest spełnione. Zweryfikowano hipotezę, że średnie wyniki matur z języka polskiego na poziomie

podstawowym w obu badanych województwach są równe, tzn. nie zależą od województwa. Za pomocą modułu Excela *Analiza danych* przeprowadzono stosowne obliczenia, których wyniki przedstawiono w tabl. 2.

TABL. 2. WYNIKI ANALIZY WARIANCJI WEDŁUG WOJEWÓDZTW

A. PODSUMOWANIE WYNIKÓW

Województwa	Licznik	Suma	Średnia	Wariancja
Kujawsko-pomorskie	18434	983280	53,34057	272,506262
Pomorskie	20059	1071551	53,41996	271,442130

B. ANALIZA WARIANCJI

Źródła wariancji	*SS*	*df*	*MS*	*F*	Wartość *p*	Test *F*
Pomiędzy województwami	60,55233347	1	60,55233	0,22265838	0,637024461	3,841700205
W obrębie województw	10467694,17	38491	271,9517			
R a z e m	**10467754,72**	**38492**				

U w a g a. *SS* — suma kwadratów, *df* — stopnie swobody, *MS* — średnia kwadratów, *F* — empiryczna wartość statystyki testowej *F*, Wartość *p* — wartość progowa testu *F*, test *F* — wartość krytyczna testu *F*.

Ź r ó d ł o: jak przy tabl. 1.

Ponieważ wartość statystyki *F* (tabl. 2) jest znacznie mniejsza od statystyki testowej, czyli na poziomie istotności 5%, to nie ma podstaw do odrzucenia postawionej hipotezy zerowej. Zatem średnie wyniki uzyskane przez maturzystów w obu województwach nie różnią się statystycznie.

Przeprowadźmy jeszcze test *z* (Mynarski, 2003, 2006) w zakresie różnicy między średnimi wynikami matur z języka polskiego na poziomie podstawowym w badanych województwach. Przeprowadzono stosowne obliczenia, a otrzymane wyniki podano w tabl. 3.

TABL. 3. TEST *Z* Z DWIEMA PRÓBAMI ŚREDNICH WEDŁUG WOJEWÓDZTW

Test *z*	Kujawsko-pomorskie	Pomorskie
Średnia	53,34056634	53,4199611
Znana wariancja	272,5	271,44
Obserwacje	18434	20059
Różnica średnich według hipotezy	0	
z	–0,471831663	
Test *z* dwustronny	1,959963985	

Ź r ó d ł o: jak przy tabl. 1.

Uzyskana wartość statystyki *z* jest mniejsza od krytycznej, zatem i tutaj nie ma podstaw do odrzucenia hipotezy zerowej o równości średnich. Ten test potwierdza, że średnie wyniki matur w obu województwach nie różnią się statystycznie.

WYNIKI MATUR Z JĘZYKA POLSKIEGO Z UWZGLĘDNIENIEM SZKÓŁ PUBLICZNYCH I NIEPUBLICZNYCH

Podstawowe parametry statystyczne danych dotyczących szkół publicznych i niepublicznych przedstawiono w tabl. 4.

TABL. 4. PODSTAWOWE WARTOŚCI PARAMETRÓW STATYSTYCZNYCH WYNIKÓW MATUR Z JĘZYKA POLSKIEGO WEDŁUG SZKÓŁ PUBLICZNYCH I NIEPUBLICZNYCH

Szkoły	*N* ważnych	Średnia	Mediana	Moda	Wariancja	Odchylenie standardowe	Skośność	Kurtoza
Publiczne	35404	54,25782	54	54	257,7918	16,05590	–0,054065	–0,036167
Niepubliczne	3089	43,34315	43	30	324,7054	18,01958	0,020942	0,038122

Ź r ó d ł o: jak przy tabl. 1.

Wyniki matury z języka polskiego z uwzględnieniem rodzaju szkoły przedstawiono również na histogramie z dopasowanym wykresem rozkładu normalnego.

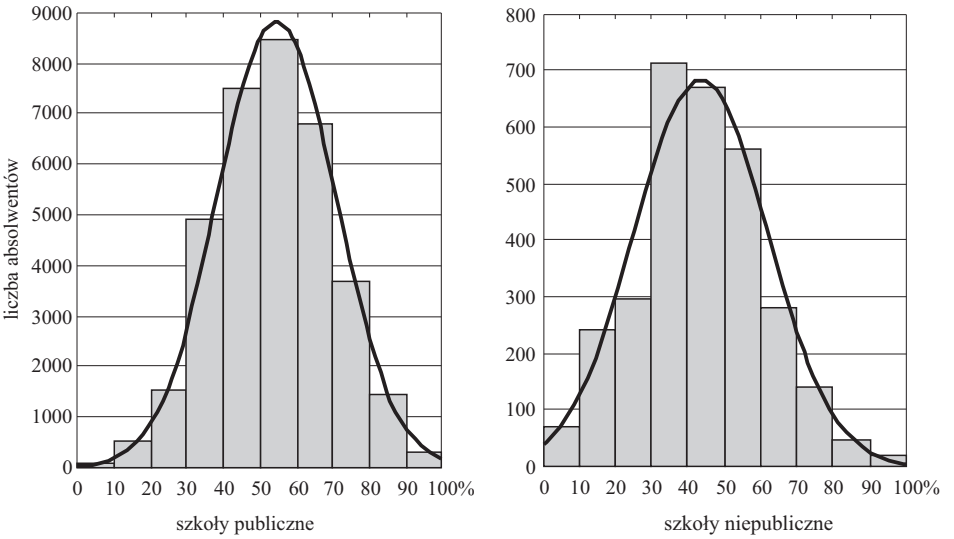
Również w tej kategorii wyników matur (wykr. 7 i 8) zastosowanie krzywej normalnej powinno dawać dobrą zgodność z rozkładem empirycznym w zakresie poszczególnych rodzajów szkół. Sprawdzono zatem formalnie stosując test zgodności Mizesa-Smirnowa, czy wyniki matur na poziomie podstawowym w szkołach publicznych i niepublicznych można opisać rozkładem normalnym odpowiednio z parametrami $N(54{,}26;\ 16{,}06)$ i $N(43{,}34;\ 18{,}02)$. Obliczone wartości statystyki testowej to odpowiednio 0,007786 i 0,011775. Tak więc w obu przypadkach wartości statystyki testowej są zdecydowanie mniejsze niż wartości krytyczne w teście zgodności Mizesa-Smirnowa, nie ma więc podstaw do odrzucenia postawionej hipotezy.

Dane empiryczne można opisać rozkładem normalnym, czyli spełnione jest podstawowe założenie analizy wariancji. Przeprowadzono zatem badanie wartości średniej wyników uzyskanych w szkołach publicznych i niepublicznych, co przedstawiono wraz z typową zmiennością na wykr. 9.

Z wykr. 9 można odczytać, że w szkołach publicznych ponad 66% abiturientów zdało egzamin maturalny na poziomie dostatecznym i dobrym. W szkołach niepublicznych było nieco gorzej. Typowy obszar zmienności cechy był tam poniżej progowych 30%, czyli odsetek niezdanych matur z języka polskiego w szkołach niepublicznych był większy niż w szkołach publicznych.

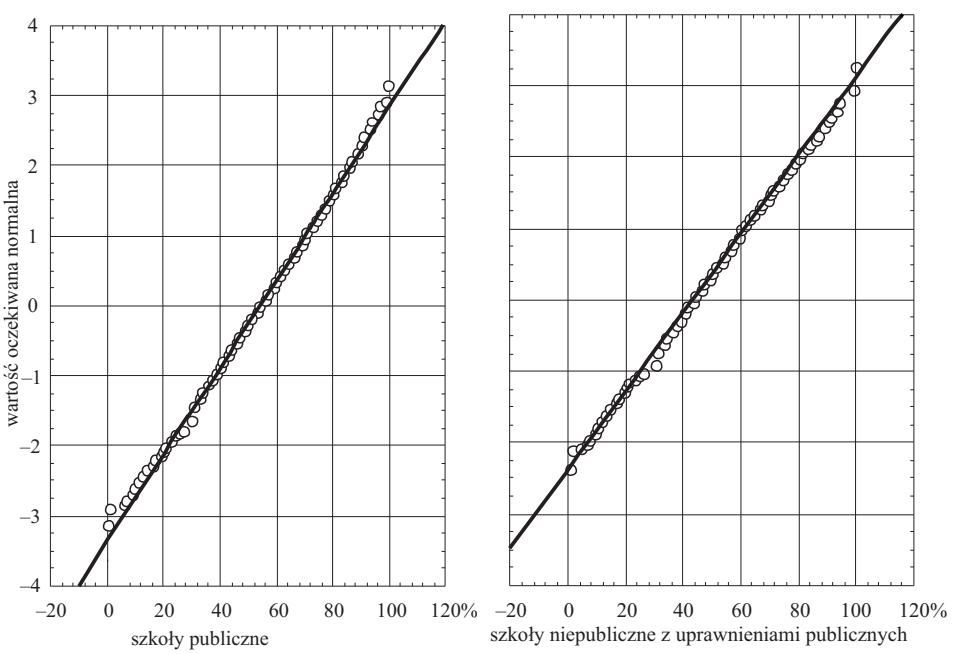
Przeprowadzając klasyczną analizę wariancji wyników uzyskanych przez uczniów szkół publicznych i niepublicznych, otrzymano rezultaty, które podano w tabl. 5.

Wykr. 7. HISTOGRAM WYNIKÓW MATUR Z JĘZYKA POLSKIEGO Z UWZGLĘDNIENIEM SZKÓŁ PUBLICZNYCH I NIEPUBLICZNYCH Z DOPASOWANYM WYKRESEM ROZKŁADU NORMALNEGO



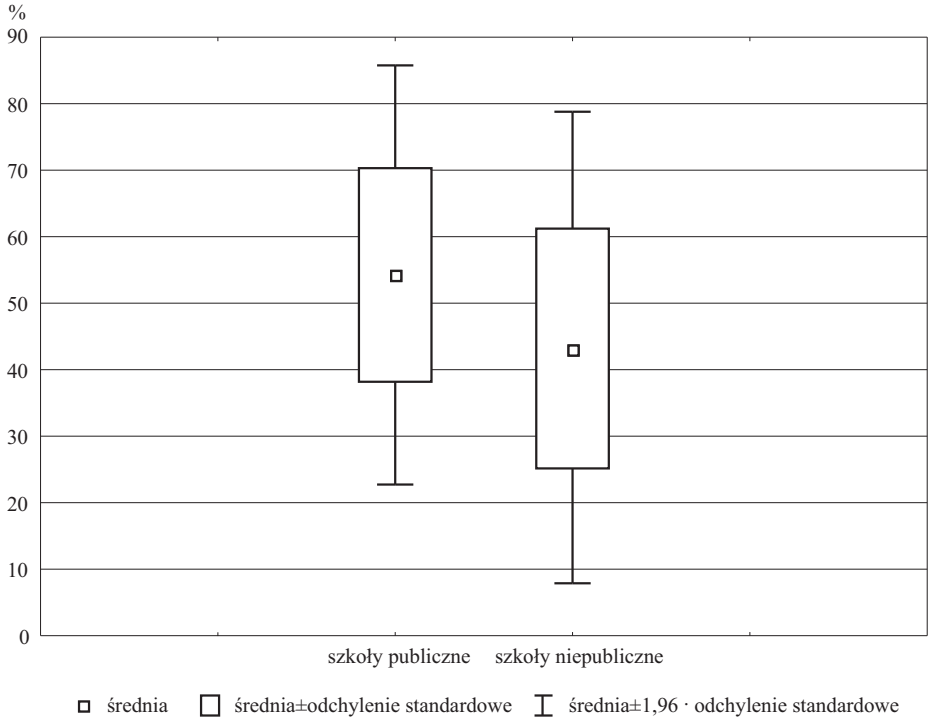
Ź r ó d ł o: jak przy wykr. 1.

Wykr. 8. WYKRES NORMALNOŚCI WYNIKÓW MATUR Z JĘZYKA POLSKIEGO Z UWZGLĘDNIENIEM SZKÓŁ PUBLICZNYCH I NIEPUBLICZNYCH



Ź r ó d ł o: jak przy wykr. 1.

Wykr. 9. ŚREDNIE WYNIKI MATUR Z JĘZYKA POLSKIEGO I ICH TYPOWEJ ZMIENNOŚCI Z UWZGLĘDNIENIEM SZKÓŁ PUBLICZNYCH I NIEPUBLICZNYCH



Ź r ó d ł o: jak przy wykr. 1.

TABL. 5. WYNIKI ANALIZY WARIANCJI Z UWZGLĘDNIENIEM SZKÓŁ PUBLICZNYCH I NIEPUBLICZNYCH

A .PODSUMOWANIE WYNIKÓW

Szkoły	Licznik	Suma	Średnia	Wariancja
Publiczne	35404	1920944	54,25782	257,7918
Niepubliczne	3089	133887	43,34315	324,7054

B. ANALIZA WARIANCJI

Źródła wariancji	*SS*	*df*	*MS*	*F*	Wartość *p*	Test *F*
Pomiędzy szkołami publicznymi i niepublicznymi	338461,9	1	338461,9	1286,145	0	3,8417
W obrębie szkół publicznych i niepublicznych	10129293	38491	263,16			
R a z e m	**10467755**	**38492**				

U w a g a. Jak przy tabl. 2.
Ź r ó d ł o: jak przy tabl. 1.

Otrzymana wartość statystyki testowej *F* jest bardzo duża, można więc powiedzieć, że rodzaj szkoły zdecydowanie wpływa na średnie wyniki maturalne uzyskiwane z języka polskiego. Ze statystycznego punktu widzenia wyniki średnie osiągane przez maturzystów zdających egzamin z języka polskiego w szkołach publicznych i niepublicznych są różne.

Potwierdzono ten fakt przeprowadzając test *z* porównujący uzyskane wyniki średnie w obu badanych grupach. Wyniki obliczeń podano w tabl. 6.

TABL. 6. TEST *Z* Z DWIEMA PRÓBAMI DLA ŚREDNICH W SZKOŁACH PUBLICZNYCH I NIEPUBLICZNYCH

Test *z*	Publiczne	Niepubliczne
Średnia	54,257824	43,34315312
Znana wariancja	257,7918	324,7054
Obserwacje	35404	3089
Różnica średnich według hipotezy	0	
z	32,556008	
Test *z* dwustronny	1,959964	

Ź r ó d ł o: jak przy tabl. 1.

Wartość statystyki *z* jest bardzo duża w porównaniu z wartością krytyczną tego testu. Potwierdza to wcześniejszy wniosek, że wyników średnich uzyskanych w obu rozważanych grupach nie można uznać za porównywalne statystycznie.

WYNIKI MATUR Z JĘZYKA POLSKIEGO Z UWZGLĘDNIENIEM TYPU SZKOŁY

Najpierw obliczono podstawowe parametry statystyczne dla tej kategorii danych.

TABL. 7. PODSTAWOWE WARTOŚCI PARAMETRÓW STATYSTYCZNYCH WYNIKÓW MATUR Z JĘZYKA POLSKIEGO Z UWZGLĘDNIENIEM TYPU SZKOŁY

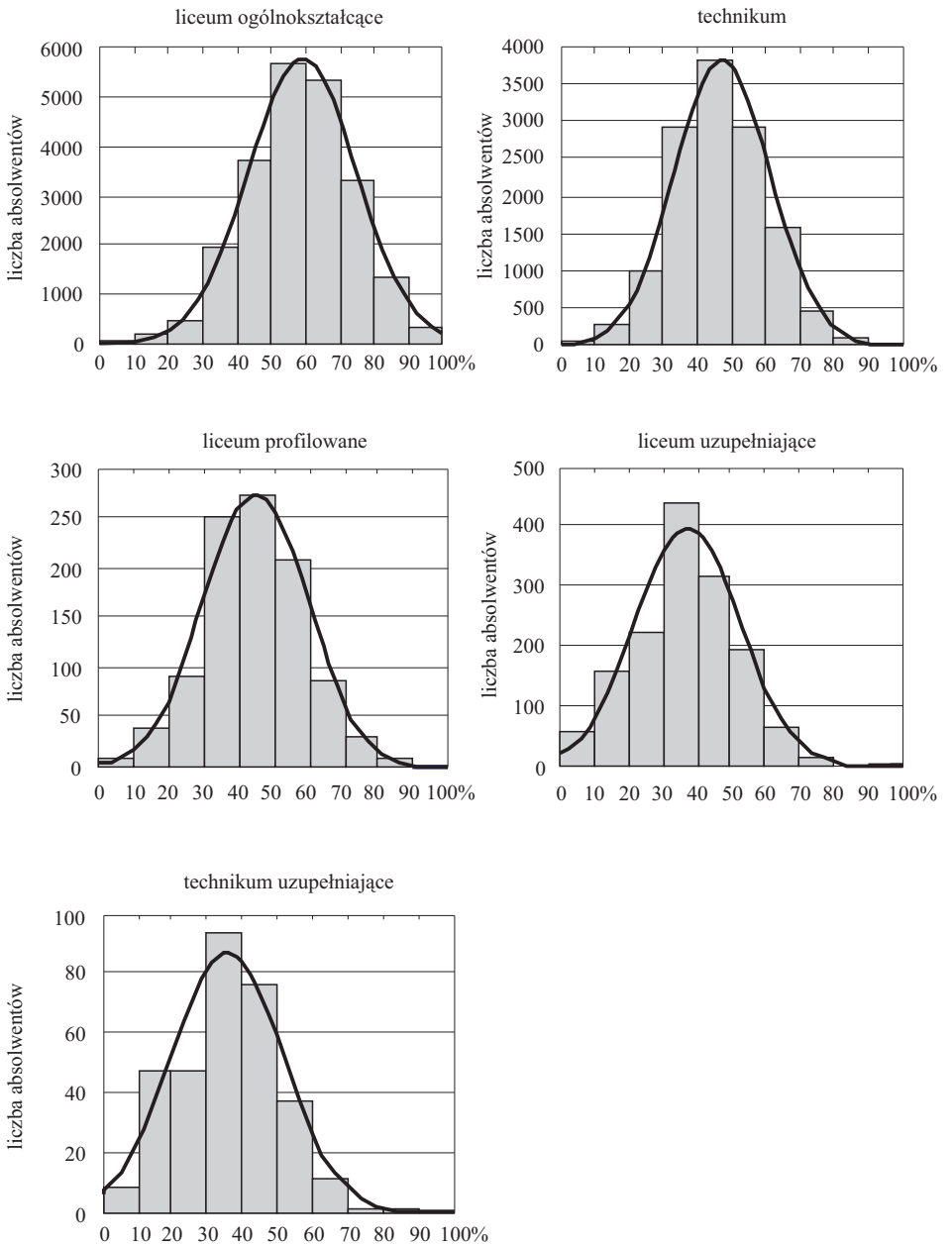
Typ szkoły	*N*	Średnia	Mediana	Moda	Wariancja	Odchylenie standardowe	Skośność	Kurtoza
Liceum ogólnokształcące	22482	58,82208	59	59	246,0199	15,68502	–0,260278	0,275152
profilowane	1008	44,49603	44	30	219,6802	14,82161	–0,073727	0,435019
uzupełniajace	1495	36,96187	37	30	228,8533	15,12790	–0,031856	–0,160606
Technikum	13174	47,09579	47	30	191,1237	13,82475	–0,018072	0,251815
uzupełniające	334	35,45808	36	30	241,0598	15,52610	–0,061662	0,052492

Ź r ó d ł o: jak przy tabl. 1.

Wyniki matury z języka polskiego z uwzględnieniem typu szkoły przedstawiono też na histogramie z dopasowanym wykresem rozkładu normalnego.

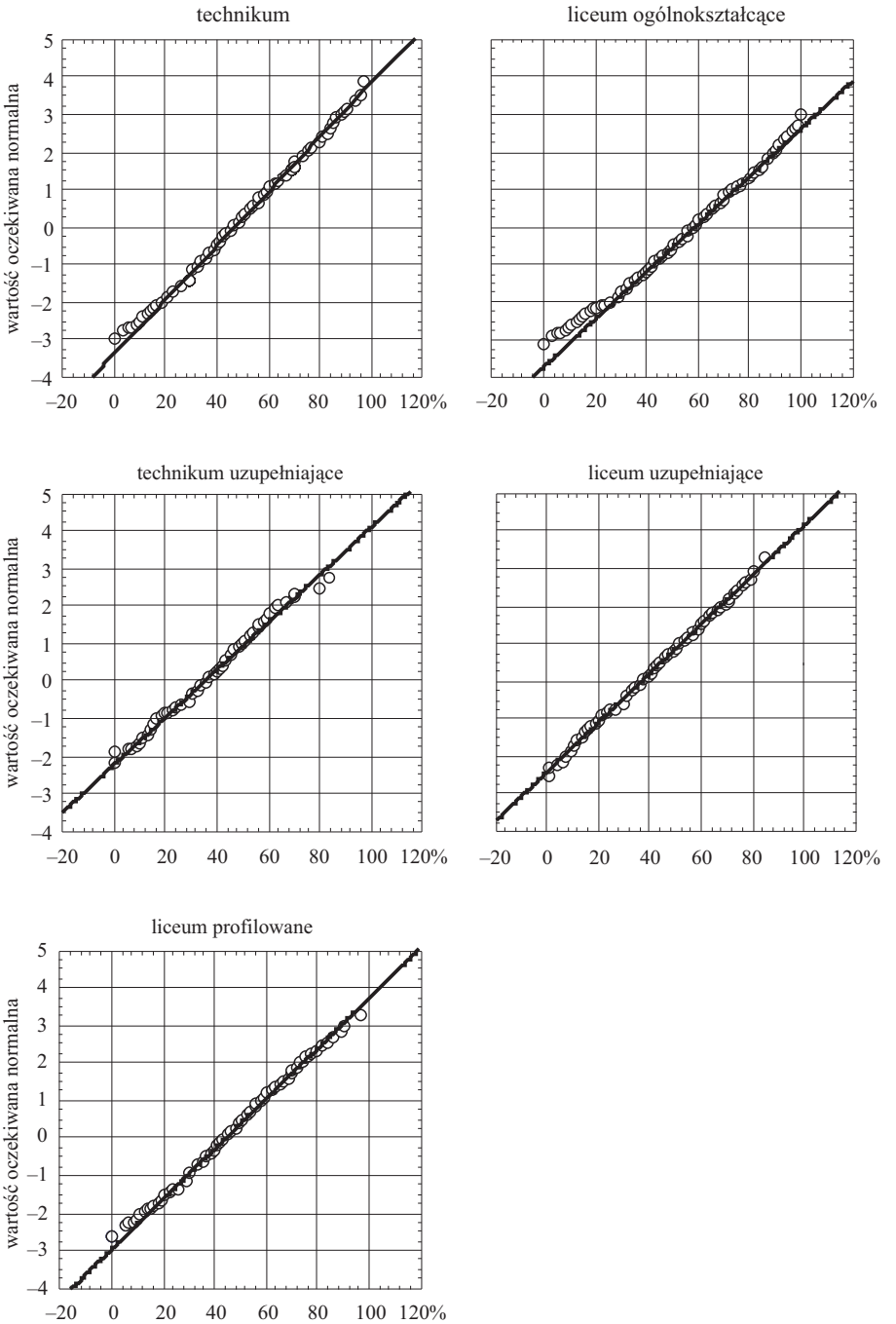
Oczekiwaną normalną w zakresie poszczególnych typów szkół pokazano również na wykr. 11.

Wykr. 10. HISTOGRAM WYNIKÓW MATUR Z JĘZYKA POLSKIEGO WEDŁUG TYPU SZKOŁY Z DOPASOWANYM WYKRESEM ROZKŁADU NORMALNEGO



Ź r ó d ł o: jak przy wykr. 1.

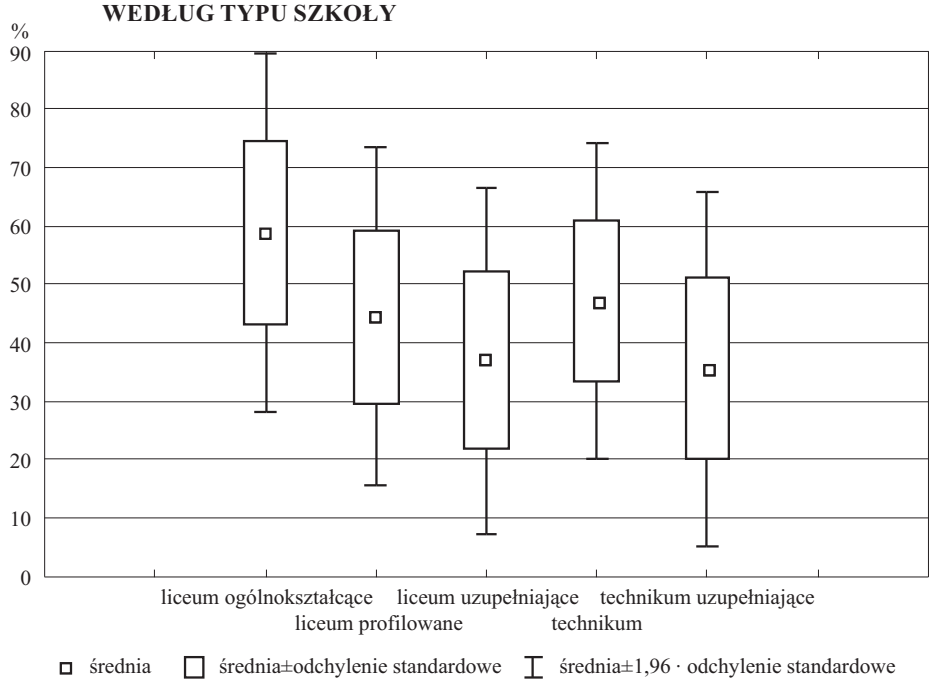
Wykr. 11. WYKRES NORMALNOŚCI WYNIKÓW MATUR Z JĘZYKA POLSKIEGO WEDŁUG TYPU SZKOŁY



Ź r ó d ł o: jak przy wykr. 1.

Na podstawie danych zilustrowanych na wykr. 10 i 11 postawiono hipotezę, że wyniki matur z języka polskiego na poziomie podstawowym w poszczególnych typach szkół można również opisać rozkładem normalnym z następującymi parametrami dla: liceum ogólnokształcącego *N*(58,82208; 15,68502), liceum profilowanego *N*(44,49603; 14,82161), liceum uzupełniającego *N*(36,96187; 15,12790), technikum *N*(47,09579; 13,82475), technikum uzupełniającego *N*(35,45808; 15,52610). Zweryfikowano tę hipotezę przeprowadzając test zgodności Mizesa-Smirnowa, otrzymane wartości statystyki testowej to odpowiednio 0,005423243, 0,012769; 0,012935; 0,011917; 0,019955. W przypadku wszystkich typów szkół porównując otrzymane wartości z wartościami krytycznymi testu Mizesa-Smirnowa nie ma podstaw do odrzucenia hipotezy zerowej. Wyniki matur z języka polskiego na poziomie podstawowym można opisać rozkładem normalnym z podanymi wcześniej parametrami. Ponieważ spełnione jest podstawowe założenie analizy wariancji, to zbadano też wartości średnie uzyskane w poszczególnych typach szkół. Otrzymane wyniki wraz z typową zmiennością podano na wykr. 12.

Wykr. 12. ŚREDNIE WYNIKI MATUR Z JĘZYKA POLSKIEGO WEDŁUG TYPU SZKOŁY



Ź r ó d ł o: jak przy wykr. 1.

Przeprowadzając klasyczną analizę wariancji wyników uzyskanych w poszczególnych typach szkół otrzymano rezultaty przedstawione w tabl. 8.

TABL. 8. WYNIKI PRZEPROWADZONEJ ANALIZY WARIANCJI WEDŁUG TYPU SZKOŁY

A. PODSUMOWANIE WYNIKÓW

Typ szkoły	Licznik	Suma	Średnia	Wariancja
Liceum ogólnokształcące	22482	1322438	58,82208	246,0199
profilowane	1008	44852	44,49603	219,6802
uzupełniające	1495	55258	36,96187	228,8533
Technikum	13174	620440	47,09579	191,1237
uzupełniające	334	11843	35,45808	241,0598

B. ANALIZA WARIANCJI

Źródła wariancji	*SS*	*df*	*MS*	*F*	Wartość *p*	Test *F*
Pomiędzy typami szkół	1775910	4	443977,4	1965,958	0	2,372163
W obrębie typu szkoły ...	8691845	38488	225,8326			
R a z e m	**10467755**	**38492**				

U w a g a. Jak przy tabl. 2.
Ź r ó d ł o: jak przy tabl. 1.

Ponieważ wartość statystyki testowej *F* jest bardzo duża można stwierdzić, że typ szkoły zdecydowanie wpływa na uzyskiwane wyniki maturalne z języka polskiego.

W przypadku tej kategorii danych nie przeprowadzono testu *z* równości średnich, bo dotyczy on porównywania średnich tylko w dwóch populacjach danych.

Wnioski

Na podstawie zbiorowości liczącej 38493 uczniów zdających maturę z języka polskiego na poziomie podstawowym stwierdzono, że uzyskane wyniki można opisać rozkładem normalnym bez względu na analizowane w opracowaniu województwo, szkoły publiczne i niepubliczne, jak i bez względu na typ szkoły.

Z przeprowadzonych analiz wynika też, że średnie wyniki maturalne z języka polskiego na poziomie podstawowym nie zależą od województwa, z którego pochodzili zdający, ale zdecydowanie zależą od typu szkoły oraz tego, czy była to szkoła publiczna czy niepubliczna.

Trzeba jednak dodać, że „zdawalność" matur zależy też w dużej mierze od realizowanego przez szkołę poziomu programu kształcenia. Może być on realizowany na poziomie rozszerzonym i jest to często praktykowane w liceach ogólnokształcących, co odzwierciedlają uzyskiwane nieco lepsze wyniki maturalne. Z kolei w pozostałych typach szkół program kształcenia jest realizowany zazwyczaj na poziomie podstawowym, a zatem może być to przyczyną uzyskiwania nieco gorszych wyników maturalnych z języka polskiego przez abiturientów techników i liceów profilowanych.

dr Sylwia Kosznik-Biernacka — *Akademia Pomorska w Słupsku*

LITERATURA

Drapella A. (2000), *Inżynieria niezawodnościowa*, Warszawa

Drapella A. (2004), *Wnioskowanie statystyczne na podstawie skośności i kurtozy*, WPAP, Słupsk

Jóźwiak J., Podgórski J. (2001), *Statystyka od podstaw*, PWE, Warszawa

Mynarski S. (2003), *Analiza danych rynkowych i marketingowych z wykorzystaniem programu STATISTICA*, WAE, Kraków

Mynarski S. (2006), *Analiza danych rynkowych i marketingowych z wykorzystaniem programu Excel*, WAE, Kraków

Plucińska A., Pluciński E. (2000), *Rachunek prawdopodobieństwa. Statystyka matematyczna. Procesy stochastyczne*, WNT, Warszawa

Stanisz A. (2007), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, „Modele liniowe i nieliniowe", t 2., StatSoft, Kraków

SUMMARY

The paper presents an analysis of the matriculation exam results of the Polish language at a basic level in Pomorskie and Kujawsko-Pomorskie voivodships in 2011. Values of the basic parameters characterizing the results are presented as well as a detailed analysis of variance between the different categories of data was made. The Author hypothesized that the matriculation exam results of the written Polish at a basic level can be described by a normal distribution. In order to verify the appropriate analysis was carried out and appropriate compliance testing. Ultimately, it turned out that there is no reason to reject the hypothesis for all categories of data concerned.

РЕЗЮМЕ

В статье был представлен анализ результатов экзаменов на аттестат зрелости по польскому языку на основном уровне в поморском и куявско-поморском воеводствах в 2011 г. Были указаны значения основных параметров характеризующих полученные результаты. Был также проведен подробный анализ дисперсии по отдельным категориям данных. Была представленная гипотеза, что результаты письменного варианта экзамена на аттестат зрелости по польскому языку на основном уровне можно охарактеризовать нормальным распределением. Для ее проверки были проведены соответствующие анализы и тесты на соответствие. В конечном итоге показалось, что нет никаких оснований отвергнуть гипотезу для всех рассматриваемых категорий данных.

Z PRAC RADY STATYSTYKI

Antoni ŻURAWICZ

Działalność Rady Statystyki w II półroczu 2012 r.

Z dniem 1 lipca 2012 r. prezes Rady Ministrów Donald Tusk powołał, na mocy ustawy o statystyce publicznej, Radę Statystyki V kadencji. W jej skład weszli:

- reprezentujący organy administracji rządowej: Grażyna Henclewska (Ministerstwo Gospodarki), Stanisław Sudak (Ministerstwo Rozwoju Regionalnego), Małgorzata Steiner (Ministerstwo Administracji i Cyfryzacji), Sławomir Dudek (Ministerstwo Finansów), Zbigniew Rykowski (Kancelaria Prezesa Rady Ministrów);
- reprezentujący organy administracji samorządowej — Andrzej Porawski;
- przedstawiciel NBP — prof. dr hab. Eugeniusz Gatnar (członek Zarządu NBP);
- reprezentujący organizacje społeczne: prof. dr hab. Joanna Kotowicz-Jawor, Marek Kłoczko i prof. dr Franciszek Kubiczek;
- przedstawiciele organizacji pracodawców reprezentatywnych w rozumieniu ustawy z 6 lipca 2001 r.: prof. dr hab. Zbigniew Strzelecki (Pracodawcy Rzeczypospolitej Polskiej), dr hab. Stanisław Kubielas (Polska Konfederacja Pracodawców Prywatnych LEWIATAN), dr Włodzimierz Biniek (Związek Rzemiosła Polskiego), dr Eugeniusz Śmiłowski (Business Centre Club Związek Pracodawców);
- reprezentujący organizacje związkowe, na mocy ustawy z 6 lipca 2001 r. o Trójstronnej Komisji do Spraw Społeczno-Gospodarczych: Jan Guz — przewodniczący OPZZ, Tomasz Krzemieński — Forum Związków Zawodowych, Maria Osowska — NSZZ „Solidarność";
- ekspertem w zakresie nauk społecznych został prof. dr hab. Henryk Domański, a ekspertem w zakresie nauk ekonomicznych — prof. dr hab. Andrzej Barczak.

11 lipca 2012 r. odbyło się w GUS inauguracyjne posiedzenie nowej Rady Statystyki, podczas którego, w tajnym głosowaniu, przewodniczącym wybrany został prof. dr Franciszek Kubiczek, a jego zastępcą Jan Guz.

Dalsza część obrad poświęcona była projektowi rozporządzenia Rady Ministrów w sprawie programu badań statystycznych statystyki publicznej na rok 2013 oraz projektowi rozporządzenia zmieniającego rozporządzenie Rady Ministrów w sprawie programu badań statystycznych statystyki publicznej na rok 2012. Związane to było z negatywnym stanowiskiem Głównego Inspektora

Ochrony Danych Osobowych (GIODO) w sprawie wprowadzenia zapisów w tematach badawczych dotyczących wtórnego wykorzystania danych osobowych uzyskiwanych na podstawie wcześniejszych badań z Narodowego Spisu Powszechnego Ludności i Mieszkań 2011.

Prezes GUS prof. dr hab. Janusz Witkowski przedstawił zmiany, jakich dokonał GUS w wyniku zaleceń stałego Komitetu RM, wniosków zgłoszonych przez resorty oraz uzgodnień zawartych podczas spotkania prezesa GUS z przedstawicielami GIODO, Ministra Finansów, Ministra Administracji i Cyfryzacji, Ministra Spraw Wewnętrznych i prezesa ZUS. Zgodnie z wnioskiem prof. dra Franciszka Kubiczka, Rada Statystyki po dyskusji przyjęła przez głosowanie uchwałę dotyczącą przyjęcia projektów odpowiednich dokumentów ze zmianami przedstawionymi przez prezesa GUS i rekomendowała je Radzie Ministrów.

Projekt rozporządzenia Rady Ministrów zmieniającego rozporządzenie w sprawie programu badań statystycznych statystyki publicznej na rok 2012 został rozpatrzony przez Stały Komitet RM 26 lipca 2012 r. i uzyskał jego rekomendację. Dokument został przyjęty przez Radę Ministrów 20 sierpnia 2012 r.

Natomiast projekt programu badań statystycznych statystyki publicznej na rok 2013, na wniosek GIODO, skierowany został przez Rządowe Centrum Legislacji do Rady Legislacyjnej przy Prezesie Rady Ministrów.

Opinia Rady Legislacyjnej była przedmiotem dyskusji w trakcie wrześniowego posiedzenia Rady Statystyki. Opinię przedstawiły Agnieszka Komar-Morawska — dyrektor Gabinetu Prezesa GUS oraz naczelnik Ewa Ossowska. Rada Legislacyjna zakwestionowała możliwość uzyskiwania danych osobowych i danych indywidualnych przez statystykę publiczną, a dotychczasowe działania w tym zakresie określiła jako niekonstytucyjne. W dyskusji zwrócono uwagę na: podważanie przez Radę Legislacyjną podstaw prawnych funkcjonowania statystyki, konieczność przekazania sprawy konstytucjonalistom oraz przeprowadzenia pilnych zmian ustawowych, potrzebę precyzyjnego określenia pojęcia danych osobowych, konieczność wykorzystania doświadczeń innych krajów czy odmienne stanowisko wielu państw europejskich ukierunkowujących się na dostępność do danych administracyjnych.

Rada Statystyki przyjęła zastosowanie w projekcie programu badań statystycznych statystyki publicznej na rok 2013 rozwiązań, analogicznych jak w programie na rok 2012, czyli niekolidujących ze stanowiskiem Rady Legislacyjnej. Projekt rozporządzenia w sprawie programu badań statystycznych statystyki publicznej na rok 2013, Stały Komitet RM rekomendował 18 października 2012 r. Radzie Ministrów, która przyjęła dokument 9 listopada 2012 r.

Kolejną (trzecią) zmianę programu badań statystycznych statystyki publicznej na rok 2012, na wniosek prezesa GUS, Rada skierowała do rozpatrzenia przez Stały Komitet RM w grudniu 2012 r. Dokument został przyjęty przez Radę Ministrów 4 stycznia 2013 r.

W październiku 2012 r. Rada podjęła prace nad projektem programu badań statystycznych statystyki publicznej na rok 2014, przyjmując przedstawiony przez GUS harmonogram prac i założenia do opracowania tego dokumentu.

Jednocześnie prof. dr Franciszek Kubiczek przedstawił propozycję szczegółowego podziału zadań pomiędzy członków Rady Statystyki w pracach nad projektem.

Podczas posiedzenia listopadowego członkowie Rady przedstawili propozycje dotyczące zmian i uzupełnień w badaniach planowanych na rok 2014, a w trakcie obrad grudniowych Anna Dobrowolska, zastępca dyrektora Departamentu Programowania i Koordynacji Badań GUS, omówiła zgłoszone przez autorów badań i użytkowników statystyki propozycje badań nowych, znacznie zmienionych oraz do usunięcia, w porównaniu do programu na rok 2013.

W II półroczu 2012 r. Rada podejmowała również inne tematy oraz dokonywała przeglądu istotnych problemów statystyki publicznej.

Podczas posiedzenia z 30 lipca 2012 r. członkowie Rady poznali raport z Narodowego Spisu Powszechnego Ludności i Mieszkań 2011, który przedstawiły Lucyna Nowak — zastępca dyrektora Departamentu Badań Demograficznych GUS oraz Dorota Szałtys — naczelnik wydziału.

Prelegentki zwróciły uwagę na obszerny rozdział zawierający metodologię spisu oraz na dane dotyczące ostatecznej faktycznej liczby ludności Polski według stanu z 31 marca 2011 r. wynoszącej 38,5 mln. Porównując dane spisowe z 2011 r. z danymi ze spisu z 2002 r. wskazały na zahamowanie urbanizacji Polski, na niewielkie zmiany struktury ludności według płci, spadek odsetka ludności w wieku przedprodukcyjnym (w wyniku niekorzystnych trendów demograficznych), wzrost odsetka ludności z wykształceniem wyższym (głównie w miastach) oraz zróżnicowanie regionalne wskaźników demograficznych.

Członków Rady Statystyki interesowała m.in. weryfikacja danych spisowych, różnice pomiędzy danymi z BAEL i spisowymi, zmiany prognozy demograficznej wynikające z nowych danych spisowych oraz udostępnianie danych spisowych.

Radzie Statystyki przedstawiono też projekt rozporządzenia Parlamentu Europejskiego i Rady Unii Europejskiej (UE) w sprawie statystyki europejskiej w dziedzinie demografii. Lucyna Nowak omówiła genezę i założenia tego dokumentu. Zaznaczyła, że podstawowym celem rozporządzenia jest uregulowanie prawne statystyki w dziedzinie demografii pod kątem dostarczania Komisji Europejskiej danych w odpowiednim terminie, właściwej jakości i porównywalnych w układzie międzynarodowym. Przedstawiając prace prowadzone przez polski rząd w sprawie projektu rozporządzenia oraz prace grupy roboczej Rady UE do spraw statystyki, wskazała na kontrowersje dotyczące braku w dokumencie jednolitej definicji liczby ludności kraju i związanego z tym zobowiązania w art. 4 do przekazywania na potrzeby parlamentu UE liczby ludności danego kraju bez emigrantów przebywających za granicą co najmniej 12 miesięcy.

Podsumowując dyskusję prof. dr hab. Zbigniew Strzelecki podkreślił trudności w zharmonizowaniu statystyki w dziedzinie demografii wynikające z różnic metodologicznych, jakie mają miejsce w krajach UE. Zaznaczył, że przy ogólnie prawidłowych założeniach projektu rozporządzenia PE i Rady w dziedzinie de-

mografii zawiera on jednak niepotrzebnie elementy pozastatystyczne. Zwrócił uwagę, że wprowadzenie dokumentu w życie związane będzie z koniecznością nowelizacji w Polsce wielu przepisów oraz spowodować może określone skutki finansowe wynikające z konieczności dodatkowego przeliczania danych statystycznych. Prof. dr hab. Zbigniew Strzelecki pozytywnie ocenił prace polskiej delegacji nad omawianym dokumentem unijnym.

Podczas posiedzenia wrześniowego Rada zajęła się potrzebami statystycznymi jednostek samorządu terytorialnego wynikającymi z ustawowych obowiązków opracowywania programów lokalnych i regionalnych. Temat przedstawił Andrzej Porawski. Koreferowali Dominika Rogalińska — dyrektor Departamentu Badań Regionalnych i Środowiska GUS oraz dr Marek Pieniążek — główny specjalista.

Rada wysłuchała też informacji o kierunkach nowelizacji rozporządzenia Parlamentu Europejskiego i Rady nr 223/2009 w sprawie statystyki europejskiej według projektu zmian Komisji Europejskiej z 17 kwietnia 2012 r. Kierunki zmian przedstawili Marek Mroczkowski — dyrektor Departamentu Statystyki Międzynarodowej GUS oraz Barbara Jarosińska — naczelnik.

Przewodniczący Rady Statystyki prof. dr Franciszek Kubiczek wskazał na pozytywne zapisy w projekcie rozporządzenia Parlamentu Europejskiego i Rady UE wzmacniające rolę statystyki oraz prezesa urzędu. Wymienił tu przede wszystkim podniesienie znaczenia krajowego urzędu statystycznego w udostępnianiu administracyjnych źródeł danych oraz wskazał na obowiązek konsultacji w tworzeniu takich źródeł, jak i ich zmian czy likwidacji. Zwrócił również uwagę na zapisy dotyczące procedur nominacji i odwoływania prezesów urzędów krajowych oraz sporządzania przez kraje członkowskie sprawozdań rocznych ze spełniania poszczególnych zobowiązań. Przewodniczący Rady podkreślił konieczność uwzględnienia zapisów z rozporządzenia Parlamentu Europejskiego i Rady w projektowanej zmianie ustawy o statystyce publicznej.

W trakcie posiedzenia październikowego Rada Statystyki zapoznała się z pracami nad Europejskimi Rachunkami Ekonomicznymi Środowiska (EREŚ), które przedstawiła Wiesława Domańska — zastępca dyrektora Departamentu Badań Regionalnych i Środowiska GUS. Idea utworzenia rachunków środowiskowych powstała w wyniku zainteresowania ekonomistów sprawami związanymi ze środowiskiem naturalnym. Celem rachunków środowiskowych, jako rachunków satelitarnych do rachunków narodowych, jest interakcja gospodarka——środowisko, stwarzająca wiele nowych możliwości analitycznych, pozwalająca na rozszerzenie wskaźników monitorujących oraz doskonalenie planowania przestrzennego.

Wiesława Domańska wskazała na podstawę prawną EREŚ, jaką jest rozporządzenie Parlamentu Europejskiego i Rady UE nr 691/2011 z 6 lipca 2011 r. Rozporządzenie to jako cel rachunków określa ustanowienie wspólnych ram dla gromadzenia, zestawiania, przekazywania oraz oceny EREŚ oraz umiejscowienie rachunków ekonomicznych środowiska jako rachunków satelitarnych w stosunku do ESA' 95. Określa też metody osiągnięcia tego celu poprzez usta-

lenie metodyki, wspólnych standardów, definicji, klasyfikacji i zasad rachunkowości, jakie mają być stosowane przy opracowywaniu rachunków ekonomicznych środowiska. Do spraw rachunków środowiskowych prezes GUS powołał grupę zadaniową, w skład której wchodzą przedstawiciele: GUS, Ministerstwa Środowiska, Ministerstwa Finansów, Ministerstwa Gospodarki oraz Uniwersytetu Warszawskiego i Urzędu Statystycznego w Krakowie.

W dyskusji podkreślono m.in. coraz większe znaczenie rachunków środowiskowych przy monitorowaniu realizacji celów założonych w dokumentach strategicznych, konieczność wypracowania wspólnej metodologii opracowywania rachunków środowiskowych we wszystkich krajach europejskich, potrzebę uwzględnienia w rachunkach wpływu środowiska na zdrowie ludności oraz znaczenie dla Polski problematyki związanej z gospodarką zasobami wodnymi.

Rada pozytywnie ustosunkowała się do przebiegu dotychczasowej współpracy GUS z Ministerstwem Środowiska i innymi jednostkami, pozwalającej optymistycznie oceniać realizację nakładanych obowiązków w zakresie EREŚ.

Podczas posiedzenia listopadowego Rada Statystyki dokonała oceny spójności rocznego programu badań statystycznych statystyki publicznej na rok 2013 z Europejskim Programem Statystycznym 2013—2017 według założeń rozporządzenia Parlamentu Europejskiego i Rady UE. Temat przedstawili dr Marek Mroczkowski, Elżbieta Sudnicka — naczelnik wydziału i Anna Dobrowolska. Zgodnie z przedstawionymi materiałami Rada przyjęła, że badania ujęte w programie badań statystycznych statystyki publicznej na rok 2013 są w większości dostosowane do wymogów UE i tym samym do celów zawartych w Europejskim Programie Statystycznym 2013—2017. Jednocześnie zaznaczono, że zakładane modyfikacje zakresu tematycznego w kolejnych rocznych programach badań statystyki publicznej umożliwią, w perspektywie do 2017 r., realizację celów wyznaczonych w Europejskim Programie Statystycznym 2013—2017.

W kolejnym punkcie posiedzenia Rada Statystyki zapoznała się ze stanem i potrzebami zmian w podziale statystycznym Polski według NUTS, z powodu upływającego terminu przedłożenia ewentualnych zmian w tym zakresie do Eurostatu. Uczestniczący w posiedzeniu Marceli Niezgoda, podsekretarz stanu w Ministerstwie Rozwoju Regionalnego, przekazał oficjalną opinię, w której nie zaleca się dokonywania zmian na poziomie NUTS 2. Zaznaczył, że ewentualna zmiana podziału na poziomie NUTS 3 powinna być ograniczona do nielicznych przypadków, uzasadnionych merytorycznie przez samorządy. Podobną opinię przedstawiła Dominika Rogalińska zaznaczając, że zapewnienie stabilności klasyfikacji, umożliwiające tym samym porównywanie danych w długich szeregach czasowych, przemawia za utrzymaniem do celów statystycznych obecnego podziału terytorialnego. W trakcie dyskusji polemizowano z powyższymi opiniami zwracając m.in. uwagę, że wszelkie zmiany w zakresie podziału NUTS 2 i 3 poprzedzone powinny być oceną aktualnego stanu pod względem funkcjonalności i przydatności z punktu widzenia planów i strategii opracowywanych

w województwach. Rada Statystyki powołała czteroosobowy zespół do opracowania propozycji związanych z korektą NUTS.

Temat ewentualnych zmian w podziale NUTS Rada kontynuowała podczas posiedzenia grudniowego. Dominika Rogalińska przedstawiła informację o uzgodnieniach z urzędami marszałkowskimi w zakresie korekty na poziomie NUTS 3, które Rada Statystyki zaopiniowała pozytywnie. Na skutek wątpliwości dotyczących korekty NUTS 2, Rada zwróciła się do ministra rozwoju regionalnego, prezesa GUS oraz marszałka województwa mazowieckiego o przeprowadzenie konsultacji w trybie międzyresortowym dla wypracowania jednolitego stanowiska w tym zakresie.

Podczas obrad grudniowych Agnieszka Komar-Morawska przedstawiła informację o toczących się pracach nad zmianami przepisów obowiązującej ustawy o statystyce publicznej. Projektowane korekty wynikają ze zmieniających się uwarunkowań funkcjonowania statystyki publicznej, w tym zmian zachodzących w prawodawstwie UE i prawie krajowym oraz konieczności doprecyzowania kwestii będących przedmiotem tej regulacji.

Obecnie prace przebiegają dwutorowo. Niezależnie od prac nad nową ustawą o statystyce publicznej, w której uwzględnione mają być m.in. przygotowywane przez Parlament Europejski i Radę UE nowe rozwiązania prawne dotyczące statystyki, trwają prace nad nowelizacją obowiązującej ustawy o statystyce publicznej, w celu zapewnienia niebudzących kontrowersji podstaw prawnych dla prac nad projektem programu badań na rok 2014.

Rada Statystyki pozytywnie oceniła prowadzone w GUS prace związane z nowelizacją przepisów ustawy o statystyce publicznej. W dyskusji zwracano uwagę przede wszystkim na konieczność doprecyzowania w nowym ujęciu prawnym kwestii związanych z pojęciem tajemnicy statystycznej, dostępem statystyki do danych jednostkowych w źródłach administracyjnych oraz udostępnianiem danych przez statystykę. Wskazano również, że zmiany w ustawie o statystyce publicznej spowodują konieczność korekt zapisów w innych ustawach.

Obrady grudniowe Rada Statystyki zakończyła przyjęciem planu pracy na rok 2013. Założono odbycie 11 posiedzeń plenarnych, podczas których obok prac nad projektami programów badań statystycznych na rok 2014 i na rok 2015, zaplanowano m.in. kontynuację tematów związanych z wykorzystaniem administracyjnych źródeł danych w statystyce publicznej, zapoznanie się z nowymi metodami pomiaru ubóstwa oraz z innymi niż PKB pomiarami rozwoju społeczno-gospodarczego, a także rozpatrzenie metody i technik komunikowania się statystyki publicznej ze społeczeństwem. Rada planuje również rozpatrzenie koncepcji wdrożenia nowej organizacji systemu badań statystyki publicznej, statystyki szarej strefy oraz dokonanie przeglądu światowych indeksów gospodarczych.

mgr Antoni Żurawicz — *GUS*

INFORMACJE. PRZEGLĄDY. RECENZJE

XXXI międzynarodowa konferencja naukowa *Wielowymiarowa analiza statystyczna*

Od 12 do 14 listopada 2012 r. w Centrum Szkoleniowo-Konferencyjnym Uniwersytetu Łódzkiego odbyła się XXXI międzynarodowa konferencja naukowa *Wielowymiarowa analiza statystyczna* (WAS 2012) (*The 31-st International Annual Conference on Multivariate Statistical Analysis, MSA 2012*). Organizatorami konferencji były Instytut Statystyki i Demografii UŁ, Katedra Metod Statystycznych UŁ oraz Polskie Towarzystwo Statystyczne (PTS). Funkcję Przewodniczącego Komitetu Organizacyjnego objął prof. dr hab. Czesław Domański, a obowiązki sekretarzy naukowych konferencji pełnili dr Artur Mikulec oraz mgr Aleksandra Kupis-Fijałkowska.

Konferencja objęta była honorowym patronatem prezesa GUS, prof. dra hab. Janusza Witkowskiego. Jej organizacja została dofinansowana przez NBP. Partnerami konferencji były też firma StatSoft Polska Sp. z o.o. oraz Polskie Towarzystwo Badaczy Rynku i Opinii (PTBRiO).

Celem WAS 2012 było stworzenie forum wymiany poglądów, dyskusji, a także prezentacja nowych kierunków rozwoju jej metod. Mają one duże znaczenie w dobie coraz większego upowszechnienia badań statystycznych i wzrastającego zapotrzebowania na dane statystyczne dotyczące różnych sfer życia społeczno-gospodarczego.

Tematyka konferencji obejmowała m.in.: rozkłady wielowymiarowe, wnioskowanie statystyczne, metody nieparametryczne, metody klasyfikacji i analizy danych, analizę dyskryminacyjną, metody Monte Carlo, metody bayesowskie, metody eksploracyjnej analizy danych oraz zastosowania metod statystycznych w naukach ekonomicznych, społecznych, przyrodniczych, finansach i ubezpieczeniach.

W konferencji wzięło udział 77 osób z różnych ośrodków akademickich, przedstawiciele urzędów statystycznych i NBP. Podczas dziewięciu sesji uczestnicy wygłosili łącznie 44 referaty oraz 1 wykład zaproszony.

Konferencję otworzył prof. dr hab. Czesław Domański. W imieniu prof. dra hab. Włodzimierza Nykiela, rektora Uniwersytetu Łódzkiego głos zabrał prof. dr hab. Antoni Różalski, prorektor ds. nauki UŁ. Uczestników konferencji przywitał również dr hab. Paweł Starosta, dziekan Wydziału Ekonomiczno-Socjologicznego UŁ. Podczas otwarcia konferencji miało miejsce także wystąpienie prof. dra hab. Eugeniusza Gatnara, członka Zarządu NBP. W imieniu prof. dra hab. Janusza Witkowskiego, prezesa GUS „adres patronacki" przekazał dr hab. Włodzimierz Okrasa, doradca prezesa GUS.

Pierwsza sesja, jubileuszowo-historyczna, poświęcona była stuleciu PTS oraz wybitnym przedstawicielom polskiej myśli statystycznej, przewodniczył jej prof. dr hab. Krzysztof Jajuga (Uniwersytet Ekonomiczny we Wrocławiu). Wykład zaproszony, *Statystyczny paradygmat Jerzego Spławy-Neymana w analizie przyczynowości*, wygłosił prof. dr hab. Tadeusz Bednarski (Uniwersytet Wrocławski). Z jednej strony był on wprowadzeniem w zagadnienia statystycznej analizy przyczynowości zjawisk, w takim ujęciu jak postrzegali je główni twórcy współczesnej statystyki — Ronald A. Fisher, Jerzy Spława-Neyman i Egon S. Pearson — z drugiej strony, okazją do wspomnień kontaktów prof. Bednarskiego z prof. Neymanem z okresu studiów doktorskich odbytych na Uniwersytecie Kalifornijskim w Berkeley w latach 1973—1976. Następnie prof. dr hab. Mirosław Krzyśko (Państwowa Wyższa Szkoła Zawodowa im. Prezydenta Stanisława Wojciechowskiego w Kaliszu) wygłosił referat *Karolina Iwaszkiewicz-Gintowt (1902—1999). Współpracowniczka Jerzego Neymana z okresu warszawskiego*, a prof. dr hab. Czesław Domański (Uniwersytet Łódzki) z okazji 120 rocznicy urodzin przybliżył postać Rajmunda Buławskiego, *Rajmund Buławski — założyciel Śląsko-Dąbrowskiego oddziału Polskiego Towarzystwa Statystycznego*, oraz Jana Piekałkiewicza.

Pierwszego dnia konferencji odbyły się też dwie **sesje równoległe**, którym przewodniczyli odpowiednio prof. dr hab. Krystyna Katulska (Uniwersytet im. Adama Mickiewicza w Poznaniu) oraz prof. dr hab. Marek Walesiak (Uniwersytet Ekonomiczny we Wrocławiu), w każdej wygłoszono po 4 referaty.

Dr hab. Grażyna Dehnel, dr Tomasz Klimanek (Uniwersytet Ekonomiczny w Poznaniu) oraz dr Jacek Kowalewski (Urząd Statystyczny w Poznaniu) przedstawili referat *Estymacja pośrednia z uwzględnieniem korelacji przestrzennej w statystyce gospodarczej*. Autorzy podjęli próbę wykorzystania metod estymacji pośredniej do oszacowania celu średnich i dużych przedsiębiorstw w woj. wielkopolskim. W analizie wykorzystano informacje pochodzące z badania DG-1, stanowiące podstawę do opracowywania większości krótkookresowych wskaźników dotyczących działalności przedsiębiorstw. W przeprowadzonym badaniu do konstrukcji odpowiednich estymatorów statystyki małych obszarów wykorzystano, oprócz modeli przekrojowych, także najnowsze podejście w estymacji pośredniej, uwzględniające zależności przestrzenne.

Dr Tomasz Żądło (Uniwersytet Ekonomiczny w Katowicach) w referacie *O estymacji MSE dla pewnego predyktora w przypadku złej specyfikacji modelu* omawiał problem predykcji wartości globalnej w podpopulacji (domenie). Analizował wykorzystanie predyktora, który jest empirycznym najlepszym liniowym nieobciążonym predyktorem, ale przy założeniu błędnego modelu. Dla rozważanego predyktora wyprowadzono postać naiwnego estymatora *MSE* dla prawidłowego modelu nadpopulacji oraz zaproponowano wykorzystanie estymatorów *MSE* typu *jackknife* i parametryczny *bootstrap*. W badaniu symulacyjnym analizowano względne obciążenia zaproponowanych estymatorów *MSE*.

Dr inż. Wojciech Gamrot (Uniwersytet Ekonomiczny w Katowicach) przedstawił referat *O pewnym sposobie zapożyczania mocy w estymacji wartości globalnej*, w którym opisana była estymacja wartości globalnej ustalonej cechy w populacji skończonej na podstawie prób losowanych złożonymi schematami losowania, o nieznanym prawdopodobieństwie inkluzji pierwszego rzędu. Do oceny tego parametru proponuje się wykorzystanie ogólnego empirycznego estymatora Horvitza-Thompsona. W sytuacji gdy dostępne są dodatkowe informacje o szacowanym prawdopodobieństwie inkluzji w postaci ograniczeń nierównościowych, definiujących prosty porządek, rozważane jest wykorzystanie nieparametrycznego estymatora jądrowego. Prowadzi to do konstrukcji nieparametrycznego empirycznego estymatora Horvitza-Thompsona wartości globalnej.

Referat *O szacowaniu liczby modeli bazowych wykorzystując testy parametryczne i permutacyjne* wygłosił mgr inż. Jacek Stelmach (Uniwersytet Ekonomiczny w Katowicach). Autor poruszył temat podejścia wielomodelowego w analizie regresji. Ważnym problemem przy agregacji modeli staje się określenie optymalnej liczby modeli bazowych, które po agregacji pozwalają na utworzenie prognozy o jak najmniejszym błędzie, przy czym analiza samego błędu predykcji może doprowadzić do przeszacowania liczby modeli bazowych. W przeprowadzonym eksperymencie autor zaproponował podejście formalne, będące adaptacją propozycji Latinne, Debein i Decaestecker, przeznaczonej dla modeli klasyfikacyjnych, w którym dokonuje się porównania prognoz otrzymywanych za pomocą modeli zagregowanych z różnej liczby modeli bazowych. W podejściu tym wykorzystano zarówno testy parametryczne, jak i testy permutacyjne. Zaproponowano także statystykę testową wykorzystującą miary odległości pomiędzy prognozami modeli zagregowanych.

Dr Daniel Kosiorowski (Uniwersytet Ekonomiczny w Krakowie) oraz Mateusz Bocian, Anna Węgrzynkiewicz i Zygmunt Zawadzki (studenci Uniwersytetu Ekonomicznego w Krakowie) przygotowali referat *Pakiet {DepthProc} w eksploracyjnej analizie wielowymiarowego szeregu czasowego*. Omówili w nim pakiet środowiska R własnego autorstwa o nazwie {*DepthProc*}, który zawiera implementacje kilku wielowymiarowych procedur statystycznych indukowanych przez statystyczne funkcje głębi. Przedstawili przykłady zastosowań procedur pakietu w eksploracyjnej analizie wielowymiarowego szeregu czasowego.

Z kolei mgr Justyna Brzezińska (Uniwersytet Ekonomiczny w Katowicach) w referacie *Iloraz szans w analizie tablic kontyngencji* omówiła metodę opisu tablic kontyngencji. Iloraz szans jest miernikiem wykorzystywanym do badania związku w tablicach kontyngencji. Można go uogólnić dla tablic wielodzielczych poprzez użycie lokalnych ilorazów szans oraz podejścia opartego na krzyżowaniu komórek. W ilustracji empirycznej metody wykorzystano funkcje zaimplementowane w pakietach *{vcd}* i *{vcdExtra}* środowiska R.

W referacie *Konstrukcja skali pomiaru postaw konsumenckich w ujęciu jednoparametrycznego modelu Rascha i danych dychotomicznych* dr Piotr Tarka

(Uniwersytet Ekonomiczny w Poznaniu) zajął się zagadnieniem dotyczącym konstrukcji skali pomiarowej służącej do badania postaw konsumenckich. Opisano specyfikę pomiaru oraz kluczowe założenia teoretyczne dotyczące konstrukcji skal pomiarowych w badaniach marketingowych. Następnie Autor zwrócił uwagę na charakter danych, wobec których istnieją różnorodne możliwości budowy modelu pomiarowego. Podał też przykład konstrukcji skali na podstawie danych dychotomicznych, gdzie analizie poddano postawy konsumpcyjno-hedonistyczne ludzi.

Referat *Wpływ formuły normalizacyjnej na wynik porządkowania liniowego* przedstawiły dr Katarzyna Dębkowska oraz mgr Marta Jarocka (Politechnika Białostocka). Znalazły się w nim rezultaty eksperymentu, którego celem była analiza wpływu różnych procedur normalizacyjnych na wynik porządkowania liniowego obiektów. W obliczeniach wykorzystano m.in. standaryzację, standaryzację Webera, unitaryzację, unitaryzację zerowaną oraz wybrane przekształcenia ilorazowe. Procedurę badawczą oparto na danych zaczerpniętych z bazy Eurostat.

Podczas **III sesji konferencji**, której przewodniczącym był prof. dr hab. Bronisław Ceranka (Uniwersytet Przyrodniczy w Poznaniu) wygłoszono 3 referaty.

Jako pierwszy głos zabrał prof. dr hab. Włodzimierz Okrasa (doradca prezesa GUS, Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie), który przedstawił *Uwagi metodologiczne nt. interakcji metod i danych w statystyce społecznej — przypadek badań statystyczno-socjologicznych*. W referacie omówiono wybrane wątki rozwoju badań statystycznych, dotyczące danych jakościowych i analizy przyczynowej, którą można wyróżnić w literaturze dzięki przyjęciu perspektywy socjologicznej.

Natomiast w referacie dr hab. Aliny Jędrzejczak (Uniwersytet Łódzki) *Ubóstwo i nierówności dochodowe w rodzinach wielodzietnych* omówiono problem pomiaru nierówności dochodowych i ubóstwa. Podjęto próbę oceny zmian sytuacji dochodowej rodzin wielodzietnych na podstawie badania budżetów gospodarstw domowych. W analizie uwzględnione zostały współczynniki koncentracji dochodów, miary ubóstwa oraz współczynniki dystansu ekonomicznego, a także oceny błędów standardowych oszacowań.

Prof. dr hab. Tadeusz Gerstenkorn (Uniwersytet Łódzki) i dr Jacek Mańko (XXXI LO w Łodzi) przygotowali referat *Prawdopodobieństwo zdarzenia rozmytego i jego zastosowanie w problemach ekonomicznych*. Autorzy zarysowali ogólną ideę tzw. zbioru rozmytego, wprowadzoną do nauki i praktyki przez Zadeha w 1965 r. Następnie omówili uogólnienie tej teorii, zaproponowane przez Atanassova. Ukazano zalety wprowadzenia do rozważań, oprócz tzw. funkcji przynależności, także funkcji nieprzynależności elementu do pewnego zbioru, tak scharakteryzowane zbiory nazywa się *intuicjonistycznymi rozmytymi* lub *dwoisto rozmytymi.* Następnie autorzy przedstawili koncepcję prawdopodobień-

stwa zdarzenia rozmytego oraz sposób obliczenia tego prawdopodobieństwa według różnych koncepcji.

Na kolejnych **sesjach równoległych**, którym przewodniczyli prof. dr hab. Mirosław Krzyśko oraz prof. dr hab. Grażyna Trzpiot (Uniwersytet Ekonomiczny w Katowicach), przedstawiono 6 referatów.

Dr Dariusz Parys (Uniwersytet Łódzki) zgłosił referat *Kroczące procedury testów wielokrotnych dla rozkładów dyskretnych*. Zaproponowano w nim nowe procedury wielokrotnego testowania w przypadku danych pochodzących z populacji o rozkładzie dyskretnym. Wybierając procedurę opartą na badaniach Tarone'a, Westfalla i Welfingera porównano ją z innymi procedurami testowania wielokrotnego i pokazano większą moc tej procedury.

Z kolei dr Justyna Wilk i dr Marcin Pełka (Uniwersytet Ekonomiczny we Wrocławiu) opracowali referat *Analiza skupień danych symbolicznych. Przegląd badań*. Autorzy omówili koncepcję, źródła i znaczenie danych symbolicznych w opisie zjawisk. Opisali oni systematykę metod klasyfikacji danych symbolicznych, z wydzieleniem metod taksonomii numerycznej i symbolicznej, a także dokonali przeglądu ich zastosowań w analizie danych symbolicznych. Rozważania uzupełniono przykładami wykorzystania proponowanych rozwiązań metodologicznych w badaniach empirycznych.

Mgr Małgorzata Szerszunowicz (Uniwersytet Ekonomiczny w Katowicach) była autorką referatu *Kolejność realizacji doświadczeń w planie eksperymentu a koszty procesu produkcyjnego*. Przedmiotem rozważań był odpowiedni dobór kolejności wykonywania doświadczeń eksperymentu w celu zmniejszenia kosztów jego realizacji. Dla tak określonego problemu zaprezentowany został algorytm realizacji planu eksperymentu wykorzystujący macierze przejścia z odpowiednio określonym prawdopodobieństwem przejścia. Proponowana metoda przeprowadzenia eksperymentu została porównana z klasycznymi metodami planowania eksperymentu, z uwzględnieniem kosztów realizacji.

Natomiast mgr Marta Małecka (Uniwersytet Łódzki) miała referat *Projektowanie eksperymentów symulacyjnych przy ocenie prognoz VaR*. Tematyka jego dotyczyła procedur testowych, służących ocenie jakości prognostycznej modeli, w tym własności statystycznych i zastosowań dwóch najbardziej popularnych testów: Kupca z 1995 r., badającego udział przekroczeń *VaR* oraz autokorelacji Christoffersena z 1998 r. Celem opracowania było omówienie proponowanych w literaturze eksperymentów symulacyjnych oraz pokazanie ich wpływu na wyniki analiz. Autorka skoncentrowała się na weryfikacji braku autokorelacji przekroczeń *VaR*. W kontekście badania własności statystycznych test Christoffersena został porównany do testu zaproponowanego w 2004 r. przez Engla i Mangianelliego. Ponadto przedstawiono własną propozycję projektu eksperymentu oceniającego moc testów autokorelacji *VaR*.

Dr Beata Bieszk-Stolorz i dr Iwona Markowicz (Uniwersytet Szczeciński) uczestnikom konferencji zaproponowały referat *Wpływ zmiennej objaśniającej*

na hazard i jego zmiany w czasie w modelu regresji Coxa. Była to analiza wpływu wieku osób bezrobotnych na okres poszukiwania przez nich zatrudnienia. Analizę taką umożliwia zastosowanie modelu nieproporcjonalnego hazardu *Coxa*, w którym zakłada się, że wpływ zmiennej objaśniającej na hazard zależy od czasu, a więc ilorazy hazardu zmieniają się. Stosując model proporcjonalnych hazardów można uzyskać stałe oceny parametrów. W referacie oszacowano modele hazardu dla funkcji $g(t)$ zdefiniowanej w różny sposób.

Dr inż. Jacek Białek (Uniwersytet Łódzki) zgłosił *Uwagi o uogólnionych formułach indeksów cen*. W referacie przedstawiono trzy ogólne formuły indeksów cen, w tym dwie autorskie. Pokazano, że proponowane formuły spełniają podstawowe testy pochodzące z aksjomatycznej teorii indeksów, a w szczególnych przypadkach dają sprowadzić się do powszechnie znanych indeksów cen. Trzecia, najbardziej ogólna, formuła jest kolejnym krokiem w tzw. podejściu Divisia dla czasu ciągłego.

Drugi dzień konferencji otworzyła **sesja plenarna**, której przewodniczącą była dr hab. Alina Jędrzejczak (Uniwersytet Łódzki).

Wystąpienie prof. dra hab. Eugeniusza Gatnara (NBP, Uniwersytet Ekonomiczny w Katowicach), *Wskaźniki włączenia finansowego w Polsce (z perspektywy Narodowego Banku Polskiego)*, dotyczyło tematyki „włączenia finansowego", definiowanego przez ONZ jako stan, w którym wszyscy ludzie mają dostęp do pełnego zestawu dobrej jakości usług finansowych, dostarczanych w dostępnych cenach, w dogodny sposób z szacunkiem dla klientów. W referacie przedstawione zostały rodzaje danych zbieranych i kompilowanych przez systemy statystyczne systemu płatniczego oraz bilansu płatniczego (NBP). Na ich podstawie obliczono wskaźniki poziomu włączenia finansowego w Polsce, których wartości porównano z odpowiednimi wskaźnikami uzyskanymi dla innych krajów Unii Europejskiej.

Referat dra hab. Grzegorza Kończaka (Uniwersytet Ekonomiczny w Katowicach), *O wykorzystaniu rozkładu wartości ekstremalnych w monitorowaniu procesów*, związany był z tematyką statystycznej kontroli jakości. W wielu analizach statystycznych ważna jest ocena poziomu przeciętnego badanej charakterystyki, tak jest np. przy monitorowaniu procesów z wykorzystaniem karty kontrolnej $\overline{X}$. Jednak w wielu zastosowaniach interesująca może być ocena wielkości maksymalnych lub minimalnych. Przedstawiono propozycję wykorzystania własności rozkładu wartości ekstremalnych w monitorowaniu procesów. Rozważania teoretyczne zostały uzupełnione analizami symulacyjnymi własności proponowanej metody.

Prof. dr hab. Wojciech Zieliński (SGGW w Warszawie) w referacie *Zastosowanie najkrótszych przedziałów ufności dla frakcji w kontrolach NIK* kontynuował w innym ujęciu poprzednią tematykę. W statystycznej kontroli jakości badane obiekty poddane są również ocenie alternatywnej i badacza interesuje odsetek obiektów ocenionych negatywnie. Jednym z przykładów takich zasto-

sowań są kontrole jakości prowadzone przez NIK, których celem jest m.in. wykrywanie nieprawidłowości w pracy urzędów skarbowych. Autor badał statystyczne własności planów badań podanych przez Karlińskiego. W referacie Autor zaproponował i omówił zastosowanie najkrótszych przedziałów ufności w kontrolach jakości.

Szóstej sesji plenarnej przewodniczył prof. dr hab. Eugeniusz Gatnar.

Dr Jerzy Korzeniewski (Uniwersytet Łódzki) przedstawił referat *Modyfikacja metody HINoV selekcji zmiennych w analizie wielokrotnych struktur skupień*. Autor zaproponował, aby przy wyborze zmiennych w analizie skupień oddzielnie dla każdej ustalonej zmiennej grupować zmienne w dwie klasy — podobnych i niepodobnych do niej w sensie podobieństwa podziału zbioru danych na zadaną liczbę skupień (od 2 do 10). Otrzymana wówczas macierz zero-jedynkowa, opisująca związki pomiędzy każdą parą zmiennych, może przy zastosowaniu kryterium optymalizującego zostać wykorzystana do wybrania zmiennych tworzących tę samą (najsilniejszą) strukturę skupień. W referacie szczegółowo omówiono zaproponowane podejście, a modyfikację oceniono na podstawie eksperymentu symulacyjnego z wykorzystaniem 9540 zbiorów danych wygenerowanych w postaci mieszanin rozkładów normalnych.

Wystąpienie dra Mariusza Kubusa (Politechnika Opolska) *Wybrane uwagi na temat podejścia wrappers bazującego na rankingu zmiennych* dotyczyło zagadnienia selekcji zmiennych w analizie dyskryminacji lub regresji. Podejście *wrappers* (wyboru zmiennych) polega na wykorzystaniu modeli budowanych na różnych podzbiorach zmiennych. Do wyboru podzbiorów zmiennych stosowane są techniki przeszukiwania (heurystyczne lub stochastyczne). W referacie porównano kilka wariantów procedury selekcji zmiennych. Zaproponowane zostało też konkurencyjne kryterium rangowania zmiennych, celem ich wyboru. W badaniach empirycznych rozważano zmienne nieistotne metryczne oraz binarne.

Dr Artur Zaborski (Uniwersytet Ekonomiczny we Wrocławiu) w referacie *Grawitacyjna analiza unfolding dla niesymetrycznej macierzy podobieństw* przedstawił różnice pomiędzy klasyczną, wielowymiarową analizą *unfolding* a grawitacyjną analizą *unfolding*, która może być również wykorzystana do graficznej prezentacji danych zawartych w niesymetrycznej macierzy podobieństw. Macierz taka zawiera informacje na temat zmian preferencji respondentów w pewnym okresie. W rezultacie nie otrzymujemy na mapie percepcyjnej, tak jak w przypadku analizy *unfolding*, konfiguracji punktów i respondentów, lecz konfigurację punktów reprezentujących obiekty w różnych okresach.

Z kolei dr Katarzyna Cheba (Zachodniopomorski Uniwersytet Technologiczny w Szczecinie) oraz dr Maja Kiba-Janiak (Uniwersytet Ekonomiczny we Wrocławiu) w referacie *Conjoint analysis jako metoda analizy preferencji konsu-*

mentów na przykładzie rynku miejskich usług transportowych przedstawiły charakterystykę analizy *conjoint* wykorzystywanej do ustalenia, w jaki sposób konsumenci wyznaczają preferencje w stosunku do określonych produktów lub usług. Autorki przedstawiły i zinterpretowały wyniki własnych badań empirycznych na temat preferencji konsumentów na przykładzie transportu miejskiego w Gorzowie Wielkopolskim, Zielonej Górze i Jeleniej Górze.

Przewodniczącymi dwóch kolejnych **sesji równoległych** byli dr hab. Józef Dziechciarz (Uniwersytet Ekonomiczny we Wrocławiu) oraz prof. dr hab. Wojciech Zieliński. Wygłoszono na nich 8 referatów.

Nieklasyczne miary ryzyka na przykładzie GPW w Warszawie — zastosowanie rozkładów alfa-stabilnych — mgr Dominik Krężołek (Uniwersytet Ekonomiczny w Katowicach). Przedmiotem referatu była prezentacja zastosowania metodologii rozkładów alfa-stabilnych w pomiarze ryzyka inwestycyjnego. Wykorzystano nieklasyczne mierniki ryzyka wykorzystującego metodologię *Value-at-Risk*: *Expected Shortfall* oraz *Median Shortfall* dla wybranych rzędów kwantyli, a badanie przeprowadzono na podstawie dziennych logarytmicznych stóp zwrotu wybranych walorów Giełdy Papierów Wartościowych w Warszawie.

Z kolei dr Iwona Konarzewska (Uniwersytet Łódzki) opracowała referat *Optymalny portfel akcji — wykorzystanie wielowymiarowej analizy statystycznej*. Zaproponowano w nim oryginalny algorytm preselekcji aktywów do portfela inwestycyjnego oraz modele optymalizacyjne służące do ustalenia składu portfela inwestycyjnego. W obu propozycjach wykorzystana została dekompozycja macierzy korelacji stóp zwrotu z akcji według wartości i wektorów własnych. Wyniki teoretyczne zilustrowano przykładami analizy dla sektora średnich spółek — komponentów indeksu mWIG40 GPW w Warszawie.

Dr Joanna Trzęsiok (Uniwersytet Ekonomiczny w Katowicach) była autorką referatu *Wybrane symulacyjne procedury porównywania nieparametrycznych metod regresji*. Przedstawiła w nim symulacyjną procedurę badawczą, pozwalającą na porównywanie różnych, zarówno klasycznych, jak i nieparametrycznych modeli regresji. W pierwszym etapie analizy zbudowane zostały rankingi modeli pod względem ich zdolności predykcyjnych, mierzonych za pomocą estymatora punktowego jakim jest błąd średniokwadratowy, obliczony metodą sprawdzania krzyżowego (MSE_{CV}). Drugi etap analizy miał na celu skorygowanie otrzymanych rankingów. Badanie przeprowadzono na zbiorach danych typowych dla analizy własności modeli regresji, a obliczenia przeprowadzono w środowisku R.

Natomiast dr Michał Trzęsiok (Uniwersytet Ekonomiczny w Katowicach) zgłosił referat *Profilowanie klas w metodzie wektorów nośnych*. Metoda wektorów nośnych (*SVM*) należy do grupy eksploracyjnych metod wielowymiarowej analizy statystycznej i jest osadzona w nurcie statystycznych metod uczących się. *SVM* buduje modele o bardzo dobrych własnościach predykcyjnych,

lecz niewielkiej interpretowalności. W celu uzyskania dodatkowej wiedzy — opisu wyznaczonych klas — stosuje się dodatkowe procedury wspomagające interpretowanie wyników modelowania. W referacie przedstawiono tego typu procedurę wykorzystującą informacje zawarte w „wektorach nośnych" — obserwacjach istotnie wpływających na postać wyznaczonej funkcji dyskryminującej.

Dr Anna Witaszczyk (Uniwersytet Łódzki) przygotowała referat *Zastosowanie teorii macierzy losowych w weryfikacji hipotez statystycznych dotyczących parametrów wielowymiarowego rozkładu normalnego*. W opracowaniu przedstawiono zastosowanie teorii macierzy losowych do weryfikacji hipotezy o równości dwóch wektorów wartości oczekiwanych i do weryfikacji hipotezy, że macierz kowariancji jest macierzą jednostkową. Teoria macierzy losowych zajmuje się badaniem rozkładów wektorów własnych i wartości własnych macierzy losowych. Wiele twierdzeń granicznych dotyczy przekształcenia Stieltjesa funkcji spektralnych macierzy losowych. W twierdzeniach granicznych zakłada się, że wraz ze wzrostem liczby obserwacji również liczba współrzędnych wektora losowego dąży do nieskończoności.

Z kolei dr Justyna Wilk (Uniwersytet Ekonomiczny we Wrocławiu) i dr Michał Bernard Pietrzak (Uniwersytet Mikołaja Kopernika w Toruniu) byli autorami referatu *Analiza migracji wewnętrznych w kontekście aspektów społeczno-gospodarczych — podejście dwuetapowe*. Autorzy zaproponowali dwuetapowe podejście w badaniu zależności między aspektami społeczno-gospodarczymi i zjawiskiem przepływu migracyjnego oraz zastosowali je w analizie migracji wewnętrznych w Polsce. W pierwszym kroku, na podstawie taksonomicznego miernika rozwoju, wyodrębniono podstawowe czynniki kształtujące sytuację społeczno-gospodarczą. W drugim kroku skonstruowany został ekonometryczny model grawitacji, w którym przyjęte aspekty posłużyły jako zmienne objaśniające wielkość przepływów między- i wewnątrzwojewódzkich. W modelu uwzględniony został także stopień urbanizacji, poziom migracji zagranicznych oraz odległość geograficzna jednostek terytorialnych.

Mgr Damian Gajda i dr Tomasz Jurkiewicz (Uniwersytet Gdański) przygotowali referat *Ubezpieczenia w sektorze małych i średnich przedsiębiorstw w latach 2010—2012 (wyniki badań ankietowych)*. Przedstawiono w nim wyniki rocznych badań reprezentacyjnych małych i średnich przedsiębiorstw. W badaniach zbierano informacje o świadomości ubezpieczeniowej przedsiębiorców, ich opiniach i zachowaniach na rynku ubezpieczeń. Dzięki powtarzalności badań i porównywalności narzędzia badawczego możliwe stało się określenie kierunków zmian w zakresie korzystania z ubezpieczeń przez małe i średnie przedsiębiorstwa.

Mgr Anna Jurek (Uniwersytet Łódzki) zgłosiła referat *Charakterystyka ubezpieczenia z tytułu zdarzeń medycznych i analiza dotychczasowej szkodowości*. Ubezpieczenie z tytułu zdarzeń medycznych jest nowym rodzajem

ubezpieczenia, funkcjonującym w Polsce od 1 stycznia 2012 r. Przeznaczone jest ono dla podmiotów zajmujących się działalnością leczniczą. Kształt tego ubezpieczenia wzorowany był na modelu szwedzkim, który zakłada odpowiedzialność niezależną od winy osób udzielających świadczeń opieki zdrowotnej.

W trzecim dniu konferencji odbyły się dwie **sesje równoległe**, którym przewodniczyli prof. dr hab. Bronisław Ceranka (Uniwersytet Przyrodniczy w Poznaniu) oraz dr hab. Grażyna Dehnel.

Dr Katarzyna Bolonek-Lasoń i prof. dr hab. Piotr Kosiński (Uniwersytet Łódzki) w referacie *Równowaga Nasha przy strategiach bez ograniczeń* rozważali podstawową grę kwantową (tzw. grę Eiserta-Wilkensa-Levensteina) zakładając, że gracze nie mają żadnych ograniczeń co do stosowanej strategii. Scharakteryzowano wszystkie stany równowagi Nasha dla takiej gry. Brak stanów Nasha dla dużych wartości parametru splątania można wytłumaczyć istnieniem dla każdej strategii jednego z graczy odpowiedniej „kontrstrategii" drugiego. Celem opracowania było wykazanie, że można wyznaczyć punkty Nasha korzystając z metod matematycznych opisywanych teorią grup.

Dr Dorota Rozmus (Uniwersytet Ekonomiczny w Katowicach) wygłosiła referat *Porównanie dokładności taksonomicznej metody propagacji podobieństwa oraz zagregowanych algorytmów taksonomicznych opartych na idei macierzy współwystąpień*. Stosując metody taksonomiczne w jakimkolwiek zagadnieniu klasyfikacji ważną kwestią jest zapewnienie wysokiej poprawności wyników grupowania. W literaturze wciąż proponowane są nowe rozwiązania, które mają przynieść poprawę dokładności grupowania w stosunku do tradycyjnych metod (np. *k*-średnich, metod hierarchicznych). Przykładem mogą być metody polegające na zastosowaniu podejścia zagregowanego, czyli łączenia wyników uzyskanych w wyniku wielokrotnego grupowania (*cluster ensemble*) oraz taksonomiczna metoda propagacji podobieństwa (*affinity propagation clustering*).

W referacie prof. dra hab. Czesława Domańskiego oraz dr Katarzyny Bolonek-Lasoń, *Uogólnienia rozkładów lambda-Tukeya*, omówiono rozkład lambda (GλD), który jest czteroparametrowym uogólnieniem rodziny rozkładów lambda-Tukeya. W literaturze można odnaleźć wiele metod estymacji parametrów GλD, jednak najbardziej popularną metodą jest metoda momentów zaproponowana przez Ramberga i Schmeisera. Wadą uogólnionego rozkładu lambda jest to, że parametry kształtu określają również skośność rozkładu. Powinny być zatem trzy parametry określające położenie, skalę i skośność oraz dwa parametry określające kształt ogonów rozkładu. Wobec tego w referacie dokonano uogólnienia GλD poprzez wprowadzenie pięcioparametrowego rozkładu lambda (FPLD).

Dr Małgorzata Misztal (Uniwersytet Łódzki) przygotowała referat *Kilka uwag o zastosowaniu metod graficznych w analizie brakujących danych*. Metody graficzne mogą być użyteczne w eksploracji i analizie struktury brakujących warto-

ści w zbiorach danych w celu zidentyfikowania mechanizmów generujących braki danych. Znajomość tych mechanizmów jest ważna z punktu widzenia doboru właściwej metody imputacji brakujących wartości. W opracowaniu pokazano przykłady zastosowania graficznych narzędzi pakietu {*VIM*}, działającego w środowisku *R*, do analizy struktury brakujących wartości w zbiorach danych rzeczywistych i symulacyjnych.

Mgr Dominika Polko i dr hab. Grzegorz Kończak zgłosili referat *O prognozowaniu szeregów czasowych z wykorzystaniem kart kontrolnych*. Metody statystyczne opracowane na potrzeby kontroli jakości produktów z powodzeniem mogą być stosowane w analizie innych zagadnień. Do najczęściej stosowanych narzędzi kontroli jakości należy zaliczyć karty kontrolne. Nieklasyczne ich zastosowanie opisane jest w literaturze przedmiotu (Pawłowski Z. (1969), *Predykcja za pomocą kart kontrolnych Prediction based on control chart*). Prognozy otrzymywane metodą predykcji mają charakter ilościowy lub jakościowy. W referacie przedstawiono propozycję modyfikacji tej metody w celu uzyskania wszystkich prognoz o charakterze ilościowym.

Referat mgr Magdaleny Chmielińskiej (Uniwersytet Ekonomiczny w Katowicach), *O wpływie niedokładności pomiarów na koszty kontroli międzyoperacyjnej*, również związany był z tematyką statystycznej kontroli jakości. Jego celem było zwrócenie uwagi na kształtowanie się kosztów kontroli międzyoperacyjnej, szczególnie w przypadku występowania błędnych pomiarów (karty kontrolnej). Postawiona została hipoteza głosząca, że niedokładność dokonywanych i rejestrowanych na karcie kontrolnej pomiarów przyczynia się do znacznego wzrostu kosztu kontroli międzyoperacyjnej. Rozważania zostały uzupełnione o symulacje komputerowe przeprowadzone w środowisku *R*.

Ostatniej sesji konferencji (plenarnej) przewodniczył dr hab. Grzegorz Kończak (Uniwersytet Ekonomiczny w Katowicach), wysłuchano w niej 3 referatów.

Prof. dr hab. Grażyna Trzpiot (Uniwersytet Ekonomiczny w Katowicach) wygłosiła referat na temat bayesowskiej przestrzennej regresji kwantylowej, w którym przedstawiła bayesowski model regresji przestrzennej. Autorka zamiast przestrzennej funkcji regresji wykorzystującej średnią rozpatrywała przestrzenną regresję kwantylową. Głównym celem opracowania było omówienie semiparametrycznego modelu bayesowskiego wraz z estymacją. Wychodząc bowiem poza tradycyjny opis klasycznej regresji dodatkowe zasoby informacji o zmienności otrzymuje się badając kwantyle. Estymacja kwantylowa w modelu przestrzennym uwydatnia zależności przestrzenne dla różnych fragmentów rozważanych rozkładów.

Prof. dr hab. Bronisław Ceranka oraz dr Małgorzata Graczyk (Uniwersytet Przyrodniczy w Poznaniu) byli współautorami dwóch referatów, w których skupili się na zagadnieniach dotyczących układów wagowych.

W opracowaniu *Optymalne układy wagowe dla estymacji całkowitej masy obiektów* przedstawiono pełną teorię estymacji całkowitej masy obiektów zarówno w sprężynowym, jak i w chemicznym układzie wagowym przy założeniu, że błędy pomiarów dokonywanych w tych układach są ujemnie skorelowane. Podano dolne ograniczenie wariancji estymatorów oraz warunki konieczne i dostateczne, przy spełnieniu których to dolne ograniczenie jest osiągnięte. Referat stanowi zebraną wiedzę na temat tego zagadnienia poszerzoną o metody konstrukcji i przykłady macierzy odpowiednich układów eksperymentalnych.

W referacie *Regularne d-optymalne sprężynowe układy z diagonalną macierzą wariancji błędów* Autorzy omówili z kolei temat wyznaczenia sprężynowego układu wagowego spełniającego kryterium D-optymalności przy założeniu, że błędy pomiarów są nieskorelowane i mają różne wariancje. Opisano warunki wyznaczające układ optymalny oraz postać macierzy układu optymalnego. Ponadto podane zostały metody konstrukcji macierzy układu i przykłady opisujące przedstawioną teorię.

Podczas konferencji odbyło się również posiedzenie Rady Głównej PTS, któremu przewodniczył prof. dr hab. Czesław Domański — prezes Towarzystwa. Na spotkaniu omawiano m.in. przebieg obchodów stulecia PTS w 2012 r. oraz dyskutowano nad harmonogramem zadań Towarzystwa na 2013 r. z okazji obchodów Międzynarodowego Roku Statystyki.

Podsumowania obrad dokonał prof. dr hab. Czesław Domański. Przewodniczący Komitetu Organizacyjnego podziękował wszystkim za udział w konferencji i poinformował, że XXXII międzynarodowa konferencja naukowa *Wielowymiarowa analiza statystyczna,* WAS 2013 (*The 32-nd International Annual Conference on Multivariate Statistical Analysis, MSA 2013*) odbędzie się w Łodzi od 18 do 20 listopada 2013 r. Przewodnictwo Komitetu Organizacyjnego edycji objął prof. dr hab. Czesław Domański, a sekretarzami naukowymi zostali mgr Marta Małecka oraz dr Artur Mikulec. Konferencja będzie wpisywała się w obchody Międzynarodowego Roku Statystyki, Statystyka 2013 (*The International Year of Statistics, Statistics 2013*).

Oprac. **Artur Mikulec, Aleksandra Kupis-Fijałkowska**

Wydawnictwa GUS — marzec 2013 r.

Wśród publikacji wydanych w marcu br. ukazały się tytuły opracowywane z częstotliwością kilkuletnią bądź wydane jednorazowo, są to: „Wyznania religijne. Stowarzyszenia narodowościowe i etniczne w Polsce 2009—2011", „Obszary wiejskie. Powszechny Spis Rolny 2010", „Sytuacja społeczno-ekonomiczna gospodarstw domowych w latach 2000—2011. Zróżnicowanie miasto—wieś" oraz „Mienie gmin i powiatów w latach 2009—2011" — zapraszamy do ich lektury.



Publikacja **„Wyznania religijne. Stowarzyszenia narodowościowe i etniczne w Polsce 2009—2011"** stanowi siódme wydanie, ukazującego się co 3 lata, informatora o wyznaniach religijnych, stowarzyszeniach narodowościowych i etnicznych istniejących i działających na obszarze Polski. Przedstawione w publikacji informacje zebrano za pomocą statystycznej *Ankiety wyznaniowej* (AW), którą GUS kieruje każdego roku do organizacji wyznaniowych.

Podobnie jak poprzednie wydanie, publikacja składa się z dwóch odrębnych części: wyznaniowej i narodowościowej. Poszczególne części zawierają opis struktur organizacyjnych: pierwsza — wspólnot wyznaniowych (kościoły, związki wyznaniowe), druga — zbiorowości narodowościowych i grup etnicznych (stowarzyszenia, związki, towarzystwa). Obydwie części poprzedzono szczegółowym omówieniem problemów metodologicznych statystycznych badań nad wyznaniami religijnymi oraz badań narodowościowych i etnicznych w Polsce. Zaprezentowano w nich m.in. źródła danych o wyznaniach i stowarzyszeniach jako podstawowych jednostkach opisu oraz krótki rys historyczny omawianego tematu. Dalszą część opracowania stanowi indywidualna prezentacja poszczególnych jednostek, obejmująca dwa główne elementy — indeksy oraz karty informacyjne.

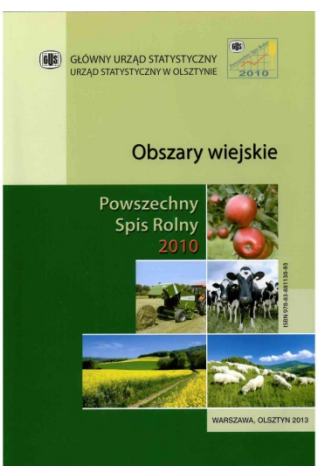
Indeksy zawierające pełen wykaz uwzględnionych w opracowaniu kościołów, związków wyznaniowych i stowarzyszeń umożliwiają Czytelnikowi łatwe wyszukiwanie danych i opisów w następujących porządkach: alfabetycznym, chronologicznym (według daty uznania w Polsce) i demograficznym (według liczby wyznawców lub członków), a w przypadku drugiej części także w układzie etnicznym.

W obydwu częściach publikacji zasadniczą część prezentacji stanowią karty informacyjne, w których zamieszczono charakterystyki poszczególnych wyznań

oraz stowarzyszeń. W części pierwszej znajdujemy przy każdym opisywanym wyznaniu podstawowe dane adresowe, rok powstania i rejestracji, charakterystykę opisową wyznania (m.in. historię, doktrynę, zasady życia religijnego wyznawców, zasady przyjmowania członków, strukturę i działalność związku wyznaniowego) oraz dane statystyczne, tj. liczbę wyznawców, duchownych, jednostek kościelnych, obiektów sprawowania kultu, udzielonych chrztów (lub innych form inicjacji religijnej) czy ślubów.

W części drugiej, zawierającej dane o stowarzyszeniach narodowościowych i etnicznych, podano informacje o adresie siedziby zarządu, przewodniczącym, roku powstania i rejestracji, liczbie członków i zasięgu terytorialnym stowarzyszenia, jednostkach terenowych, wydawnictwach, placówkach kulturalnych, naukowych i oświatowych, przedsięwzięciach i działaniach. Bloki kart informacyjnych zamyka zestawienie krótkich notek o wyznaniach i stowarzyszeniach, które formalnie zaprzestały działalności. Zamieszczony materiał statystyczny wzbogacono ilustracjami graficznymi w postaci tablic, wykresów oraz map.

Publikacja w polskiej wersji językowej (przedmowa oraz spis treści w języku angielskim) dostępna również na płycie CD oraz na stronie internetowej GUS.



Kolejne opracowanie z serii prezentującej wyniki Powszechnego Spisu Rolnego 2010 pt. „**Obszary wiejskie. Powszechny Spis Rolny 2010”** przedstawia kompleksową analizę przemian, jakie zaszły na obszarach wiejskich na poziomie krajowym i regionalnym w latach 2002 i 2010. W niniejszej publikacji wyniki Powszechnego Spisu Rolnego 2010 dla obszarów wiejskich zaprezentowano na tle informacji dotyczących sytuacji społeczno-demograficznej oraz środowiska naturalnego w zakresie dotychczas przedstawianym w opracowaniu *Obszary wiejskie w Polsce*.

Tegoroczna edycja składa się z wersji książkowej oraz stanowiącej integralną część publikacji — płyty CD z tablicami statystycznymi. Pierwsza z nich zawiera obszerny komentarz analityczny, poprzedzony uwagami metodycznymi, w którym Czytelnicy znajdą omówienie źródeł i zakresu prezentowanych danych oraz opis typologii regionów według Eurostatu oraz Organizacji Współpracy Gospodarczej i Rozwoju (OECD).

Komentarz analityczny stanowiący opis wybranych aspektów rozwoju obszarów wiejskich opracowano przede wszystkim na podstawie wyników Powszechnego Spisu Rolnego 2010 oraz danych odnoszących się do finansowania obszarów wiejskich. Zagadnienia te zaprezentowano w powiązaniu z informacjami społeczno-demograficznymi oraz z zakresu rolnictwa i środowiska naturalnego,

ze szczególnym uwzględnieniem zmian dokonujących się na obszarach wiejskich po przystąpieniu Polski do Unii Europejskiej. Uzupełnieniem tych informacji są tablice przeglądowe obejmujące szerokie spektrum zagadnień: od danych demograficznych, działalności pozarolniczej, budżetu i wyposażenia gospodarstw domowych, dostępu do placówek edukacyjnych, ośrodków kultury i opieki zdrowotnej, pomocy społecznej po dane o sytuacji mieszkaniowej, infrastrukturze komunalnej i ochronie środowiska. Zgromadzone dane przedstawiono w ujęciu retrospektywnym dla lat 2002 i 2010, a także w porównaniu z danymi ogólnopolskimi oraz z danymi charakteryzującymi obszary miejskie o liczbie ludności do 5 tys. Wybraną część danych liczbowych dotyczących obszarów wiejskich zilustrowano również na mapach i wykresach. Ponadto opracowanie zawiera aneks z mapami przedstawiającymi najważniejsze zagadnienia społeczno-gospodarcze na poziomie gmin i podregionów.

Płyta CD zawiera opis ważniejszych definicji zastosowanych w opracowaniu oraz tablice z danymi na poziomie województw i podregionów. Prezentowane w tablicach informacje dotyczące obszarów wiejskich wyodrębniono m.in. na podstawie podziału terytorialnego według Krajowego Rejestru Urzędowego Podziału Terytorialnego Kraju (TERYT).

Podział publikacji na część drukowaną i elektroniczną umożliwił udostępnienie Czytelnikom szerokiego zakresu danych ukazujących najważniejsze przemiany i potencjał tkwiący w obszarach wiejskich.

Publikacja w wersji polsko-angielskiej, dostępna również na stronie internetowej GUS.



Wydana jednorazowo publikacja **„Sytuacja społeczno-ekonomiczna gospodarstw domowych w latach 2000—2011. Zróżnicowanie miasto—wieś”** jest opracowaniem prezentującym zróżnicowanie sytuacji społeczno-ekonomicznej gospodarstw domowych zamieszkujących miasta i obszary wiejskie w latach 2000—2011. Publikację przygotowano na podstawie danych pochodzących z badania budżetów gospodarstw domowych, dostosowanego do zaleceń Eurostatu oraz Sytemu Rachunków Narodowych. Badanie to spełnia istotną rolę w analizowaniu warunków życia ludności, gdyż dostarcza cennych informacji na temat sytuacji budżetowej oraz innych ważnych aspektów jakości życia określonej grupy mieszkańców.

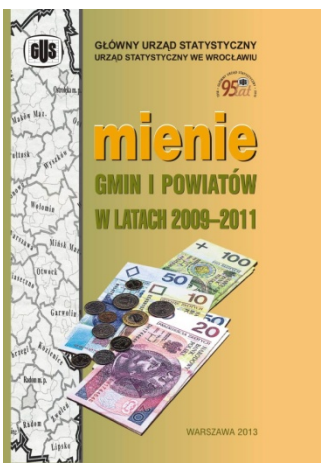
Opracowanie składa się z komentarza analitycznego oraz części tabelarycznej, poprzedzonych opisem metodologii badania. W uwagach metodycznych Czytelnicy znajdą omówienie podmiotu i przedmiotu badania oraz jego organizacji,

metod i narzędzi badawczych, sposobu uogólnienia wyników oraz wyjaśnienie podstawowych pojęć, definicji oraz kategorii ekonomicznych stosowanych w badaniu budżetów gospodarstw domowych.

W części analitycznej przedstawiono analizę uzyskanych wyników w następujących blokach tematycznych: „Charakterystyka gospodarstw domowych", „Dochody i wydatki gospodarstw domowych", „Spożycie produktów żywieniowych", „Warunki mieszkaniowe", „Wyposażenie w dobra trwałego użytkowania" oraz „Subiektywna ocena własnej sytuacji materialnej gospodarstw domowych". W przypadku niektórych zagadnień analizę rozszerzono w zakresie poszczególnych klas miast lub za punkt odniesienia dla obszarów wiejskich przyjęto, obok miast ogółem, największe ośrodki miejskie liczące co najmniej 500 tys. mieszkańców. Ponadto wybraną część danych liczbowych zilustrowano na wykresach oraz mapach.

Bogatym uzupełnieniem komentarza analitycznego jest część tabelaryczna zawierająca zbiór szczegółowych danych liczbowych odnoszących się do wyżej wymienionych zagadnień w podziale na miasta 100—199, 200—499, 500 i więcej tysięcy mieszkańców oraz wieś. Zakończenie publikacji stanowi słownik podstawowych terminów stosowanych w badaniu budżetów gospodarstw.

Publikacja w polskiej wersji językowej (wstęp, spis treści oraz słownik terminów w języku angielskim) dostępna również na płycie CD oraz na stronie internetowej GUS.



W marcu br. ukazała się piąta edycja — opracowywanej co trzy lata — publikacji **„Mienie gmin i powiatów w latach 2009—2011"** przygotowanej przez Urząd Statystyczny we Wrocławiu oraz Departament Badań Regionalnych i Środowiska GUS. Celem opracowania jest kompleksowa prezentacja struktury i tendencji rozwojowych w zakresie majątkowych aspektów funkcjonowania organów samorządu terytorialnego szczebla gminnego i powiatowego z uwzględnieniem elementów, które służą realizacji zadań własnych tych jednostek. Układ tematyczny oraz forma prezentacji danych pozostają niezmienione w porównaniu do wydania poprzedniego z 2009 r.

Publikacja składa się z krótkiej prezentacji uwarunkowań prawnych w zakresie mienia gmin i powiatów, uwag metodologicznych oraz części tabelarycznej. W rozdziale metodologicznym omówiono podstawowe definicje i pojęcia stosowane w sprawozdawczości GUS. Informacje o finansach samorządów zaprezentowano na podstawie sprawozdań Ministerstwa Finansów, a dane odnoszące się do edukacji pochodzą z Systemu Informacji Oświatowej. Gminne i powiato-

we samorządowe jednostki organizacyjne zarejestrowane w rejestrze REGON ujęto według wybranych sekcji Polskiej Klasyfikacji Działalności (PKD 2007).

Część tabelaryczną opracowania rozpoczynają tablice przeglądowe przedstawiające podstawowe dane o mieniu gmin i powiatów w latach 2000, 2005 oraz 2009—2011 w przekroju województw. Szczegółowy zbiór danych statystycznych, charakteryzujących stan majątkowy jednostek samorządu terytorialnego, zaprezentowano w 25 tablicach tematycznych w układzie według województw, powiatów i gmin, zgodnie z podziałem terytorialnym kraju według stanu na dzień 1 stycznia 2011 r. Prezentowane w publikacji wskaźniki umożliwiają porównanie stopnia zaspokajania potrzeb mieszkańców przez jednostki samorządu terytorialnego gminy i powiatu, które są realizowane poprzez zadania wynikające z ustaw kompetencyjnych, w tym w szczególności w zakresie: ochrony zdrowia, opieki społecznej, edukacji publicznej, kultury, infrastruktury komunalnej, utrzymania czystości i porządku, wysypisk i unieszkodliwiania odpadów komunalnych.

Publikacja dostępna jest wyłącznie na stronie internetowej GUS. Wszystkie tablice w wersji elektronicznej przygotowano w formacie MS Excel, co ułatwi Czytelnikom przeprowadzanie dodatkowych analiz i porównań.

Oprac. **Justyna Wójtowicz**

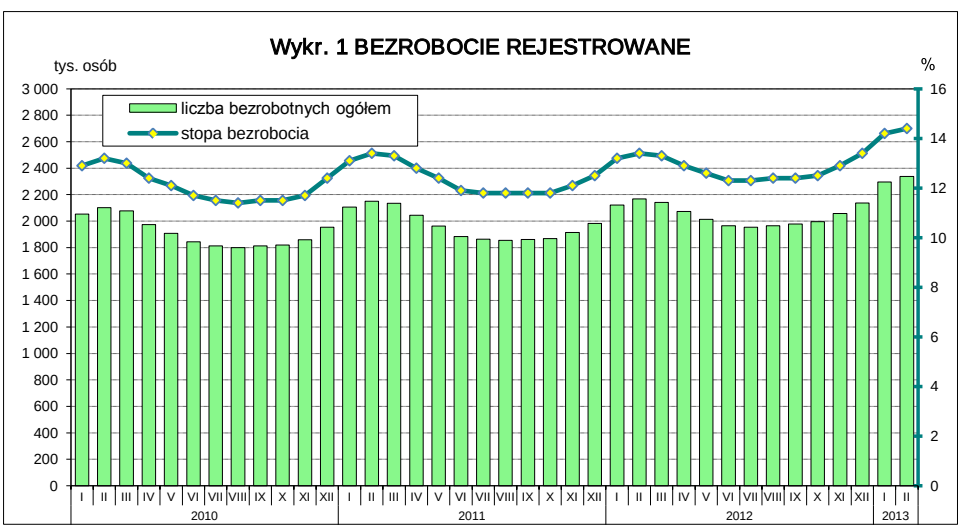
W marcu 2013 r. — poza omówionymi nowościami wydawniczymi — ukazały się następujące tytuły opracowane przez GUS: **„Bezrobocie rejestrowane. I—IV kwartał 2012 r.", „Biuletyn Statystyczny nr 2/2013", „Budownictwo mieszkaniowe I—IV kwartał 2012 r.", „Ceny w gospodarce narodowej — luty 2013 r.", „Ceny robót budowlano-montażowych i obiektów budowlanych — styczeń 2013 r.", „Informacja o sytuacji społeczno-gospodarczej kraju — luty 2013 r.", „Informator GUS 2013", „Kapitał ludzki w Polsce w 2013 r.", Produkcja ważniejszych wyrobów przemysłowych — II 2013 r.", „Rocznik Statystyczny Rolnictwa 2012", i „Wiadomości Statystyczne nr 3/2013".**

Informacja o sytuacji społeczno-gospodarczej kraju — luty 2013 r.

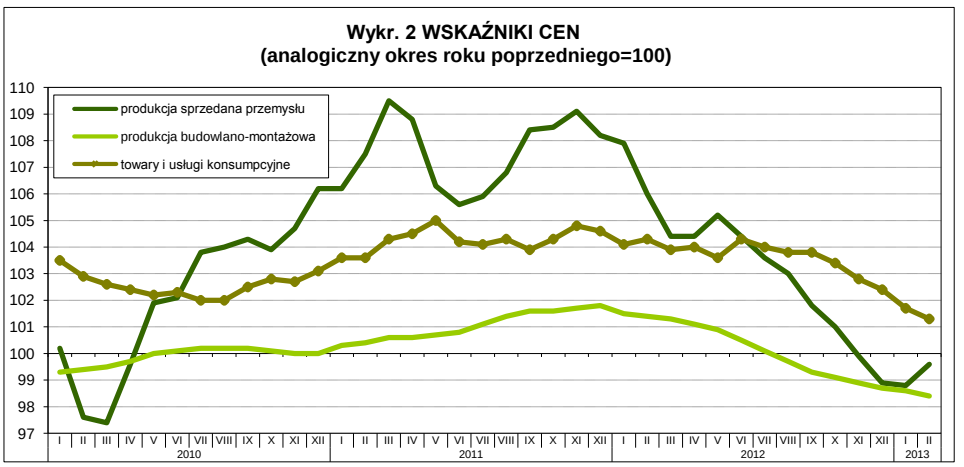
W lutym br. w większości obszarów gospodarki obserwowano tendencje wskazujące na utrzymywanie się spowolnienia gospodarczego. Po niewielkim wzroście w styczniu br., produkcja sprzedana przemysłu ponownie ukształtowała się poniżej poziomu ubiegłorocznego. W budownictwie notowano dalszy, chociaż mniejszy niż w dwóch poprzednich miesiącach, spadek produkcji. Niższa niż przed rokiem była sprzedaż detaliczna. Wolniej niż w styczniu br. wzrosła sprzedaż usług w transporcie.

Dalszemu spowolnieniu uległo tempo wzrostu cen towarów i usług konsumpcyjnych. Przeciętne nominalne wynagrodzenia brutto w sektorze przedsiębiorstw wzrosły bardziej niż przed miesiącem, a ich siła nabywcza, po raz pierwszy od kilku miesięcy, zwiększyła się w skali roku. Dynamika nominalnych i realnych emerytur i rent w obydwu systemach była wyższa niż wynagrodzeń.

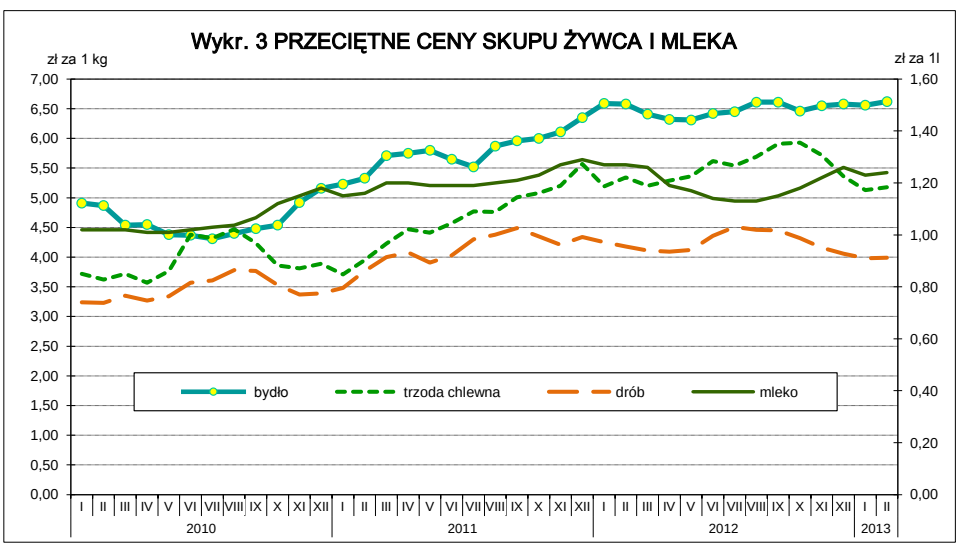
Utrzymuje się trudna sytuacja na rynku pracy. W czwartym miesiącu z kolei przeciętne zatrudnienie w sektorze przedsiębiorstw kształtowało się poniżej poziomu sprzed roku. Spadek notowano m.in. w wytwarzaniu i zaopatrywaniu w energię elektryczną, gaz, parę wodną i gorącą wodę, budownictwie oraz przetwórstwie przemysłowym. Nadal wyższe niż przed rokiem było zatrudnienie m.in. w administrowaniu i działalności wspierającej oraz działalności profesjonalnej, naukowej i technicznej. Stopa bezrobocia rejestrowanego zwiększyła się do 14,4% (wykr. 1).



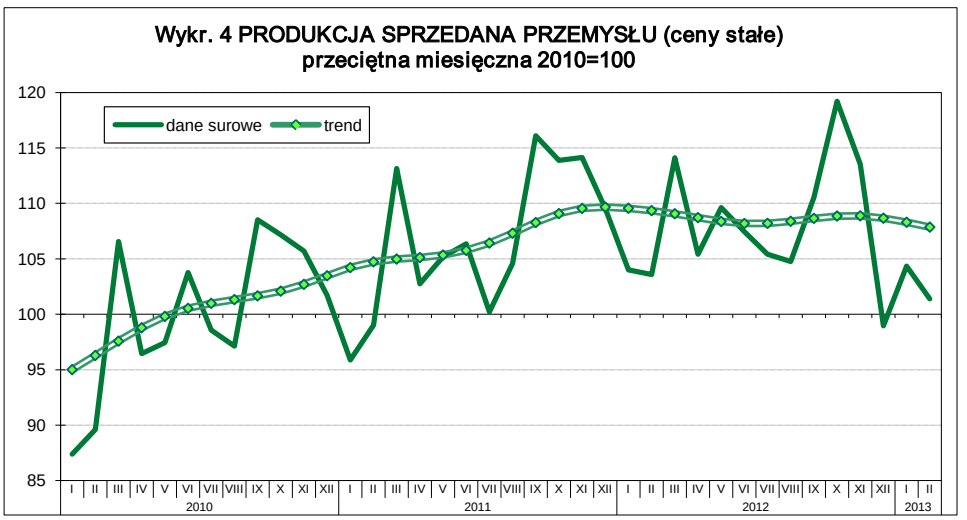
Tempo wzrostu cen towarów i usług konsumpcyjnych zwolniło w lutym br. do 1,3% w skali roku. Wpłynęła na to m.in. słabsza niż w poprzednich miesiącach dynamika cen żywności i napojów bezalkoholowych oraz cen towarów i usług związanych z mieszkaniem (wykr. 2). Nadal obserwowano spadek cen producentów zarówno w przemyśle, jak i w budownictwie.

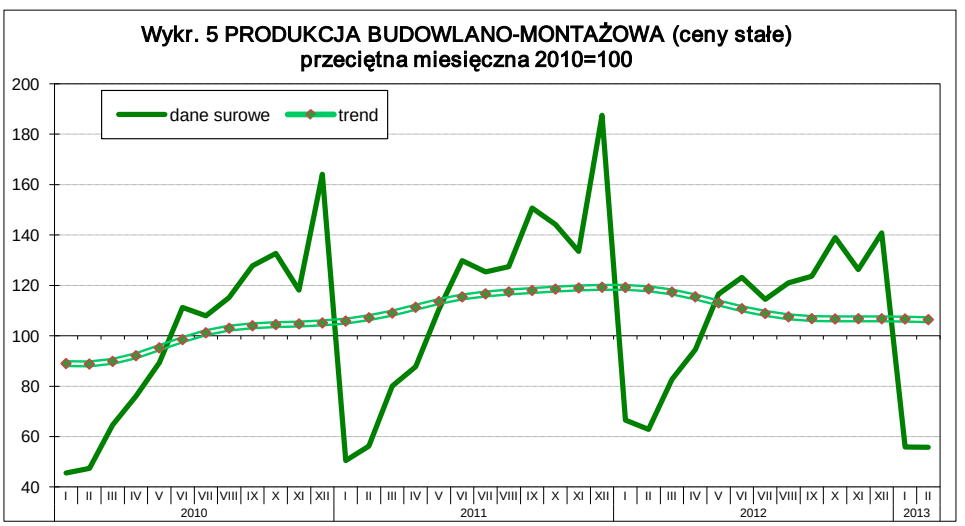


Na rynku rolnym w lutym br. ceny skupu większości produktów pochodzenia zwierzęcego kształtowały się na poziomie niższym niż przed rokiem (wykr. 3). Znacznie wyższe były ceny pszenicy, przy dalszym spadku cen żyta.



Produkcja sprzedana przemysłu w lutym br. była o 2,1% niższa niż przed rokiem, a po wyeliminowaniu czynników o charakterze sezonowym — o 2,4% (wykr. 4). Spadek obserwowano we wszystkich sekcjach przemysłu, z wyjątkiem górnictwa i wydobywania. Wśród głównych grupowań przemysłowych wzrost produkcji sprzedanej notowano jedynie w zakresie dóbr konsumpcyjnych (w tym szybszy nietrwałych niż trwałych). W największym stopniu obniżyła się produkcja dóbr związanych z energią oraz inwestycyjnych. Produkcja budowlano-montażowa była niższa niż przed rokiem o 11,4% (po wyeliminowaniu czynników o charakterze sezonowym o 12,2%) (wykr. 5). Spadek sprzedaży wystąpił we wszystkich działach budownictwa, najgłębszy w jednostkach wykonujących głównie roboty budowlane specjalistyczne.





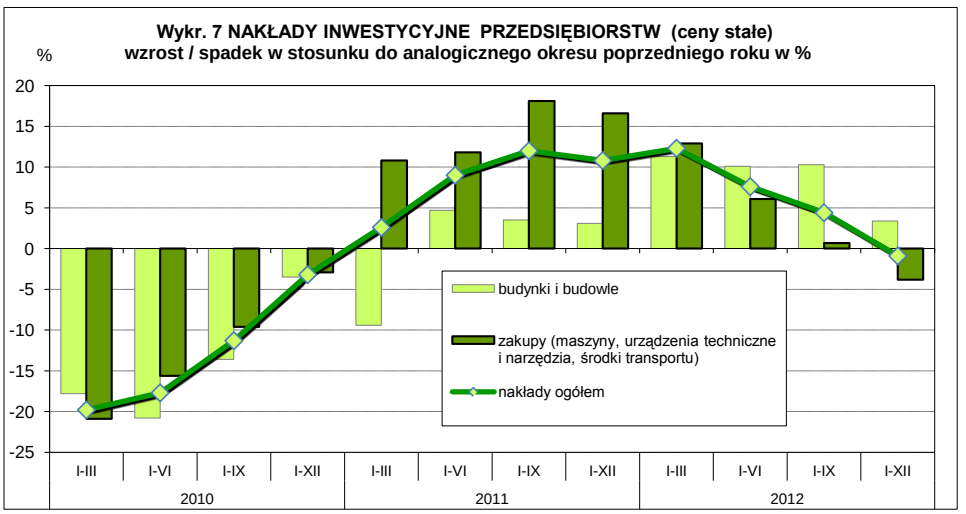
W marcu br. nieco poprawiły się negatywne oceny ogólnego klimatu koniunktury w przetwórstwie przemysłowym. Mniej niekorzystne niż przed miesiącem są bieżące wskazania w zakresie portfela zamówień ogółem. Bardziej pozytywne są przewidywania dotyczące produkcji i portfela zamówień (głównie krajowego). Podmioty budowlane również oceniają koniunkturę nieco mniej negatywnie niż w lutym br., na co wpływa poprawa niekorzystnych ocen bieżących oraz przyszłych w zakresie portfela zamówień i produkcji. W obydwu sekcjach poprawiły się negatywne przewidywania dotyczące sytuacji finansowej, przy utrzymujących się niekorzystnych ocenach bieżących. Nieco mniej pesymistyczne niż przed miesiącem są nastroje jednostek handlu detalicznego, w tym ich prognozy dotyczące sytuacji gospodarczej, sprzedaży, popytu na towary oraz sytuacji finansowej.

Sytuacja finansowa przedsiębiorstw niefinansowych w 2012 r. była trudniejsza niż przed rokiem. Niższy był wynik na działalności gospodarczej oraz wyniki finansowe brutto i netto. W rezultacie wolniejszego wzrostu przychodów z całokształtu działalności niż kosztów ich uzyskania pogorszył się wskaźnik poziomu kosztów (z 94,6 do 95,8%) (wykr. 6). Osłabieniu uległy podstawowe relacje ekonomiczno-finansowe przedsiębiorstw — wskaźniki rentowności brutto i netto zmniejszyły się o ponad 1 p.proc. (odpowiednio do 4,2 oraz 3,4%). Obniżyły się wysokie wskaźniki płynności I i II stopnia. Zmniejszył się udział przedsiębiorstw wykazujących zysk netto w ogólnej liczbie przedsiębiorstw. Relacje ekonomiczno-finansowe eksporterów również były gorsze niż przed rokiem, ale nadal kształtowały się korzystniej niż dla ogółu badanych przedsiębiorstw.



Po wysokim wzroście przed rokiem, w 2012 r. notowano niewielki spadek (w cenach stałych) nakładów inwestycyjnych badanych przedsiębiorstw (o 0,9% w skali roku). Niższe niż przed rokiem były nakłady na zakupy, przy wzroście

nakładów na budynki i budowle (wykr. 7). Słabszą niż przeciętnie dynamikę nakładów wykazały podmioty z kapitałem zagranicznym. Zmniejszyła się liczba inwestycji rozpoczętych ogółem oraz (znacznie) ich wartość kosztorysowa.



Po dwóch miesiącach br. budżet państwa zamknął się deficytem w wysokości 21,7 mld zł, co stanowiło 60,9% kwoty założonej w ustawie budżetowej na 2013 r. Dochody wyniosły 42,8 mld zł, a wydatki — 64,5 mld zł (odpowiednio 14,3% oraz 19,3% planu).

Departament Analiz i Opracowań Zbiorczych

SPIS TREŚCI

STUDIA METODOLOGICZNE

Henryk Gurgul, *Marcin Suder* — Rozkład prawdopodobieństwa dziennych wypłat z bankomatów **1**

Tomasz Górecki, Ewa Łaźniewska — Funkcjonalna analiza składowych głównych PKB **23**

Dominik Śliwicki, Marek Ręklewski — Ekonometryczna analiza czynników aktywności zawodowej kobiet i mężczyzn **35**

BADANIA I ANALIZY

Małgorzata Juchniewicz, Urszula Tomczyk — Kapitał intelektualny sektora przedsiębiorstw na tle produktu krajowego brutto według województw **47**

Mirosław Gorczyca — Wydatki na mieszkanie w budżetach gospodarstw domowych **60**

STATYSTYKA REGIONALNA

Sylwia Kosznik-Biernacka — Analiza wyników matur z języka polskiego w wybranych województwach **66**

Z PRAC RADY STATYSTYKI

Antoni Żurawicz — Działalność Rady Statystyki w II półroczu 2012 r. **82**

INFORMACJE. PRZEGLĄDY. RECENZJE

XXXI międzynarodowa konferencja naukowa *Wielowymiarowa analiza statystyczna* (oprac. *Artur Mikulec*, *Aleksandra Kupis-Fijałkowska*) **88**

Wydawnictwa GUS (marzec 2013 r.) (oprac. *Justyna Wójtowicz*) **100**

Informacja o sytuacji społeczno-gospodarczej kraju — luty 2013 r. (oprac. *Departament Analiz i Opracowań Zbiorczych*, *GUS*) **105**

CONTENTS

METHODOLOGICAL STUDIES

Henryk Gurgul, *Marcin Suder* — Probability distribution of daily withdrawals from ATMs **1**

Tomasz Górecki, Ewa Łaźniewska — Functional analysis of the main components of GDP **23**

Dominik Śliwicki, Marek Ręklewski — Econometric factor analysis of the women and men economic activities **35**

SURVEYS AND ANALYSES

Małgorzata Juchniewicz, Urszula Tomczyk — Intellectual capital of the business sector against the gross domestic product by voivodships **47**

Mirosław Gorczyca — Expenditure on housing in household budgets **60**

REGIONAL STATISTICS

Sylwia Kosznik-Biernacka — Analysis of the matriculation exam results of the Polish language in selected voivodships **66**

INFORMATION SOCIETY

Antoni Żurawicz — The activity of the Statistical Council in the second half of 2012 ..... **82**

INFORMATION. REVIEWS. COMMENTS

XXXI international scientific conference *Multivariate statistical analysis* (by *Artur Mikulec*, *Aleksandra Kupis-Fijałkowska*) **88**

CSO Publications (March) (by *Justyna Wójtowicz)* **100**

Information on the socio-economic situation of Poland in February 2013 (by *Aggregated Studies Department, CSO*) **105**

TABLE DES MATIÉRES

ÉTUDES MÉTHODOLOGIQUES

Henryk Gurgul, *Marcin Suder* — Distribution de probabilité des retraits journaliers aux distributeurs **1**

Tomasz Górecki, *Ewa Łaźniewska* — Analyse en composantes principales du PIB **23**

Dominik Śliwicki, *Marek Ręklewski* — Analyse économétrique des facteurs de l'activité professionnelle des femmes et des hommes **35**

ÉTUDES ET ANALYSES

Małgorzata Juchniewicz, *Urszula Tomczyk* — Capital intellectuel relatif au secteur d'entreprises par le produit intérieur brut selon les voievodies **47**

Mirosław Gorczyca — Dépenses de logement relatives aux budgets des ménages **60**

STATISTIQUES RÉGIONALES

Sylwia Kosznik-Biernacka — Analyse des résultats de l'examen du baccalauréat de la langue polonaise relative aux voievodies choisies **66**

DES TRAVAUX DU CONSEIL DE LA STATISTIQUE

Antoni Żurawicz — Activité du Conseil de la Statistique relative au II semestre 2012 ..... **82**

INFORMATIONS. REVUES. COMPTE-RENDUS

XXXI Conférence scientifique Internationale „Analyse statistique multidimensionnelle" (par *Artur Mikulec*, *Aleksandra Kupis-Fijałkowska*) ... **88**

Publications du GUS (mars 2013) (par *Justyna Wójtowicz*) **100**

Information sur la situation socio-économique du pays — février 2013 (par *Département d'Analyses et d'Élaboration Agrégées*, *GUS*) **105**

СОДЕРЖАНИЕ

МЕТОДОЛОГИЧЕСКИЕ ИЗУЧЕНИЯ

Хенрик Гургул, Марцин Судэр — Распределение вероятностей ежедневных выплат денег через банкоматы **1**

Томаш Гурецки, Эва Лазьневска — Функциональный анализ основных компонентов ВВП **23**

Доминик Сливицки, Марек Ренклевски — Эконометрический анализ факторов профессиональной активности женщин и мужчин **35**

ОБСЛЕДОВАНИЯ И АНАЛИЗЫ

Малгожата Юхневич, Уршуля Томчик — Интеллектуальный капитал сектора предприятий на фоне валового внутреннего продукта по воеводствам **47**

Мирослав Горчица — Расходы на жилье в бюджетах домашних хозяйств **60**

РЕГИОНАЛЬНАЯ СТАТИСТИКА

Сильвия Кошник-Бернацка — Анализ результатов экзамена на аттестат зрелости по польскому языку в избранных воеводствах **66**

ИЗ ДЕЯТЕЛЬНОСТИ СОВЕТА СТАТИСТИКИ

Антони Журавич — Деятельность Совета статистики во второй половине 2012 г. **82**

ИНФОРМАЦИИ. ОБЗОРЫ. РЕЦЕНЗИИ

XXXI международная научная конференция *Многомерный статистический анализ* (разраб. *Артур Микулец, Александра Купис-Фиялковска*) **88**

Публикации ЦСУ (март 2013 г.) (разраб. *Юстина Вуйтович*) **100**

Информация о социально-экономическом положении страны — февраль 2013 год (разраб. *Отдел анализа и сводных разработок, ЦСУ*) **105**

Do Autorów

Szanowni Państwo!

- W „Wiadomościach Statystycznych" publikowane są artykuły poświęcone teorii i praktyce statystycznej, omawiające metody i wyniki badań prowadzonych przez GUS oraz przez inne instytucje w kraju i za granicą, jak również zastosowanie informatyki w statystyce oraz zmiany w systemie zbierania i udostępniania informacji statystycznej. Zamieszczane są też materiały dotyczące zastosowania w kraju metodologicznych i klasyfikacyjnych standardów międzynarodowych oraz informacje o działalności organów statystycznych i Polskiego Towarzystwa Statystycznego, a także o rozwoju myśli statystycznej i kształceniu statystycznym.

- Artykuły proponowane do opublikowania w „Wiadomościach Statystycznych" powinny zawierać oryginalne opisy zjawisk oraz autorskie wnioski i sugestie dotyczące rozwoju badań i analiz statystycznych. Dla zwiększenia właściwego odbioru nadsyłanych tekstów Autorzy powinni wyraźnie określić cel opracowania artykułu oraz jasno przedstawić wyniki, a w przypadku prezentacji przeprowadzonych badań — opisać zastosowaną metodę i osiągnięte wyniki. Przy prezentacji nowych metod analizy konieczne jest podanie przykładów ich zastosowania w praktyce statystycznej.

- Artykuły zamieszczane w „Wiadomościach Statystycznych" powinny wyrażać opinie własne Autorów. Autorzy ponoszą odpowiedzialność za treść zgłaszanych do publikacji artykułów. W razie zastrzeżeń ze strony czytelników w sprawie tych treści Autorzy zostają zobligowani do merytorycznej odpowiedzi na łamach miesięcznika.

- Po wstępnej ocenie przez Redakcję „Wiadomości Statystycznych" tematyki artykułu pod względem zgodności z profilem czasopisma, artykuły mające charakter naukowy przekazywane są dwóm niezależnym, zewnętrznym recenzentom specjalizującym się w poszczególnych dziedzinach statystyki, którzy w swojej decyzji kierują się kryterium oryginalności i jakości opracowania, w tym treści i formy, a także potencjalnego zainteresowania czytelników. Recenzje są opracowywane na drukach zaakceptowanych przez Kolegium Redakcyjne „Wiadomości Statystycznych". Recenzenci są zobowiązani do poświadczenia (na karcie recenzji) braku konfliktu interesów z Autorem. Wybór recenzentów jest poufny.

- Lista recenzentów oceniających artykuły w danym roku jest publikowana w pierwszym numerze elektronicznej wersji czasopisma.

- Autorzy artykułów, którzy otrzymali pozytywne recenzje, wprowadzają zasugerowane przez recenzentów poprawki i dostarczają redakcji zaktualizowaną wersję opracowania. Autorzy poświadczają w piśmie uwzględnienie wszystkich poprawek. Jeśli zaistnieje 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.

- Decyzję o publikacji artykułu podejmuje Kolegium Redakcyjne „Wiadomości Statystycznych". Podstawą tej decyzji jest szczegółowa dyskusja poświęcona omówieniu zgłoszonych przez Autorów artykułów, w której uwzględniane są opinie przedstawione w recenzjach wraz z rekomendacją ich opublikowania.

- Redakcja „Wiadomości Statystycznych" przestrzega zasady nietolerowania przejawów nierzetelności naukowej autorów artykułów polegającej na:
  a) nieujawnianiu współautorów, mimo że wnieśli oni istotny wkład w powstanie artykułu, określanemu w języku angielskim terminem „ghostwriting";
  b) 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 (wzór oświadczenia zamieszczono na stronie internetowej) dotyczące:
  a) stwierdzenia, że zgłoszony artykuł jest własnym dziełem i nie narusza praw autorskich osób trzecich,
  b) wykazania wkładu w powstanie artykułu przez poszczególnych współautorów,
  c) poinformowania, że zgłoszony artykuł nie był dotychczas publikowany i nie został złożony w innym wydawnictwie.

  Główną odpowiedzialność za rzetelność przekazanych informacji, łącznie z informacją na temat wkładu poszczególnych współautorów w powstanie artykułu, ponosi zgłaszający artykuł.

- Artykuły opublikowane są dostępne w wersji elektronicznej na stronie internetowej czasopisma.

- Wersję pierwotną czasopisma stanowi wersja elektroniczna.

Redakcja zastrzega sobie prawo dokonywania w artykułach zmian tytułów, skrótów i przeredagowania tekstu i tablic, bez naruszenia zasadniczej myśli Autora.

Informacje ogólne

- Artykuły należy dostarczać pocztą elektroniczną (lub na płycie CD). Prosimy również o przesłanie dwóch egzemplarzy jednostronnego wydruku tekstu na adres:

a.swiderska@stat.gov.pl lub e.grabowska@stat.gov.pl
Redakcja „Wiadomości Statystycznych"
Główny Urząd Statystyczny
al. Niepodległości 208, 00-925 Warszawa

- Konieczne jest dołączenie do artykułu skróconej informacji (streszczenia) o jego treści (ok. 10 wierszy) w języku polskim i, jeżeli jest to możliwe, także w językach angielskim i rosyjskim. Streszczenie powinno być utrzymane w formie bezosobowej i zawierać: ogólny opis przedmiotu artykułu, określenie celu badania, przyjętą metodologię badania oraz ważniejsze wnioski.
- Pytania dotyczące przesłanego artykułu, co do jego aktualnego statusu itp., należy kierować do redakcji na adres: a.swiderska@stat.gov.pl lub e.grabowska@stat.gov.pl lub tel. 22 608-32-25.
- Korespondencję do redaktora naczelnego należy kierować na adres t.walczak@stat.gov.pl.

Wymogi edytorskie wydawnictwa

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

1. Edytor tekstu — Microsoft Word, format *.doc lub *.docx.
2. Czcionka:
   - autor — Arial, wersalik, wyrównanie do lewej, 12 pkt.,
   - tytuł opracowania — Arial, wyśrodkowany, 16 pkt.,
   - tytuły rozdziałów i podrozdziałów — Times New Roman, wyśrodkowany, kursywa, 14 pkt.,
   - tekst główny — Times New Roman, normalny, wyjustowany, 12 pkt.,
   - przypisy — Times New Roman, 10 pkt.
3. Marginesy przy formacie strony A4 — 2,5 cm z każdej strony.
4. Odstęp między wierszami półtorej linii oraz interlinia przed tytułami rozdziałów.
5. Pierwszy wiersz akapitu wcięty o 0,4 cm, enter na końcu akapitu.
6. Wyszczególnianie rozmaitych kategorii należy zacząć od kropek, a numerowanie od cyfr arabskich.
7. Strony powinny być ponumerowane automatycznie.
8. Wykresy powinny być załączone w osobnym pliku w oryginalnej formie (Excel lub Corel), tak aby można było je modyfikować przy opracowaniu edytorskim tekstu. W tekście należy zaznaczyć miejsce ich włączenia. Należy także przekazać dane, na podstawie których powstały wykresy.
9. Tablice należy zamieszczać w tekście, zgodnie z treścią artykułu. W tablicach nie należy stosować rastrów, cieniowania, pogrubiania czy też podwójnych linii itp.
10. Pod wykresami i tablicami należy podać informacje dotyczące źródła opracowania.
11. Stosowane są skróty: tablica — tabl., wykres — wykr.
12. Przypisy do tekstu należy umieszczać na dole strony.
13. Przytaczane w treści artykułu pozycje literatury przedmiotu należy zamieszczać podając nazwisko autora i rok wydania publikacji według wzoru: (Kowalski, 2002). Z kolei przytaczane z podaniem stron pozycje literatury przedmiotu należy zamieszczać w przypisie dolnym według wzoru: Kowalski (2002), s. 50—58.
14. Wykaz literatury należy zamieszczać na końcu opracowania według porządku alfabetycznego według wzoru: Kowalski J. (2002), *Tytuł publikacji*, Wydawnictwo X, Warszawa (bez podawania numerów stron). Literatura powinna obejmować wyłącznie pozycje przytoczone w artykule.