

POLSKA AKADEMIA NAUK  
KOMITET STATYSTYKI I EKONOMETRII

# PRZEGLĄD STATYSTYCZNY

STATISTICAL REVIEW

NUMER SPECJALNY

**2**

2012

WARSZAWA 2012

## WYDAWCA

Komitet Statystyki i Ekonometrii Polskiej Akademii Nauk

## RADA REDAKCYJNA

Andrzej S. Barczak, Marek Gruszczyński, Krzysztof Jajuga, Stanisław M. Kot, Tadeusz Kufel,  
Władysław Milo (Przewodniczący), Jacek Osiewalski, Aleksander Welfe, Janusz Wywiał

## KOMITET REDAKCYJNY

Magdalena Osińska (Redaktor Naczelny)  
Marek Walesiak (Zastępca Redaktora Naczelnego)  
Piotr Fiszeder (Sekretarz Naukowy)  
Michał Majsterek (Redaktor)  
Anna Pajor (Redaktor)

Wydanie publikacji dofinansowane przez Ministra Nauki i Szkolnictwa Wyższego

© Copyright by Komitet Statystyki i Ekonometrii PAN

Strona WWW „Przegląd Statystyczny”:  
<http://www.przegladstatystyczny.pan.pl>

Nakład: 250 egz.

PAN Warszawska Drukarnia Naukowa  
00-056 Warszawa, ul. Śniadeckich 8  
Tel./fax: +48 22 628 76 14



**Z okazji jubileuszu 100-lecia  
Polskiego Towarzystwa Statystycznego**

**Honorowy Patronat Prezydenta Rzeczypospolitej Polskiej  
Bronisława Komorowskiego**

*Publikacja dofinansowana ze środków  
Narodowego Banku Polskiego*

**NBP**

N a r o d o w y   B a n k   P o l s k i



JAN PARADYSZ

## STATYSTYKA REGIONALNA: STAN, PROBLEMY I KIERUNKI ROZWOJU

### 1. DEFINICJA STATYSTYKI REGIONALNEJ

Na użytek niniejszego opracowania określimy sobie statystykę regionalną jako ilościowy opis tożsamości administracyjnych i funkcjonalnych części danego kraju zwanych regionami. Istotą statystyki regionalnej jest to, że stanowi ona immanentną część statystyki ogólnokrajowej, jako hierarchicznego układu geograficzno-infrastrukturalnych powiązań gospodarczych, demograficznych i społecznych. Przestrzeń w statystyce regionalnej odgrywa kluczową rolę, ale rozróżnienie statystyki regionalnej od lokalnej jest już dyskusyjne. Dyskusyjne bywa też pojęcie regionu, na co wskazuje się we wstępie do europejskiego systemu statystyki regionalnej i miejskiej podkreślając rolę takich cech specyficznych jak krajobraz, klimat, języki mieszkańców, ich pochodzenie etniczne czy przeszłość historyczna, por. Eurostat (2009, s. 2). Na początku lat siedemdziesiątych w państwach Wspólnoty Europejskiej rozpoczęto prace nad jednolitym i koherentnym systemem podziału poszczególnych krajów na regiony celem opracowywania porównywalnych danych statystycznych. Zasadnicze zręby systemu statystyki regionalnej wypracowywano w ciągu 30 lat w wyniku żmudnych uzgodnień pomiędzy Eurostatem a państwami członkowskimi wspólnoty. Ostatecznie Rozporządzeniem<sup>1</sup> Parlamentu Europejskiego i Rady Unii z dnia 26 maja 2003r. wprowadzono ujednolicony system geokodowania jednostek terytorialnych dla celów statystycznych (Nomenclature d'unités territoriales statistiques – NUTS). Przedmiotem regulacji są NUTS1, NUTS2 i NUTS3, które noszą nazwy makroregionów (w Polsce mamy 6 grup województw), regionów (16 województw) i podregionów (66, do 2007 – 45). Dwa następne stopnie terytorialnego podziału kraju NUTS4 i NUTS5 są zaliczane do poziomu lokalnego (Local Administrative Units – LAU) i na razie nie są objęte regulacją unijną, ale podejmowane są prace także w tym kierunku.

---

<sup>1</sup> Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS) (Official Journal L 154, 21.6.2003). Kolejne rozporządzenia wydawano w latach 2005 i 2008 za każdym rozszerzeniem Unii Europejskiej.

## 2. WŁAŚCIWOŚCI INFORMACJI W UKŁADZIE REGIONALNYM

Niezależnie od pojęcia statystyki regionalnej i lokalnej wydaje się celowym sformułowanie pewnych pożądanych właściwości danych w układzie przestrzennym. Wy różnimy następujące cechy informacji statystycznej: zgodność definicyjna, zgodność statystyki struktur ze statystyką przepływów, hierarchiczność, geokodowalność, teleinformatyczność, aktualność, obciążone niskimi błędami systematycznymi oraz możliwie wysoką precyzją.

Pierwsza z tych właściwości oznacza zgodność definicyjną między pojęciami w statystyce lokalnej, regionalnej i ogólnokrajowej. Nie jest to ani oczywistość ani truizm, jeśli uwzględnimy różnorodność źródeł danych statystycznych, szczególnie w warunkach wykorzystywania zasobów administracyjnych, pozyskiwanych w sposób autonomiczny przez ich różnych gestorów. Szczególnym przypadkiem tej zgodności definicyjnej jest wymóg adekwatności statystyki struktur ze statystyką przepływów. W analizie demograficznej oznacza to, że struktura ludności musi być ujęta według zameldowania na pobyt stały, gdyż ruch naturalny i wędrowniczy jest klasyfikowany w ten sposób w układach przestrzennych. W przypadku statystyki aktywności zawodowej ludności właściwą strukturą będzie tak zwana ludność faktycznie zamieszkała, bo według tej kategorii występują przepływy na rynku pracy w systemie BAEL. Rozkłady ludności według wieku i płci zameldowanych na pobyt stały oraz tak zwani faktycznie zamieszkali różnią się zasadniczo. Dotyczy to szczególnie osób w wieku 20-30 lat, którzy wyróżniają się w tym okresie dużą mobilnością przestrzenną. Ponieważ rozkład migracji, zarówno na pobyt czasowy ponad 3 miesiące jak i tych na stałe, jest funkcją odległości – wraz ze wzrostem odległości spada liczba migrantów – to waga ewentualnej niezgodności definicyjnej rośnie odwrotnie proporcjonalnie do wielkości regionu. Dużo większe ta niezgodność ma znaczenie dla gmin i powiatów niż województw i makroregionów. Dla Polski ogółem obydwa rozkłady są takie same.

Hierarchiczność oznacza, że informacje regionalne są sumą danych lokalnych lub punktem wyjścia do dezagregacji na poziom statystyki lokalnej. Suma informacji regionalnych niższego poziomu agregacji przestrzennej stanowi poziom wyższy. Statystyka ogólnokrajowa jest sumą danych na poziomie makroregionów, czyli NUTS1. NUTS1 z kolei jest sumą informacji wojewódzkich (NUTS2) i tak dalej. Nie wszystkie informacje posiadają zapewnioną własność hierarchiczności w każdym przypadku. Jedynie systemy zasilane od dołu, najczęściej ma to miejsce w przypadku rejestrów administracyjnych oraz sprawozdawczości Głównego Urzędu Statystycznego – na przykład, ewidencja ruchu ludności i pełne spisy – mają automatycznie zapewnioną hierarchiczność. Dane otrzymywane za pomocą badań reprezentacyjnych tylko dla najwyższych poziomów przestrzennego podziału kraju dostarczają informacji, które mogą się sumować<sup>2</sup>.

<sup>2</sup> Na przykład, tak duże próby losowe, z jakimi mamy do czynienia w Badaniu Aktywności Ekonomicznej Ludności (BAEL), budżetach gospodarstw domowych czy ankiety mikroprzedsiębiorstw SP3, dostarczają, i to niewiele informacji, na poziomie województw. Należy, bowiem, wziąć pod uwagę, że im mniejszy region, tym mniej szczegółowe tablice statystyczne mogą być dla niego wyprodukowane. W tych

Geokodowalność informacji statystycznej może występować na poziomie danych jednostkowych i jest cechą bardzo pożądaną we wszystkich rejestrach administracyjnych oraz w układzie jednostek przestrzennych. W Polsce geokodowalność statystyczną zapewnia numer w systemie TERYT. W chwili obecnej dostęp do bazy TERYT jest bezpłatny na stronach GUS. Jeszcze kilka lat temu osoby fizyczne z tego nie mogły korzystać, co utrudniało rozwój innych rejestrów administracyjnych.

Teleinformatyczność oznacza możliwość przesyłu danych statystycznych na odległość. Dotyczy to zbierania danych za pomocą portali sprawozdawczych oraz urządzeń przenośnych typu handheld, za pomocą których przeprowadzano wywiady w ostatnich dwóch spisów w Polsce (PSR 2010 oraz NSP 2011), oraz ich udostępniania (Bank Danych Lokalnych – BDL i baza DEMOGRAFIA w Głównym Urzędzie Statystycznym).

Aktualność danych statystycznych jest często podnoszonym postulatem, który jednak należy widzieć we właściwej perspektywie czasowej i odpowiednich uwarunkowaniach. Na przykład, wyniki ankiety koniunktury gospodarczej wymagają możliwie szybkiej publikacji i wraz z upływem czasu tracą swoją wartość diagnostyczną. Jednakże są takie badania, jak *Dzielnictwo kobiet*, które zostało przeprowadzone w czasie Narodowego Spisu Powszechnego w 1970 i do dzisiaj nie straciło na znaczeniu. Przeciwnie, wraz z upływem lat, gdy poszczególne roczniki kobiet starzały się i odchodziły do wieczności, rzeczony źródło danych nabierało szczególnego znaczenia odnośnie rozwoju rodziny i jej *dzielnictwa* przed 1950 r. na Ziemiach Polskich, które po *Drugiej Wojnie Światowej* znalazły się w obecnych granicach naszego Państwa. Fatalnym zbiciem okoliczności<sup>3</sup> po badaniu *dzielnictwa* w ramach NSP 1970 pozostały dwa tomy tablic statystycznych, które w niewielkim stopniu zostały wykorzystane dla polskiej demografii. W szczególności drugi tom dotyczący ówczesnych województw, które poza trzema na Północy Polski, stosunkowo niewiele różnią się od obecnych, mogłyby być dobrym punktem wyjścia do rekonstrukcji generacji i kohort rzeczywistych w przekrojach regionalnych<sup>4</sup>. Innymi przykładami względności pojęcia aktualności danych statystycznych są kohortowe tablice trwania życia we Francji<sup>5</sup> i w krajach skandynawskich<sup>6</sup> oraz

przypadkach coraz częściej sięga się do informacji wykorzystujących estymację pośrednią. Jednakże nie oznacza to automatycznego spełnienia własności hierarchiczności.

<sup>3</sup> Niestety, zadziałało tutaj fałszywie pojmowane wówczas pojęcie aktualności. Należy zaznaczyć, że względy ekonomiczne spowodowały zarejestrowanie na komputerowych taśmach magnetycznych tylko dzieci 3 pierwszych kolejności urodzeń, które później około 1980r. komisyjnie zniszczono, podobnie jak ponad 412 tysięcy jednostronicowych załączników do kwestionariusza spisuowego, por. Paradyś (1990, s. 27).

<sup>4</sup> Powróćmy do tego problemu pisząc dalej o translacji demograficznej w układzie regionalnym.

<sup>5</sup> Są to tablice zrekonstruowane przez Vallin i Meslé (2012) od początku XIX wieku z projekcją na całe bieżące stulecie.

<sup>6</sup> Por. Kolari (1980) dla Finlandii i SS (2010) dla Szwecji. Jeszcze dalej wstecz w tych krajach sięga rekonstrukcja płodności w generacjach rzeczywistych i hipotetycznych, na przykład dla Finlandii Lutz (1987). Jak widać, informacje o rozwoju ludności mają znaczenie niekiedy nawet przez kilka wieków dla bieżącej oceny sytuacji demograficznej.

opracowywanie niektórych tematów po spisach powszechnych w USA<sup>7</sup>. Nie sądzę też, żeby informacje ekonomiczne były zawsze skazane na „aktualność” a dane sprzed kilku czy kilkunastu lat mogły liczyć tylko na zainteresowanie historii gospodarczej. Jak wskazywałem w jednym z ostatnich moich tekstów, por. Paradysz (2011), w statystyce gospodarczej też mamy do czynienia z problemem analizy kohortowej<sup>8</sup>, który stanie kiedyś na porządku dziennym i pomoże nam objaśnić bieżące zachowania osób na rynku pracy czy może nawet koniunktury ekonomicznej.

Kolejnymi pożądanymi cechami regionalnych informacji statystycznych, z wymienionych na wstępie, są braki obciążenia błędami systematycznymi i wysoka precyzja badań, czyli niskie błędy estymatora. Tak przynajmniej postuluje się w literaturze przedmiotu. Nasze wątpliwości, co do poprawności takiego podejścia budzi fakt, że przedmiotem estymacji w dużych badaniach statystycznych są tabele statystyczne, najczęściej wielowymiarowe, a nie pojedyncze parametry rozkładu (modalna, średnia, mediana i inne kwantyle, a także liczba jednostek wyróżnionych, frakcja czy wartość globalna). Ponadto, jak wynika z naszego doświadczenia z oceną jakości Narodowego Spisu Powszechnego 2002 oraz z przygotowań do NSP 2011, im mniejsze regiony, tym większe są udziały błędów systematycznych i błędów losowych (por. opracowania CSR, 2008; Józefowski, Rynarzewska-Pietrzak, 2010 i Paradysz, 2010). W przypadku błędów systematycznych, mechanizmem ich powstania w NSP 2002 prawdopodobnie była szara w mieszkalnictwie i materialne zainteresowanie rachmistrzów spisaniem możliwie największej liczby gospodarstw domowych niezależnie od stanu prawnego mieszkających w nich osób<sup>9</sup>. Co się tyczy redukcji błędów losowych w badaniach reprezentacyjnych, to nowoczesne techniki estymacji pośredniej, takie jak kalibracja, statystyka małych obszarów, wyrównywanie wielowymiarowe wyników badania oraz benchmarking, pozwolą na otrzymywanie wiarygodnych tabel także w przekrojach regionalnych.

Kolejnym krokiem w kierunku budowy systemu statystyki regionalnej jest szukanie możliwości rekonstrukcji bardziej złożonych jednostek statystycznych oraz agregowalność tematyczna. Jak wynika z doświadczeń międzynarodowych i polskich związanych z nową koncepcją spisów opartych na administracyjnych źródłach danych oraz metodologią estymacji pośredniej<sup>10</sup> przejście od pojęcia osoby do bardziej złożonych struktur, jak rodzina i gospodarstwo domowe, zwielokrotnia zasoby informacji o regionie. Jesz-

<sup>7</sup> W tym ostatnim przypadku bardzo obszerne tomy poświęcone odstępom intergenetycznym, które przebadano w czasie spisu powszechnego w 1970, ukazały się dopiero po upływie 5 lat, por. US (1975).

<sup>8</sup> Demografia jako nauka była rozwijana przez prawie trzy wieki zanim zdano sobie sprawę z niemożności objaśnienia takiego zjawiska jak Baby Boom bez pomocy analizy kohortowej.

<sup>9</sup> Należy zauważyć, że we wszystkich poprzednich spisach ludności aż do NSP 2002 włącznie obowiązywał stan deklarowany przez reprezentanta gospodarstwa domowego, najczęściej kobiety w średnim wieku, bez okazywania odpowiednich dokumentów (dowodów tożsamości, świadectw potwierdzających poziom wykształcenia itp.). Do nielicznych wyjątków należało odnotowywanie przez urząd gminny stanu prawnego posiadanej przez rolników ziemi, co odbywało się bez angażowania respondentów.

<sup>10</sup> Autorem nowej koncepcji spisów PSR 2010 i NSP 2011 jest Dygaszewicz (2007a, 2007b).



cze większym pożytkiem byłaby agregowalność tematyczna, przez którą przedstawimy na przykładzie rynku pracy. Badanie aktywności ekonomicznej ludności bardzo zyskałoby na wartości, gdyby go połączyć z dojazdami do pracy oraz z analizą płac. Niestety, informacja o dojazdach do pracy, która jest pozyskiwana dzięki statystyce podatkowej jest opracowywana w oderwaniu od badania aktywności ekonomicznej ludności bez uwzględniania tak istotnych cech rynku pracy jak wiek i płeć. Postulat analizy płac bądź dochodów w tym kontekście – o ile mi wiadomo – jeszcze się nie pojawiał.

Dalszymi cennymi właściwościami danych statystycznych jest możliwość ich użycia w obu rodzajach analiz, to znaczy w transversalnej i kohortowej. Chodzi o to, żeby na podstawie raportowanych systematycznie, co roku, faktów ekonomicznych i społecznych odtworzyć losy jednostek statystycznych (osoby, rodziny, gospodarstwa domowe, przedsiębiorstwa i instytucje) zgodnie z ich rzeczywistym rozwojem. W demografii oznacza to możliwość rekonstrukcji kohort rzeczywistych w przypadku bieżącej rejestracji zdarzeń demograficznych, ekonomicznych lub społecznych. Znacznie rzadziej spotykany jest drugi typ analizy. Kiedy na podstawie przeprowadzonego badania anamnetycznego, najczęściej za pomocą ankiety reprezentacyjnej, odtwarza się wskaźniki w ujęciu transversalnym, które dotyczą konkretnych lat lub okresów. Tego rodzaju przechodzenie od ujęć poprzecznych do wzdłużnych a niekiedy także na odwrót, daje możliwość wyodrębnienia efektu translacji w układach regionalnych.

### 3. ANALIZA SWOT POLSKIEJ STATYSTYKI REGIONALNEJ

Analiza SWOT nie jest specjalnie wyrafinowanym narzędziem poznania, umożliwia jednak w miarę całościowe ujęcie badanego przedmiotu w postaci czterodzielnej tablicy, gdzie staramy się wypunktować jego mocne i słabe strony oraz szanse i zagrożenia.

Zestawienie w tabeli 1 mocnych i słabych strony statystyki regionalnej oraz ukazanie jej szans i zagrożeń stanowi bilans moich przemysłów i wystąpień publicznych<sup>11</sup>. Stawiając na pierwszym miejscu wśród mocnych stron polskie podejście spraw trudnych chciałem podkreślić nasze bardzo specyficzne zaangażowanie przy tworzeniu statystyki regionalnej przy wychodzenia z socjalizmu<sup>12</sup>. Na początku lat dziewięćdziesiątych w kilku ośrodkach naukowych w Polsce rozpoczęto prace w zakresie utworzenia instytucji zajmujących się statystyką regionalną, które niekiedy konkurowały ze sobą, ale częściej

<sup>11</sup> Zarówno, jako organizatora i dotychczasowego dyrektora Centrum Statystyki Regionalnej, uczestnika debaty publicznej na temat przemian programów kształcenia ekonomicznego po 1989, członka Rady Statystyki przy Prezesie Rady Ministrów w latach 2000-2007 oraz przewodniczącego podgrupy do spraw metod statystyczno-matematycznych w Powszechnym Spisie Rolnym (PSR 2010) i w Narodowym Spisie Powszechnym (NSP 2011).

<sup>12</sup> Trochę to przypominało tworzenie polskiej statystyki przed pierwszą wojną światową, która dała możliwość naszym przodkom stworzyć szybciej Główny Urząd Statystyczny niż powstało Państwo Polskie.

Tabela 1.

## Analiza SWOT polskiej statystyki regionalnej

| Mocne strony |                                               | Słabe strony |                                                                          |
|--------------|-----------------------------------------------|--------------|--------------------------------------------------------------------------|
| 1            | Polskie podejście do spraw trudnych           | 1            | Dziedzictwo PRL                                                          |
| 2            | Taksonomia numeryczna                         | 2            | Przywiązanie do tradycji GUS z czasów II RP                              |
| 3            | Konferencje na Południu Polski i SKAD         | 3            | Niski poziom świadomości statystycznej                                   |
| 4            | Bank Danych Lokalnych                         | 4            | Brak koncepcji informatyzacji kraju                                      |
| 5            | Centrum Statystyki Regionalnej                | 5            | Partykularyzm resortowy                                                  |
| 6            | Estymacja pośrednia i EURAREA                 | 6            | Rozmiary szarej strefy                                                   |
| 7            | PSR 2010 i NSP 2011                           | 7            | Nieudana reforma administracyjna kraju                                   |
| Szanse       |                                               | Zagrożenia   |                                                                          |
| 1            | Akcesja do Unii Europejskiej                  | 1            | Rezygnacja z administracyjnych źródeł danych                             |
| 2            | Rejestry administracyjne                      | 2            | Powrót do estymacji bezpośredniej                                        |
| 3            | Sukces PSR 2010 i NSP 2011                    | 3            | Regres statystyki na rynku edukacyjnym                                   |
| 4            | Portale statystyczne i marketing statystyczny | 4            | Niski poziom przygotowania kadr administracji publicznej                 |
| 5            | Strategie rozwoju lokalnego                   | 5            | Dalszy rozwój patologii społecznych i próby ich ukrycia przed statystyką |
|              |                                               | 6            | Uchronienie niezależności statystyki publicznej                          |

Źródło: opracowanie własne.

współpracowały<sup>13</sup>. W 1995 powstało Centrum Statystyki Regionalnej (CSR). O podobnym charakterze instytucja powstała w Jeleniej Górze<sup>14</sup>, która zajęła się tworzeniem

<sup>13</sup> CSR zostało utworzone w 1995 roku na mocy porozumienia między Prezesem Głównego Urzędu Statystycznego i Rektorem Akademii Ekonomicznej w Poznaniu odpowiadając na zapotrzebowania polskiego społeczeństwa i gospodarki w okresie przemian ustrojowych, w tym w zakresie systemu funkcjonowania statystyki publicznej. Centrum Statystyki Regionalnej skoncentrowało się na metodologii w zakresie statystyki regionalnej i lokalnej ze szczególnym uwzględnieniem nowych technik próbkowania, statystyki małych obszarów, wykorzystania administracyjnych źródeł danych, integracji baz danych i komputerowych systemów informacyjnych. CSR zajmowało się także promocją małych ojczyzn (internetowa panorama miast, gmin i regionów) i szkoleniem w zakresie metody reprezentacyjnej, statystyki małych obszarów i statystyki regionalnej. Wykonywało prognozy demograficzne na niskich poziomach agregacji przestrzennej (miasta i gminy) oraz uczestniczyło w dużych projektach badawczych, jak strategia rozwoju województwa wielkopolskiego w 2000r. oraz w budowaniu strategii i programu operacyjnego rozwoju Polski Zachodniej latach 2010-2011. CSR współuczestniczyło w przygotowywaniu Powszechnego Spisu Rolnego 2010 oraz Narodowego Spisu Powszechnego 2011. W minionych 15 latach CSR kształtowało świadomość statystyczną na poziomie regionalnym i lokalnym przez organizację integracyjnych spotkań środowiskowych z udziałem samorządu lokalnego, statystyki publicznej oraz instytucji badawczych i szkoleniowych z innych uczelni. Zorganizowało 5 międzynarodowych konferencji naukowych na temat statystyki regionalnej (1996, 1998 2000, 2002, 2005).

<sup>14</sup> Wśród wielu cennych opracowań powstałych zarówno w ówczesnym Wojewódzkim Urzędzie Statystycznym w Jeleniej Górze jak na Wydziale zamiejscowym wrocławskiej Akademii Ekonomicznej warto wymienić książki Borysa (2005), Strahl (2003), Markowskiej i Strahl (2009), Walesiaka (2002).

Banku Danych Lokalnych, który został następnie zaimplementowany w Głównym Urzędzie Statystycznym<sup>15</sup>.

Mocną też stroną polskiej statystyki regionalnej stała się estymacja pośrednia, zwana też często statystyką małych obszarów. Zastosowania metod estymacji dla małych obszarów ma już długą historię w Polsce, która się rozpoczęła od międzynarodowej konferencji w Głównym Urzędzie Statystycznym zorganizowanej w 1992 r. z inicjatywy ówczesnego wiceprezesa GUS J. Kordosa<sup>16</sup>. Jemu także zawdzięczamy promocję estymacji pośredniej w Polsce także w późniejszym okresie na łamach założonego przez niego kwartalnika *Statistics in Transition*. Dzięki J. Kordosowi, Polska, jako jeden z 6 krajów europejskich został zaproszony do międzynarodowego konsorcjum EURAREA. Przystępując do prac przygotowawczych PSR 2010 oraz NSP 2011, kierownictwo GUS postanowiło skorzystać z doświadczenia wyniesione z konsorcjum EURAREA. Jednakże pełne wykorzystanie tego dorobku w praktyce statystyki publicznej wymaga pogłębionej analizy wyników empirycznych. W ramach projektu EURAREA, sformułowano następujący pewne ogólne wnioski<sup>17</sup>. Poza oceną standardu 7 estymatorów EURAREA prowadzi się prace w zakresie benchmarkingu (por. Datta, Ghosh, Steorts, Maples, 2009; Paradysz, 2008 oraz Paradysz, Paradysz, 2011, 2012) i kalibracji. Współczesna koncepcja kalibracji danych także znalazła się w obszarze zainteresowań organizatorów ostatnich spisów powszechnych. Idea kalibracji została zaproponowana przez Deville'a i Särndala (1992) i następnie rozwijana w pracach Estevao i Särndal (2000, 2006), Särndal i Lundström (2005) a w Polsce Szymkowiak (2008, 2009, 2009a, 2010) i Żądło (2011). Zaleca się wykorzystanie tych najnowszych metod do kalibracji wyników badania reprezentacyjnego, w szczególności w warunkach braków odpowiedzi oraz zniekształceń próby losowej innymi rodzajami błędów systematycznych. Szczególnego znaczenia kalibracja nabrała w przypadku kombinowanych form spisu ludności opartego na rejestrach administracyjnych (por. Wallgren, Wallgren, 2007 oraz Schulte, Nordholt, 2009).

Do słabych stron polskiej statystyki regionalnej należy, naszym zdaniem, dziedzictwo PRL, kiedy popyt na informację w przekrojach regionalnych był niewielki. Z niewielką dozą przesady można by powiedzieć, że głównym odbiorcą statystyki regionalnej był Centralny Urząd Planowania, który w latach 1949-1989 funkcjonował

<sup>15</sup> Oczywiście, tego rodzaju inicjatywy miały ułatwione zadanie dzięki podbudowie teoretycznej w pracach J. Czekanowskiego, H. Steinhausa i skupionych wokół niego uczniów (m. in. J. Kordos), co zaowocowało aplikacjami w naukach społecznych diagramu Czekanowskiego i taksonomii wrocławskiej. Duży wkład w przechowanie tego dziedzictwa, a nawet twórczego jego rozwoju, miały konferencje Polski Południowej, coroczne konferencje zakopiańskie prof. A. Zeliasia i konferencje taksonometryczne, które z czasem zmieniły nazwę na Sekcję Klasyfikacji i Analizy Danych.

<sup>16</sup> Także w XXI wieku już po przejściu na emeryturę prof. J. Kordos promował metodologię statystyki małych obszarów w Polsce, por. Kordos (2004, 2005, 2006).

<sup>17</sup> Por. <http://www.statistics.gov.uk/eurarea/conclusions.asp> i prowadzi się dalsze dyskusje w ramach biennale SAE (Jyväskylä, 2005; Piza, 2007; Elche, 2009; Trewir, 2011). W przygotowaniu jest konferencja w Poznaniu w 2014r., której gospodarzem będzie Katedra Statystyki Uniwersytetu Ekonomicznego.

jako Państwowa Komisja Planowania Gospodarczego<sup>18</sup>, a po 1997 r. zmienił nazwę na Rządowe Centrum Studiów Strategicznych. Innych użytkowników statystyki regionalnej nie było wielu i zmiany w zakresie jej kształtu były wymuszane głównie przez środowisko naukowe na zasadzie naśladownictwa międzynarodowego. Przynajmniej na początku przemian systemowych trudno się było przebić z postulatami zmiany kategoriami myślenia z tamtej epoki. Dobrym przykładem wydaje się wspomniany wyżej brak agregowalności tematycznej informacji o rynku pracy. Brak zrozumienia, że zatrudnienie, bezrobocie, dojazdy do pracy i wynagrodzenia powinny stanowić jednolity i koherentny system informacji o rynku pracy jest też w pewnym sensie dziedzictwem PRL. Wydaje się nie do pomyślenia by w nowoczesnym systemie informacyjnym o rynku pracy zabrakło danych o wykształceniu (lub dokładniej o przygotowaniu zawodowym), wieku i płci. Tego rodzaju system musi być budowany na rejestrach administracyjnych odpowiednich gestorów (przede wszystkim Ministerstwa Finansów, Zakład Ubezpieczeń Społecznych, Spraw Wewnętrznych, Administracji i Cyfryzacji). Ponadto, porównując tablice ruchu ludności opublikowane w przedwojennym Kwartalniku Statystycznym bądź w ówczesnych małych i dużych rocznikach oraz współczesne wydania roczników demograficznych stwierdzimy, że wiele rozwiązań tabelarycznych pochodzi z okresu międzywojennego, kiedy na skutek niskiej mechanizacji obliczeń, GUS był zmuszony do ograniczania szczegółowości tablic<sup>19</sup>.

Dalsze słabe strony wymieniamy tylko hasłowo, chociaż prasa codzienna dostarcza przykładów na ich istnienie w szerokich obszarach polskiej gospodarki. Niski poziom świadomości statystycznej, brak koncepcji informatyzacji kraju, czy partykularyzm resortowy można by zilustrować na tych samych przykładach. Zapewne najbardziej znaczącym z nich byłoby perypetie związane z systemem informatycznym POLTAX<sup>20</sup>, który do dzisiaj dostarcza zaledwie ułamek procenta potencjalnie możliwych do uzyskania informacji o dochodach polskich obywateli<sup>21</sup>. Podjęte w ostatnich latach prace przy okazji PSR 2010 oraz NSP 2011 nad wykorzystaniem tych jak i innych danych z rejestrów administracyjnych pozwalają jednak żywić nadzieję, że w niedługim czasie staną się one systematycznym źródłem zasilania w publicznej statystyce regionalnej. Nie oceniając całościowo reformy administracyjnej, nasz krytycyzm wynika z tego, że nie wykorzystano wówczas szansy powrotu do podziału terytorialnego kraju sprzed 1975 r.

<sup>18</sup> W województwach istniały ekspozytury regionalne, m.in. pod nazwą Wojewódzkich Komisji Planowania Gospodarczego.

<sup>19</sup> Znamienym przykładem jest tutaj statystyka zawierania małżeństw i rozwodów zamieszczana w rocznikach demograficznych, która w tej postaci uniemożliwia rekonstrukcję kohort i generacji rzeczywistych. Należy zaznaczyć, że w wersji internetowej w bazie DEMOGRAFIA w Banku Danych Lokalnych istnieją współczesne „europejskie” warianty statystyki nowożeńców i rozwiedzionych, jednakże zaczynając się od 1995 lub niekiedy tylko od 2002 roku, nie nadają do rekonstrukcji wcześniej powstałych kohort rzeczywistych.

<sup>20</sup> Wśród licznych artykułów i różnych innych opracowań można wymienić artykuł Nowickiego (1995) oraz interpelacje poselskie.

<sup>21</sup> Jedną z nielicznych prób była oryginalna koncepcja K. Kruszki bezpośredniego wykorzystania rejestrów podatkowych w statystyce dojazdów do pracy, por. Kruszka (2010 – red. )

Co prawda, dzięki oporowi regionalnych polityków, uniknięto gorszej wersji podziału, która nie uwzględniała powstanie województw: kujawsko-pomorskiego, lubuskiego, opolskiego i świętokrzyskiego, ale nie potrafią obronić środkowo-pomorskiego (kaszalińskiego). Autorzy reformy administracyjnej nie potrafili przezwyciężyć lokalnych animozji, na skutek czego niektóre z miejscowości, gmin a nawet powiatów pogranicza dawnych województw zmieniło przynależność. Dlaczego tak mocno podkreślamy problem trwałości granic podziału administracyjnego? Po pierwsze, stabilność podziału terytorialnego jest jednym z elementów infrastruktury statystycznej. Zmieniając granice regionów utrudniamy sobie porównywalność w czasie. Po drugie, traci się możliwość obserwacji wzdłużnej i analizy kohortowej w przekrojach regionalnych na wiele lat<sup>22</sup>. O wadze problemu świadczy fakt, że w dobrze zorganizowanych społeczeństwach podział terytorialny kraju jest dużo trwalszy. Na przykład, francuski podział na gminy, kantony i departamenty sięga końca XVIII wieku i jest względnie stabilny. Pewnym usprawiedliwieniem autorów polskiej reformy terytorialnej jest to, że europejski system statystyki regionalnej był pod koniec XX wieku *in statu nascendi* i nie mogli w pełni korzystać z jego doświadczeń.

Wśród szans polskiej statystyki regionalnej na pierwszym miejscu trzeba wymienić akcesję do Unii Europejskiej. Kierownictwo polskiej statystyki publicznej wielokrotnie z satysfakcją podkreślało swoje dobre przygotowanie do wejścia do Unii Europejskiej i że statystyka, jako pierwszy „koszyk”, uzyskał jej akceptację. Jednakże bez tej akcesji mielibyśmy znacznie gorszą statystykę regionalną, przede wszystkim, w sensie organizacji badań i pozyskiwania nowych źródeł danych<sup>23</sup>.

Wielką szansą polskiej statystyki regionalnej są rejestry administracyjne. Mogą być wykorzystane wszechstronnie i wieloaspektowo<sup>24</sup>. Przez długi czas postulat wykorzy-

<sup>22</sup> Dla odtworzenia płodności w jednej tylko generacji potrzebujemy 30 lat (od 15 do 49 lat życia kobiety). W przypadku umieralności, potrzeba nawet 100 lat, żeby zrekonstruować jedną tylko generację.

<sup>23</sup> Eurostatowi zawdzięczamy poprawę statystyki zawierania małżeństw w rocznych przedziałach według wieku, generacji i stanu cywilnego oraz statystyki miejskich po przez wdrożenie programu Urban Audit. Szerzej na ten temat Kruska (2003), Paradysz (2005) oraz Młodak, Stachowiak (2007) i Stachowiak, Stawikowska (2009). GUS jest także zobowiązany do zasilania europejskiego systemu statystyki regionalnej, por. Eurostat (2009).

<sup>24</sup> Jak trafnie wypunktowała Walburg (2007, s. 232-233) rejestry administracyjne mogą być wykorzystane w sposób bezpośredni i pośredni w następujących formach: 1) Jako bezpośrednie źródło danych do badań obecnie realizowanych w oparciu o formularze statystyczne. Częściowe źródło zasilania badań, redukcja liczby pytań na formularzach statystycznych. Całkowite zastępowanie tradycyjnej metody zbierania danych. Badania oparte na formularzach statystycznych będą uzupełniającym źródłem informacji w dziedzinach, w których rejestry administracyjne nie są prowadzone lub dane dla statystyki nie mogą być pozyskane ze źródeł zewnętrznych. 2) Jako bezpośrednie źródło danych do nowych badań. 3) Jako zmienne dodatkowe wykorzystywane w technice estymacji pośredniej dla małych domen. Nowy rodzaj zastosowania danych administracyjnych. Nowy rodzaj statystyki publicznej charakteryzujący się wykorzystaniem wielu źródeł danych statystycznych i administracyjnych. 4) Jako źródło informacji do uzupełniania brakujących danych z badań statystycznych. 5) Jako narzędzie do kontroli jakości danych z badań statystycznych. 6) Jako operat do badań reprezentacyjnych. 7) Jako źródło danych do aktualizacji statystycznego rejestru podmiotów Bazy Jednostek Statystycznych (BJS).

stania rejestrów administracyjnych był trudny do zrealizowania. Jednym z pierwszych, eksperymentalnych i polskich przedsięwzięć było wykorzystanie danych podatkowych do aktualizacji rejestru podmiotów gospodarczych (baza jednostek statystycznych), por. Dehnel (2009, 2010) i Gołata (2009). Jeśli uwzględnić wyjątkowe trudności wykorzystania danych podatkowych w badaniu przedsiębiorczości, to należy tę pierwszą próbę uznać za udaną a stan polskich rejestrów administracyjnych za wystarczająco dobry. Jednakże za naprawdę poważne prace nad oceną rejestrów do zasilania statystyki publicznej należy uznać początek przygotowań do spisów powszechnych (PSR 2010 oraz NSP 2011)<sup>25</sup>.

Znaczną szansą publicznej statystyki regionalnej są portale statystyczne, z których w pierwszym rzędzie należy wymienić strony GUS. Na stronach GUS znajdujemy przykłady doskonałych baz danych. Bank Danych Lokalnych (BDL) przeszedł długą drogę od mało przyjaznego w odbiorze informatorium o gminach, gdzie na początku bezpłatnie pozwalano skopiować zaledwie około 70 zmiennych dla każdej jednostki terytorialnej. Obecnie jest to nowoczesny system z kreatorem tablic dla wszystkich jednostek podziału terytorialnego zgodnego z układem NUTS. Wszystkie informacje są dostępne bezpłatnie dla każdego z użytkowników Internetu. Drugim przykładem bardzo dobrego źródła danych jest bank danych DEMOGRAFIA. Niestety, obejmuje ona okres od 2002 r. i byłoby dobrze gdyby GUS zechciał uzupełnić to doskonałe informatorium o bieżącej statystyce ruchu ludności i struktur demograficznych także o lata poprzedzające rok 2002<sup>26</sup>. Jeśli chodzi o przestrzenny monitoring statystyczny, to bardzo interesującym przedsięwzięciem ze strony GUS jest Statystyczne Vademecum Samorządowca (SVS). W odróżnieniu od BDL oraz DEMOGRAFII, SVS nastawiony jest na szeroki krąg odbiorców, którzy nie muszą dysponować odpowiednim przygotowaniem fachowym, żeby móc z tych danych korzystać. Trzy powyższe bazy danych są przykładami dobrego marketingu statystycznego, gdzie odpowiednia informacja i w dobrym „opakowaniu” trafia do właściwego odbiorcy.

Dużą szansą unowocześnienia statystyki regionalnej są strategie rozwoju lokalnego i regionalnego oraz operacyjne programy działania. Z reguły wiąże się to z wykonywaniem pogłębionych analiz statystycznych i prognoz przy zaangażowaniu wysoko kwalifikowanego środowiska naukowego. Strategie i plany operacyjne mogą być inspiracją dla statystyki publicznej, a niekiedy na niej wymuszają produkcję odpowiednich informacji. Tutaj warto przytoczyć znowu rejestry administracyjne, jako przykład mo-

<sup>25</sup> Można je uznać za prawdziwą rewolucję w funkcjonowaniu GUS. Przewidując duży opór materii przed wprowadzaniem nowych zasad spisów powszechnych, postanowiliśmy zacząć od oceny jakości poprzedniego NSP 2002, który był przeprowadzony według tradycyjnych metod pozyskiwania danych. Analiza jakości NSP 2002 pokazała duże rozmiary błędów systematycznych w wieku największej mobilności 20-30 lat (por. opracowania CSR, 2008; Józefowski, Rynarzewska-Pietrzak, 2010 i Paradysz, 2010).

<sup>26</sup> Przy okazji warto podnieść problem ewentualnej korekty podstawowych struktur demograficznych według płci i wieku w układach regionalnych, która powinna zostać poprzedzona pogłębioną analizą jakości spisów powszechnych.

dernizacji źródeł zasilania, którymi wcześniej zajęto się przy konstruowaniu strategii rozwoju lokalnego niż podjęto odpowiednie działania w tym zakresie w GUS.

Wśród zagrożeń w analizie SWOT oddzieliliśmy punkty związane z rezygnacją z administracyjnych źródeł danych i powrotu do estymacji bezpośredniej, chociaż w dyskusji naukowej obydwie zagadnienia często się łączy. Faktycznie, dobry stan rejestrów administracyjnych jest czynnikiem gwarantującym sukces w estymacji dla małych obszarów. Jednakże w krajach wysoko rozwiniętych pod względem infrastruktury statystycznej, rejestry administracyjne bezpośrednio dostarczają w zakresie procesów ekonomicznych, społecznych i demograficznych. Ich użyteczność, jako zasób zmiennych wspomagających, jest tu wartością dodaną. Można sobie wyobrazić funkcjonowanie statystyki publicznej opartej na rejestrach, czego przykładem była istniejąca jeszcze przed drugą wojną światową w Polsce statystyka podatkowa na poziomie powiatów. Trudniejsze byłoby funkcjonowanie estymacji pośredniej bez wykorzystywania w niej administracyjnych źródeł danych. Po gigantycznym wysiłku w pracach nad adaptacją rejestrów do systemu statystyki publicznej, jakiego dokonał GUS z okazji przygotowań do PSR 2010 oraz NSP 2011, rezygnacja z nich oraz powrót do estymacji bezpośredniej należałoby uznać za katastrofę narodową.

*Uniwersytet Ekonomiczny w Poznaniu*

#### LITERATURA

- [1] Borys T., (2005-red), *Wskaźniki zrównoważonego rozwoju*, Warszawa-Białystok.
- [2] CSR, (2008), Wielowymiarowa analiza błędów systematycznych w NSP 2002 oraz statystyczna analiza zmiennych wspomagających z NSP 2002 pod wykorzystanie dla małych obszarów. Sformułowanie wymagań funkcjonalnych bazy mikrodanych spisowych pod względem potrzeb estymacji dla małych obszarów. Raport Centrum Statystyki Regionalnej Akademii Ekonomicznej w Poznaniu, Poznań. Maszynopis powielony.
- [3] Datta G.S., Ghosh M., Steorts R., Maples J., (2009), *Bayesian Benchmarking with Applications to Small Area Estimation*, Statistical Research Division U.S. Census Bureau. Washington, Research report series (Statistics #2009-01).
- [4] Dehnel G., (2009), *Rejestr podatkowy w estymacji pośredniej dla małych firm na podstawie badania SP3*, [w:] Statystyka małych obszarów w badaniach podmiotów gospodarczych. Pod red. J. Paradysza. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu. Poznań, s. 27-37.
- [5] Dehnel G., (2010), *Rozwój mikroprzedsiębiorczości w Polsce w świetle estymacji dla małych domen*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu. Poznań.
- [6] Deville J-C., (2000), *Generalized calibration and application to weighting for non-response*, [w:] J. G. Bethlehem i P.G. M van der Heijden (eds.) COMPSTAT – Proceedings in Computational Statistics.
- [7] Deville J-C., Särndal C-E., (1992), *Calibration estimators in survey sampling*, Journal of the American Statistical Association. T. 87, s. 376-382.
- [8] Domański Cz., Pruska K., (2001), *Metody statystyki małych obszarów*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- [9] Dygaszewicz, J., (2007a), *Narodowy Spis Powszechny Ludności i Mieszkań 2011*. Założenia metodyczne. Materiałna posiedzenie Rady Programowej narodowego spisu powszechnego ludności i mieszkań 2011 r. 4 września 2007. Warszawa, sierpień 2007.

- [10] Dygaszewicz, J., (2007b), *Powszechny Spis Rolny 2010*. Założenia metodyczne. Materiałna posiedzenie Rady Programowej powszechnego spisu rolnego 2010 r. 5 września 2007. Warszawa, sierpień 2007.
- [11] Eurostat, (2009), *European Regional and Urban Statistics Reference Guide*, Methodologies and Working papers. KS-RA-09-008-EN.pdf.
- [12] Estevao V.M., Särndal C-E., (2000), *A Functional Form Approach to Calibration*, Journal of Official Statistics, Vol. 16, No. 4.
- [13] Estevao V.M., Särndal C-E., (2006), *Survey Estimates by Calibration on Complex Auxiliary Information*, International Statistical Review, Vol. 74, 127-147.
- [14] Gołata E., (2004), *Estymacja pośrednia bezrobocia na lokalnym rynku pracy*, Wydawnictwo Akademii Ekonomicznej w Poznaniu. Prace habilitacyjne nr 11, Poznań.
- [15] Gołata E., (2009), *Mikroprzedsiębiorstwa w badaniu SP3 oraz dodatkowych źródłach informacji*, [w:] Statystyka małych obszarów w badaniach podmiotów gospodarczych. Pod red. J. Paradysza. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu. Poznań, s. 11-26.
- [16] Józefowski T., Rynarzewska-Pietrzak B., (2010), *Ocena możliwości wykorzystania rejestru PESEL w spisie ludności*, [w:] „Pomiar i informacja w gospodarce”. Zeszyt Naukowy nr 149 pod red. E. Gołaty. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań 2010, s. 67-84.
- [17] Klimanek T., Szymkowiak M., (2011), *Taksonomiczne aspekty estymacji pośredniej uwzględniającej korelację przestrzenną Taksonomia 18: Klasyfikacja i analiza danych – teoria i zastosowania*, Uniwersytet Ekonomiczny we Wrocławiu, s. 228-235.
- [18] Kolari, R., (1980), *Cohort mortality in Finland from 1851*, In Studies No 57 1980. Central Statistical Office of Finland.
- [19] Kordos J., (2004), *Niektóre problemy jakości w statystyce małych obszarów*, [w:] A. Zeliaś, Tradycje i obecne zadania statystyki w Polsce, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków, s. 95-124.
- [20] Kordos J., (2005), *Some Aspects of Small Area Statistics and Data Quality*, “Statistics in Transition”, vol. 7, Number 1, s.102-131
- [21] Kordos J., (2006), *Impact of different factors on research in small area estimation in Poland*, “Statistics in Transition”, Vol. 7 No. 4, s. 863-879.
- [22] Kruszka K., (2003), *URBAN AUDIT II dla krajów kandydujących do UE*. Maszynopis powielony.
- [23] Kruszka K., (2010 - red.), *Dojazdy do pracy w Polsce*, Urząd Statystyczny w Poznaniu, Poznań. [http://www.stat.gov.pl/cps/rde/xbcr/poznan/ASSETS.dojazdy\\_do\\_pracy\\_w\\_polsce.2010.pdf](http://www.stat.gov.pl/cps/rde/xbcr/poznan/ASSETS.dojazdy_do_pracy_w_polsce.2010.pdf)
- [24] Lutz W., (1987), *Finnish fertility since 1722*, Väestöntutkimuslaitos (The Population Research Institute), Helsinki.
- [25] Markowska M., Strahl D., (2009), *Miejsce Polski w europejskiej przestrzeni regionalnej*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- [26] Młodak A., Stachowiak D., (2007), *Rola programu URBAN AUDIT w statystyce europejskiej* [w:] Statystyka regionalna w jednoczącej się Europie, Pod red. J. Paradysza, Internetowa Oficyna Wydawnicza, Centrum Statystyki Regionalnej Poznań, s. 22-42.
- [27] Nowicki W., (1995), *Pieniądże wrzucone w Poltax*, Computerworld Polska, Wydanie: 16/1995, 17 kwietnia 1995.
- [28] Paradysz J., (1990), *Reprodukcja ludności w Polsce*, Studium metodologiczno-poznawcze. Szkoła Główna Planowania i Statystyki, Warszawa.
- [29] Paradysz J., (2005), *Audyt miejski*, [w:] Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych. Pod red. A. Zeliasia. Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków, s. 205-214.
- [30] Paradysz J., (2008), *Kryteria dobroci estymacji dla małych obszarów*, [w:] Statystyka społeczna – dokonania, szanse, perspektywy. Pod red. K. Jakóbika. Biblioteka Wiadomości Statystycznych, tom 57, Główny Urząd Statystyczny, Warszawa, s. 74-84.



- [31] Paradysz J., (2010), *Konieczność estymacji pośredniej w spisach powszechnych*, [w:] „Pomiar i informacja w gospodarce”. Zeszyt Naukowy nr 149 pod red. E. Gołaty. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań, s. 45-66.
- [32] Paradysz J., (2011), *Apologia demografii, czyli istnienie demografii bez ludności*, Wiadomości Statystyczne, Nr 2 (597), Luty 2011 s. 27-36.
- [33] Paradysz J., Paradysz K., (2011), *Benchmarking w statystyce małych obszarów*. „Taksonomia 18. Klasyfikacja i analiza danych – teoria i zastosowanie” pod red. K. Jajugi i M. Walesiaka, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2011, s. 86-93.
- [34] Paradysz J., Paradysz K., (2012), *Obszary bezrobocia w Polsce - problem benchmarkowy*. „Taksonomia 19. Klasyfikacja i analiza danych – teoria i zastosowanie” pod red. K. Jajugi i M. Walesiaka, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław. Särndal C-E., Lundström S., (2005), *Estimation in surveys with nonresponse*. John Wiley and Sons, Chichester.
- [35] Schulte Nordholt E., (2009), *Data Integration Activities on the Way to the Dutch Virtual Census of 2011*. Proceedings of MSP2009 Modernisation of Statistics Production [http://www.scb.se/Grupp/Produkter\\_Tjanster/Kurser/ModernisationWorkshop/final\\_papers/G\\_2\\_registers\\_Nordholt\\_final.pdf](http://www.scb.se/Grupp/Produkter_Tjanster/Kurser/ModernisationWorkshop/final_papers/G_2_registers_Nordholt_final.pdf)
- [36] SS, (2010), *Cohort mortality in Sweden*. Mortality statistics since 1861. Demographic reports 2010:1, Statistics Sweden.
- [37] Stachowiak D., Stawikowska M., (2009), *Sytuacja demograficzna w miastach europejskich na podstawie programu Audyt miast 2006*, Wiadomości Statystyczne, nr 5, s. 87-93.
- [38] Strahl D., (2003), *Benchmarking w badaniach regionalnych*. W: *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, pod red. A. Zelasia, Akademia Ekonomiczna, Kraków, s. 127-141.
- [39] Szymkowiak M., (2008), *Testowanie estymatorów kalibracyjnych dla małych obszarów w oparciu o dane z NSP 2002 oraz z bieżących badań reprezentacyjnych budżetów gospodarstw domowych 2002*. Raport z badań, GUS-US Poznań
- [40] Szymkowiak M., (2009), *Imputacja i kalibracja – nowe możliwości estymacji w badaniach statystycznych z brakami odpowiedzi*, [w:] *Statystyka małych obszarów w badaniach podmiotów gospodarczych*. Zeszyt Naukowy nr 116 pod red. J. Paradysza. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań, s. 90-105.
- [41] Szymkowiak M., (2009a), *Testowanie własności estymatorów kalibracyjnych dla małych obszarów na podstawie danych z NSP 2002*, [w:] „Metody i źródła pozyskiwania informacji w statystyce publicznej”. Zeszyt Naukowy nr 128 pod red. E. Gołaty. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań 2009, s. 46-64.
- [42] Szymkowiak M., (2010), *Konstrukcja estymatorów kalibracyjnych wartości globalnej z uwzględnieniem wektora zmiennych instrumentalnych*, [w:] „Pomiar i informacja w gospodarce”. Zeszyt Naukowy nr 149 pod red. E. Gołaty. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań, s. 122-144.
- [43] US, (1975), *Childspacing and current fertility*, Washington: US Department of Commerce. Bureau of the Census, 1970 Census of Population.
- [44] Walburg, (2007), *Rejestry administracyjne w statystyce publicznej*, [w:] *Statystyka regionalna w jednoczącej się Europie*, Pod red. J. Paradysza, Internetowa Oficyna Wydawnicza, Centrum Statystyki Regionalnej Poznań, s. 226-238.
- [45] Walesiak M., (2011), *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- [46] Wallgren A., Wallgren B., (2007), *Register Based Statistics: Administrative Data for Statistical Purposes*. Hoboken, NJ: John Wiley and Sons.
- [47] Vallin J., Meslé F., (2012), *Tables de mortalité françaises 1806-1997 et projections jusqu' en 2102*. Contenu du CD-Rom. [http://www.ined.fr/cdrom\\_vallin\\_mesle/contenu.htm](http://www.ined.fr/cdrom_vallin_mesle/contenu.htm) (Odczytane w dniu 5 lipca 2012).

- [48] Żądło, T., (2008), *Elementy Statystyki Małych Obszarów z Programem R*, Wydawnictwo AE Katowice, Katowice.
- [49] Żądło T., (2011), *On some calibration estimators of subpopulation total for longitudinal data*, Spatial Econometrics and Regional Economic Analysis (ed. B. Suhecki), Acta Universitatis Lodziensis, Folia Oeconomica, 252, 191-204.

## STATYSTYKA REGIONALNA: STAN, PROBLEMY I KIERUNKI ROZWOJU

### S t r e s z c z e n i e

Artykuł przedstawia stan polskiej statystyki regionalnej z uwzględnieniem wytycznych zawartych w dokumentach EUROSTATu i najnowszych praktyk Głównego Urzędu Statystycznego. Przedmiotem naszej uwagi są źródła zasilania informacyjnego ze szczególnym naciskiem położonym na rejestry administracyjne oraz na estymację dla małych obszarów. Problemy i wyzwania polskiej statystyki regionalnej przedstawiono w formie analizy SWOT. Stwierdzono znaczną poprawę stanu metodologii i źródeł zasilania informacyjnego w ostatnich latach. Aktywne uczestnictwo dużej liczby polskich statystyków w biennale estymacji dla małych obszarów stwarza nadzieję na dalszą poprawę sytuacji w zakresie statystyki regionalnej w Polsce.

**Słowa kluczowe:** statystyka regionalna, estymacja dla małych obszarów, benchmarking statystyczny, estymatory kalibracyjne

## REGIONAL STATISTICS: CURRENT STATUS, PROBLEMS AND DEVELOPMENT DIRECTIONS

### A b s t r a c t

This paper presents the state of Polish regional statistics, taking into account the guidelines contained in the EUROSTAT and the latest practices of the Central Statistical Office. The object of our attention are the sources of information, with particular emphasis on administrative records and the estimation for small areas. Problems and challenges of Polish regional statistics presented in the form of a SWOT analysis. We find a significant improvement in the methodology and sources of information supply in recent years. Active participation by a large number of Polish statisticians in meetings of small area estimation gives hope for further improvement in the regional statistics in Poland.

**Key words:** Regional statistics, small area estimation, statistical benchmarking, calibration estimators

WOJCIECH ROSZKA

## SYSTEM STATYSTYKI PUBLICZNEJ OPARTY NA ZINTEGROWANYCH ŹRÓDLACH DANYCH

### 1. WSTĘP

Wspieranie rozwijającej się gospodarki aktualną, rzetelną i wielowymiarową informacją jest rosnącym wyzwaniem dla służb statystycznych. Postulaty udostępniania wiarygodnych szacunków na niskim poziomie agregacji przestrzennej oznaczają, w dotychczasowym podejściu do badań statystycznych, zwiększanie liczebności próby, co podnosi koszty i czas ich przeprowadzenia. Te trudności, a także zwiększone obciążenie respondentów mogą uniemożliwiać spełnienie żądań odbiorców informacji statystycznych.

Rozwiązaniem problemów kosztowności i czasochłonności badań wydaje się być wykorzystanie rejestrów administracyjnych w systemie statystyki publicznej. Zawierają one informacje o bardzo dużej liczbie jednostek, opisują bardzo wiele sfer życia i aktywności ludności, a organy statystyki publicznej mogą je pozyskać od dysponujących nimi instytucji i urzędów. Dodatkowo, dynamicznie rozwijająca się informatyzacja urzędów i wzrost mocy obliczeniowej komputerów powodują, że przesyłanie i przetwarzanie danych z rejestrów jest nie tylko możliwe, ale również stosunkowo szybkie i tanie. Ponadto podmioty administracji państwowej gromadzące dane do rejestrów są gwarantem rzetelności.

W Narodowym Spisie Powszechnym 2011 (a także Państwowym Spisie Rolnym 2010) wykorzystano informacje z dużej liczby źródeł administracyjnych<sup>1</sup>. Poprzez efektywne wykorzystanie rejestrów, zredukowano koszty spisów, obciążenia społeczne związane z pozyskiwaniem danych, poprawiono bezpieczeństwo przechowywanych informacji jednostkowych, a także zwiększono ich wiarygodność i spójność. Dane pochodzące z wielu źródeł umożliwiają oprócz tego wysokie pokrycie informacyjne dotyczące badanych zagadnień spisowych. Rejestry jako bezpośrednie źródła danych zostały wykorzystane w Polsce po raz pierwszy, tworząc załączek systemu statystycznego opartego na rejestrach (Dygaszewicz 2012; Janczur-Knappek 2012).

Wielozróżdowość baz danych niesie ze sobą jednak pewne trudności, takie jak brak zmiennych mogących służyć jako klucz połączeniowy (lub błędy w takich zmiennych), różne definicje zmiennych zawartych w różnych rejestrach (nawet jeżeli zmienna

<sup>1</sup> W PSR 2010 i NSP 2011 wykorzystano informacje pochodzące z 27 rejestrów należących do 15 gestorów, jak również wykorzystano 3 zbiory danych poza administracyjnych (Janczur-Knappek, 2012).

nazywa się tak samo), różne sposoby wypełniania tych rejestrów, jak i istnienie zduplikowanych rekordów. Stanowi to duże wyzwanie dla badaczy.

Artykuł opisuje nowatorskie podejście do statystyki publicznej opartej na zintegrowanych administracyjnych źródłach danych. Przedstawia zalety połączonych źródeł informacji w stosunku do tradycyjnych systemów statystycznych opartych o badania reprezentacyjne oraz wady związane z różnym pochodzeniem zbiorów. Przedstawiono także koncepcję systemu statystycznego opartego na zintegrowanych źródłach danych oraz omówiono metody stochastyczne stosowane w procesie integracji.

## 2. REJESTRY ADMINISTRACYJNE W SŁUŻBIE STATYSTYKI PUBLICZNEJ

Informacja w systemie statystyki publicznej pozyskiwana jest drogą badań statystycznych, zarówno próbkowych, jak i pełnych. W „klasycznym” podejściu do badania zjawisk społeczno-ekonomicznych punktem wyjścia są potrzeby informacyjne odbiorców. Dążąc do ich zaspokojenia, służby statystyczne projektują badania obejmujące różne zagadnienia, wśród których można wymienić: aktywność ekonomiczną ludności, budżety gospodarstw domowych, czy działalność przedsiębiorstw. W kolejnym etapie następuje utworzenie narzędzia badawczego (kwestionariusza), sporządzany jest operat losowania<sup>2</sup> i na jego podstawie losuje się próbę. Następnie prowadzona jest obserwacja statystyczna. Zebrany materiał statystyczny poddany zostaje kontroli i po odpowiednim przetworzeniu (imputacji braków danych, przeważeniu, edycji danych itp.) jest podstawą szacunków, których wyniki są publikowane w formie tabel i wykresów statystycznych. Publikacje te służą wielu różnym celom, wśród których można wymienić wspomaganie organów rządowych i samorządowych w formułowaniu strategii rozwoju oraz prowadzenia polityki społecznej i gospodarczej.

Problemem w „klasycznym” podejściu jest fakt, że dla różnych celów przeprowadzane są oddzielne badania, przez co opracowania statystyczne obejmują pojedyncze tematy. Uniemożliwia to wielowymiarowe szacunki obejmujące różne zagadnienia społeczno-gospodarcze. Dodatkowo, ograniczenia budżetowe<sup>3</sup> powodują, że liczebność próby w badaniu jest zwykle zbyt niska, by szacunki mogły być dokonywane dla małych jednostek terytorialnych. Może to powodować brak zaspokojenia potrzeb informacyjnych samorządów (np. powiatów) dotyczących szczegółowej informacji o kształtowaniu się zjawisk na ich terytorium. Jednocześnie niewielka próba powoduje trudności w wykryciu i badaniu zjawisk rzadkich w skali kraju, które w skali małej domeny mogą być dużym problemem.

Uwzględnienie potrzeb wynikających z globalizacji gospodarki wymaga połączenia informacji z różnych dziedzin. Badania statystyczne obejmujące szeroki zakres

<sup>2</sup> Do tworzenia operatów wykorzystywane są również rejestry administracyjne.

<sup>3</sup> Badanie Aktywności Ekonomicznej Ludności w 2012 roku obejmujące około 54,7 tys. gospodarstw domowych kosztowało około 41 mln złotych, zaś Badanie Budżetów Gospodarstw Domowych w tym samym okresie aż 58 mln złotych (Program Badań Statystycznych Statystyki Publicznej na 2012 rok, 2011).

merytoryczny analizowanych zagadnień są jednak bardzo kosztowne. Ich przeprowadzenie wiąże się jednocześnie z bardzo dużym obciążeniem respondentów (van der Laan, 2000) i wynikającym z tego wzrostem liczby braków odpowiedzi i odmów wypełnienia kwestionariusza<sup>4</sup>, nawet przy zastosowaniu nowoczesnych, obciążających respondentów w mniejszym stopniu, metod zbierania informacji (CATI<sup>5</sup>, CAWI<sup>6</sup> itp.). Koszty zbierania informacji z długich kwestionariuszy mogłyby również prowadzić do zmniejszenia próby dodatkowo utrudniając wnioskowanie dla małych domen.

Informacja na niskim poziomie agregacji przestrzennej dostępna jest z badań pełnych. W badaniu takim pomiarem objęte są wszystkie jednostki należące do populacji docelowej. Ze względu jednak na zasięg takiego badania, jest ono dużym wyzwaniem zarówno finansowym, jak i organizacyjnym. W przypadku spisu powszechnego, nawet bogate kraje nie mogą sobie pozwolić na powtarzanie go częściej niż raz na kilka lat. W okresach międzyspisowych powstaje luka informacyjna, której badania reprezentacyjne nie są w stanie wypełnić. Zapewnienie precyzyjnej informacji dla małych domen jest wyzwaniem dla organów statystyki publicznej pod względem metodologicznym (np. związane z zastosowaniem metod statystyki małych obszarów).

Należy wyraźnie podkreślić, że wykorzystanie rejestrów administracyjnych może zapewnić uogólnianie wyników na niskim poziomie agregacji z dużą częstotliwością. W art. 13 ust. 1 ustawy z dnia 29 czerwca 1995 roku o statystyce publicznej ustawodawca nakazuje organom administracji rządowej i samorządowej przekazywanie danych administracyjnych służbom statystyki publicznej w terminach i formie każdorazowo wymienianej w programie badań statystycznych statystyki publicznej. Zbiory te opisują jednak pojedyncze zagadnienia, takie jak: bezrobocie rejestrowane, ruch naturalny i wędrowniczy ludności, czy działalność podmiotów gospodarczych, nie dając możliwości dokonywania wielowymiarowych szacunków obejmujących różnorodne relacje i zależności w funkcjonowaniu społeczeństwa, gospodarki i państwa jako całości. Dodatkowo definicje cech zawartych w rejestrach mogą się różnić od przyjętych w systemie statystyki publicznej. Rejestry administracyjne, z definicji, stworzone są do wypełniania zadań publicznych (Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne, Dz.U. Nr 64, poz.565, z późn. zm.), nie zaś bezpośrednio do celów statystycznych.

Bezpośrednie wykorzystanie rejestrów w statystyce publicznej nie może nastąpić w sposób automatyczny. Wynika to z odrębności systemów statystycznych i administracyjnych (por. tabela 1).

Informacje zawarte w rejestrach służą jako podstawa w podejmowaniu decyzji administracyjnych, które wpływają na funkcjonowanie jednostek. Natomiast informacje

<sup>4</sup> Przytoczyć tu można stale zmniejszający się poziom realizacji próby w badaniu Polski Generalny Sondaż Społeczny. W pierwszej edycji badania, w 1992 roku, wynosił 82,4%. W następnych latach ulegał stałemu spadkowi, by w 2008 roku wynosić już zaledwie 51,8% (Cichomski *et al.* 2009).

<sup>5</sup> *Computer-Assisted Telephone Interview* – wywiad wspomagany telefonicznie.

<sup>6</sup> *Computer-Assisted Web Interview* – wspomagany komputerowo wywiad przy pomocy strony www.

Tabela 1.

## System administracyjny a system statystyczny

| Charakterystyka               |           | System administracyjny                                                                                        | System statystyczny                                                                                                              |
|-------------------------------|-----------|---------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| 1. Cel powstania              |           | Podjmowanie decyzji administracyjnych                                                                         | Dokonywanie szacunków i analiz                                                                                                   |
| 2. Zbiorowość                 |           | Wszystkie podmioty (jednostki) prawnie podległe danemu gestorowi                                              | Wszystkie podmioty (jednostki) objęte badaniem statystycznym                                                                     |
| 3. Jednostka                  |           | Pojedynczy element zbiorowości; pobierane dane niezbędne są do podejmowania decyzji administracyjnych         | Pojedynczy element zbiorowości; pozyskiwane informacje są podstawą do szacowania statystyk dotyczących populacji lub jej podgrup |
| 4. Cecha                      | Definicja | Wynika z aktów prawnych – mogą być odrębne dla różnych rejestrów                                              | Wynika z ustaleń organizacji międzynarodowych – porównywalne dla wszystkich badań.                                               |
|                               | Warianty  | Nie muszą być zestandaryzowane                                                                                | Zharmonizowane i porównywalne.                                                                                                   |
| 5. Błędy                      |           | Błędy nielosowe. Brak kontroli statystycznej                                                                  | Błędy losowe i nielosowe. Kontrola statystyczna                                                                                  |
| 6. Użyteczność                |           | Dobre źródło statystyk małych obszarów                                                                        | Jakość i możliwości szczegółowej analizy ograniczone wielkością próby                                                            |
| 7. Terminowość i punktualność |           | Zróznicowane w zależności od źródła. Niektóre bardzo aktualne, inne mniej terminowej niż badania statystyczne | Często mają charakter retrospektywny                                                                                             |
| 8. Dostępność i przejrzystość |           | Wpływ uregulowań prawnych. Możliwe bariery techniczne i instytucjonalne                                       | Bezpośrednia kontrola urzędu statystycznego                                                                                      |
| 9. Porównywalność             |           | Zależy od zmian w czasie definicji administracyjnych                                                          | Bezpośrednia kontrola urzędu statystycznego                                                                                      |

Źródło: opracowanie własne na podstawie (Penneck, 2007)

zawarte w systemach statystycznych służą do analiz, na podstawie których formułowane są wnioski o całej populacji.

W systemach administracyjnych jednostka jest przedmiotem decyzji i działań poszczególnych organów wykonawczych, a baza danych służy do pobrania informacji na temat określonego podmiotu. W systemach statystycznych jednostka jest traktowana raczej jako część zbiorowości, dla której tworzone są pewne informacje agregatowe – jednostka nie leży więc w centrum zainteresowania (wyłączając badania monograficzne).

Definicja cechy w systemie administracyjnym wynika z aktów prawnych i może być różna dla różnych rejestrów. Warianty cechy nie muszą być spójne, ponieważ zwykle system administracyjny nie jest podstawą tworzenia zestawień statystycznych. W systemie statystycznym definicje cech, podobnie jak warianty, są spójne dla wszystkich badań (często wynika to z przyjętych ustaleń organizacji międzynarodowych).

Z punktu widzenia jakości danych, wszystkie dane w rejestrach muszą być zgodne i nie mogą zawierać błędów, jednak nie ma konieczności by zapis danej kategorii był taki sam w każdym rekordzie (np. kod pocztowy pisany z myślnikiem lub bez, pełny zapis nazw ulic lub skrócony itp.). Występujące nieścisłości (np. brak numeru PESEL, czy NIP) nie mają charakteru losowego i zazwyczaj wynikają z awarii systemów kon-

troli (które często występują w konkretnej jednostce administracyjnej). W badaniach statystycznych poszczególne warianty cech muszą być ujednolicone, by możliwe było tworzenie spójnych komunikatów. Prowadzony również w badaniach statystycznych rachunek błędów powoduje, że nieścisłości w wynikach są kontrolowane i podejmowane są działania w celu ich redukcji. Różnice w podejściu do gromadzenia danych administracyjnych i statystycznych mogą powodować rozbieżności w publikowanych komunikatach.

Wysokie pokrycie rejestrów administracyjnych może stanowić podstawę do tworzenia statystyk dla małych obszarów (również jako źródło pomocnicze dla estymacji pośredniej). W przypadku systemów statystycznych, możliwości agregacji szacunków są w dużej mierze ograniczone przez wielkość próby (pokrycie rzadko przekracza 1% populacji). Również terminowość rejestrów może być większa od badań statystycznych (mają zwykle charakter retrospektywny). Duża część zbiorów będących w posiadaniu organów publicznych jest uzupełniana na bieżąco i komunikaty tworzone na ich podstawie mogłyby być publikowane z częstotliwością miesięczną (a nawet i większą).

Dostępność i porównywalność danych publikowanych przez organy statystyki publicznej jest ściśle kontrolowana, co wynika w dużej mierze z uregulowań organizacji międzynarodowych. W przypadku systemów administracyjnych, duży wpływ mają regulacje prawne i zmiany definicji administracyjnych w czasie. Przystosowanie poszczególnych rejestrów do konkretnych celów publicznych może tworzyć również problemy techniczne i instytucjonalne – architektura baz i hurtowni danych może być bardzo odmienna, a zapisy prawne, zwłaszcza dotyczące ochrony danych osobowych, mogą zniechęcać instytucje do udostępniania swoich repozytoriów.

Należy również zaznaczyć, że wielu gestorów rejestrów posiada własne służby statystyczne (np. Zakład Ubezpieczeń Społecznych, Narodowy Bank Polski) tworzące sprawozdania na podstawie danych administracyjnych. Publikowane na ich podstawie informacje są przeznaczone jednak na potrzeby tych instytucji i nie uwzględniają potrzeb innych odbiorców.

Natomiast system statystyki oparty na zintegrowanych źródłach tworzony jest poprzez łączenie baz danych w taki sposób, by możliwa była łączna obserwacja zmiennych ze wszystkich integrowanych baz danych dla wszystkich jednostek.

Wallgren i Wallgren (2007) jako podstawowe zalety systemu statystyki opartego na zintegrowanych źródłach wymieniają przede wszystkim: niskie koszty i szybkość pozyskania danych, wysoki stopień pokrycia, jak i bogatą zawartość informacyjną rejestrów (por. tabela 2). Przeciwdziałają im jednak pewne wady, wśród których wyszczególniają: niższą niż w tradycyjnych badaniach jakość danych, opóźnienia w przekazywaniu danych od gestorów wynikające często z ograniczeń prawnych, brak kontroli nad poprawnością statystyczną na etapie zbierania informacji, a także, a wręcz przede wszystkim, niebezpieczeństwo ujawnienia informacji osobowych, co mogłoby doprowadzić do utraty zaufania społecznego dla wykorzystania rejestrów w sprawozdawczości statystycznej.

Tabela 2.

Zalety i wady wykorzystania rejestrów administracyjnych w statystyce publicznej

| Zalety                          | Wady                                                                                                       |
|---------------------------------|------------------------------------------------------------------------------------------------------------|
| Niskie koszty pozyskania danych | Jakość wyników jest gorsza niż w badaniach tradycyjnych                                                    |
| Szybkość pozyskiwania danych    | Opóźnienia w przekazywaniu rejestrów od gestorów do organów statystyki publicznej                          |
| Wysokie pokrycie informacyjne   | Różne definicje zmiennych, problemy z kompatybilnością danych, brak kontroli nad poprawnością statystyczną |
| Pokrycie całej populacji        | Niebezpieczeństwo ujawnienia wrażliwych danych osobowych                                                   |

Źródło: opracowanie własne na podstawie (Wallgren, Wallgren, 2007).

### 3. ISTOTA INTEGRACJI ADMINISTRACYJNYCH BAZ DANYCH NA POTRZEBY SPRAWOZDAWCZOŚCI STATYSTYCZNEJ

Pionierem w wykorzystaniu źródeł administracyjnych w systemie statystyki publicznej były Finlandia (Statistics Finland, 2004) oraz Norwegia (Tonder, 2008). Już w spisie w 1970 wykorzystano w tych krajach rejestr ludności, stopniowo wprowadzając w kolejnych falach spisowych kolejne rejestry. W 1981 roku do tych krajów dołączyła Dania (Borchsenius, 2000), która od razu wprowadziła wszystkie dostępne rejestry w system spisowy (Finlandia wprowadziła spis w pełni oparty o rejestry dopiero w 1990, a Norwegia w 2011). Innymi krajami wykorzystującymi administracyjne źródła danych w systemie spisowym są m.in. Austria (Statistics Austria, 2008), Australia (Ralphs, Tutton, 2011), Holandia (Nordholt, 2004), Izrael (Kamen, 2005), Kanada (Ballano, 2009), Nowa Zelandia (Bycroft, 2011), Stany Zjednoczone (Prevost, Leggieri, 1999), Szwajcaria (Swiss Federal Statistical Office, 2008) oraz Szwecja (Bruhn, 2001; Wallgren, Wallgren, 2007).

Należy zauważyć, że we wszystkich tych krajach, rejestry administracyjne zostały wykorzystane wyłącznie w spisach powszechnych. Problemy z bieżącym dostępem do zbiorów (głównie prawne oraz techniczne) powodują, że możliwe jest tylko ich okresowe wykorzystanie w sprawozdawczości.

Rejestry przeznaczone do integracji mogą się od siebie różnić, co wynika m.in. z odmienności przechowywania danych (różne architektury baz danych, por. Dygaszewicz, 2010), różnych celów, dla których poszczególne rejestry są tworzone, niezgodności momentów referencyjnych (dnia, na który rejestr jest aktualny), czy odmienności w definicjach zmiennych. Dodatkowo, populacje, definicje, czy warianty poszczególnych cech mogą odbiegać od przyjętych w systemie statystyki publicznej. Istnieje zatem potrzeba przetworzenia rejestrów administracyjnych w taki sposób, by odpowiadały potrzebom badań statystycznych i mogły zostać włączone w system badań statystycznych.

Gill (2001) oszacował, że czas potrzebny na przeprowadzenie całego procesu integracji można podzielić w następujący sposób:



- 75% czasu pracy – przygotowanie zbiorów,
- 5% czasu pracy – przeprowadzenie łączenia rekordów w bazach,
- 20% czasu pracy – sprawdzanie poprawności wyników.

Przygotowanie zbiorów do integracji oraz weryfikacja poprawności połączenia to praca w dużej mierze manualna, wymagająca odpowiedniej wiedzy dotyczącej zarówno algorytmu integracji, jak również dziedzin, z których integrowane bazy zawierają informacje (np. definicje zmiennych, specyfikę populacji itp.). Przeprowadzanie łączenia zwykle wykonywane jest za pomocą odpowiednich programów komputerowych. Stąd duża dysproporcja między czasem potrzebnym na przeprowadzenie poszczególnych etapów łączenia.

Pierwszym krokiem integracji jest zebranie informacji na temat danych źródłowych (por. rysunek 2). Może to zostać osiągnięte poprzez zapoznanie się z dostępną dokumentacją, a także poprzez kontakt z gestorami poszczególnych zbiorów danych, niekoniecznie mając dostęp do plików. Informacje zdobyte w ten sposób umożliwiają oszacowanie na ile przeznaczone do łączenia zbiory spełniają wymogi formalne, merytoryczne i jakościowe procesu integracji.

Nowozelandzki Urząd Statystyczny (Data Integration Manual, 2006), korzystając z wieloletniego doświadczenia w integracji, sporządził listę trzech punktów składających się na wstępne zbieranie informacji o zbiorach:

1. Zestawienie istniejącej wiedzy wewnętrznej

W przypadku, gdy te same lub podobne zbiory były używane wcześniej, wiedza i doświadczenie, w postaci dokumentów lub procesów, pracowników biorących udział przy ich łączeniu może być bardzo pomocna. Wraz ze wzrostem liczby integrowanych zbiorów, wypracowanie wewnętrznego systemu gromadzenia wiedzy i dzielenia się nią prowadzi do zwiększenia efektywności pracy nad kolejnymi projektami.

2. Przegląd ogólnie dostępnych informacji

Bardzo często potrzebne i użyteczne informacje na temat integrowanych zbiorów znajdują się na stronach internetowych poszczególnych gestorów. W wielu przypadkach również dostępne są też bardziej szczegółowe informacje w postaci tzw. „często zadawanych pytań” (*Frequently Asked Questions, FAQ*) lub udostępnionych słowników danych<sup>7</sup>.

3. Spotkanie z dostawcami danych

Spotkanie z dostawcami danych umożliwia efektywny transfer wiedzy od ludzi, którzy z określonymi repozytoriami pracują na co dzień. Podczas takich spotkań zadawane pytania oraz przekazywane dokumenty bardzo często w szybki sposób rozwiewają wątpliwości, które mogły się wyłonić podczas pracy nad danymi.

---

<sup>7</sup> Centralny element systemu zarządzania bazą danych, w którym przechowuje się m.in. opisy relacji i perspektyw, deklaracje kluczy głównych, grup użytkowników i uprawnień, informacje o indeksach, plikach i ich strukturach (Adamczewski, 2005).

Identyfikacja populacji docelowej, a także należących do niej jednostek statystycznych, w obu zbiorach, jest ważnym punktem zbierania informacji o danych źródłowych, gdyż może się zdarzyć, że pokrycie populacji jest różne w każdej z baz. Każde repozytorium danych odnosi się do pewnej populacji docelowej, będącej populacją teoretyczną, o której informacje baza powinna zawierać. Zaś poszczególne rekordy danych zawierają informacje o populacji rzeczywistej. Celem integracji jest stworzenie nowego, bogatszego, zbioru danych, który również będzie odnosił się do pewnej populacji docelowej, którą również należy zdefiniować.

Należy również zweryfikować zgodność typów jednostek w obu zbiorach (np. czy bazy zawierają dane o obywatelach, gospodarstwach domowych, czy przedsiębiorstwach) oraz wykonanie wszelkich potrzebnych transformacji danych. Jednostki z baz źródłowych mogą się różnić od jednostek w zintegrowanym zbiorze, jak również różnice mogą wystąpić pomiędzy jednostkami w obu łączonych bazach. Np. w zmiennej „stan cywilny” w jednej bazie może być kategoria „żonaty/zamężna”, zaś w drugiej obie te warianty mogą występować oddzielnie. Przy dużej liczbie zmiennych, przekodowanie wszystkich tak, by zawierały dokładnie te same może okazać się czasochłonnym, ale koniecznym procesem.

Ostatnią częścią etapu zbierania informacji o danych źródłowych jest sporządzenie formularza zawierającego „dane o danych” – tzw. metadanych („informacji o danych”) (Data Integration Manual, 2006). Jest on ważnym elementem, zapewniającym odpowiednią jakość przyszłych wyników, przy czym jakość rozumiana jest sześciowymiarowo, poprzez: przydatność, dokładność, terminowość, dostępność, interpretowalność oraz spójność. Wymienione wymiary jakości mogą być użyte jako kryterium oceny wszystkich szacunków uzyskanych za pomocą zintegrowanych baz danych, odnosząc się zarówno do baz zintegrowanych, jak i źródłowych.

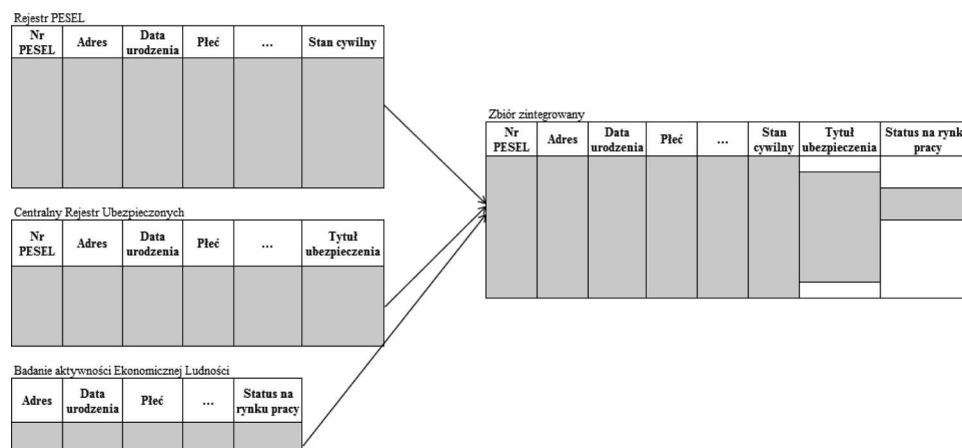
Ważnym zagadnieniem jest również określenie regulacji prawnych umożliwiających wykorzystanie rejestrów w statystyce publicznej. Art. 13 pkt. 4 ustawy z dnia 29 czerwca 1995 r. o statystyce publicznej nakłada na gestorów rejestrów administracyjnych obowiązek nieodpłatnego przekazywania danych, w tym całych zbiorów, w zakresie, formie i terminach określanych w programie badań statystycznych statystyki publicznej. Dane te jednak służą do tworzenia komunikatów „jednowymiarowych”, nie wykorzystujących zalet integracji. Dopiero ustawa o narodowym spisie powszechnym ludności i mieszkań w 2011 roku (art. 8) nakazuje utworzenie tzw. Bazy Danych NSP 2011 złożonej ze zintegrowanych administracyjnych źródeł informacji. Wykorzystanie połączonych rejestrów w okresach międzyspisowych wymaga dodatkowych aktów prawnych.

Efektom etapu badania infrastruktury rejestrów jest wiedza umożliwiająca dokonanie integracji rejestrów, edycji danych i w końcowym etapie tworzenie spójnych i rzetelnych komunikatów statystycznych.

Następnym etapem jest łączenie rejestrów na podstawie unikatowych identyfikatorów jednostek. W źródłach administracyjnych występują zwykle zmienne jak, np. numer PESEL dla osób lub REGON dla podmiotów gospodarczych. W badaniach

reprezentacyjnych jednostkę można zidentyfikować na podstawie zestawu zmiennych, takich jak wiek, płeć i adres zamieszkania<sup>8</sup>. Są to tzw. unikalne klucze połączeniowe. Na ich podstawie możliwe jest jednoznaczne (deterministyczne) wskazanie, które rekordy w łączonych bazach odnoszą się do tej samej jednostki.

Jak pokazano na rysunku 1, do rejestru PESEL, zawierającego m.in. dane demograficzne dodano, stosując numer PESEL jako unikalny klucz, informacje z Centralnego Rejestru Ubezpieczonych (CRU). W ten sposób uzyskano łączną informację na temat stanu cywilnego (zmienna obserwowana w zbiorze PESEL) i tytułu ubezpieczenia społecznego (zmienna obserwowana w CRU). Następnie, wykorzystując kombinację wartości zmiennych płeć, wiek oraz adres zamieszkania jako klucz, dodano informację o statusie na rynku pracy (obserwowana w BAEL). Zbiory CRU i BAEL charakteryzują się mniejszym pokryciem niż PESEL<sup>9</sup>, stąd łączna obserwacja cech (kolor szary) możliwa jest tylko dla tych jednostek, które występowały w każdym ze zbiorów. Dla pozostałych występują braki danych (kolor biały). Zintegrowana w ten sposób baza może być bogatym, tanim (nie ma potrzeby pomiaru statystycznego – dane już istnieją) źródłem informacji społeczno-gospodarczych.



Rysunek 1. Integracja zbiorów danych pochodzących z różnych źródeł

Źródło: opracowanie własne

Efektem zastosowania łączenia deterministycznego jest surowa baza danych zawierająca wszystkie zmienne z łączonych baz. Są to zmienne niezharmonizowane, tj. o definicjach, wariantach, momentach referencyjnych, itp. pochodzących z baz wejściowych.

<sup>8</sup> Adres zamieszkania nie jest przedmiotem pomiaru w badaniach próbkowych, jednak jest on znany z operatu losowania.

<sup>9</sup> Np. do badania BAEL wylosowanych jest około 200 tys. jednostek z około 30-milionowej populacji, co daje pokrycie rzędu ok. 0,7%.

W przypadku, gdy zmienne kluczowe o unikatowych wartościach nie są dostępne, lub zawierają wspomniane wyżej błędy, niemożliwe jest wykorzystanie metod deterministycznych. W takich przypadkach możliwe jest stosowanie metod probabilistycznego łączenia rekordów (*probabilistic record linkage*). Idea tej metody sprowadza się do wyboru kilku zmiennych (nazywanych zmiennymi parującymi), które zawarte są w obu zbiorach i wykorzystanie ich zgodności, częściowej zgodności lub niezgodności do oszacowania prawdopodobieństwa, że poszczególne rekordy należą (lub nie) do tej samej jednostki.

W integrowanych bazach, ze względu na brak kompatybilności, często te same wartości zapisywane są w niejednolity sposób (np. adresy, imiona i nazwiska, nazwy własne itp.), które mogą wynikać zarówno z przyjętych przez gestorów baz różnych standardów zapisu lub wynikających z różnorodnego rodzajów błędów zapisu (np. ortograficznych, typograficznych, wynikających z niedoskonałości sprzętu i oprogramowania skanującego itp.). Za pomocą metodyki probabilistycznego łączenia rekordów można również zintegrować takie jednostki porównując pewne wartości w zmiennych występujących w obu bazach, które choć różnią się sposobem zapisu należą do tej samej obserwacji.

Najczęściej, w literaturze przedstawia się probabilistyczne łączenie rekordów jako proces kilkustopniowy. Pierwszym krokiem jest wybór zmiennych, na podstawie których przeprowadzone zostanie łączenie (tzw. zmienne parujące). W kolejnym kroku przygotowuje się bazy do procesu integracji poprzez usunięcie duplikatów oraz standaryzację wariantów cech parujących. Następnie dokonuje się operacji grupowania (nazywanej również blokowaniem) mającej na celu podział integrowanych baz na podzbiory, w których znajdują się jednostki w jakiś sposób do siebie podobne (np. mieszkańcy jednego powiatu lub przedstawiciele jednej gałęzi przemysłu). Grupowania dokonuje się z w celu optymalizacji algorytmu poprzez zredukowanie liczby połączeń. Następnie na podstawie pewnych algorytmów wykonuje się łączenie baz oraz sprawdzenie jego efektywności.

Efektem zastosowania probabilistycznego łączenia rekordów jest surowa baza danych uzupełniona o rekordy nieprzyłączone na etapie deterministycznego łączenia rekordów.

Połączenie zbiorów jest dopiero pierwszym krokiem do pełnej integracji. W celu uzyskania zbioru danych odpowiadającego celom statystycznym należy ujednolicić zawarte w nich informacje. Ten etap w literaturze nosi nazwę przetwarzania danych zintegrowanych (*micro-integration processing*). Wallgren, Wallgren (2007) oraz Linder (2004) wymieniają następujące etapy przetwarzania zintegrowanych rejestrów:

- kodowanie zmiennych – ujednolicanie wariantów cech,
- edycja braków danych – imputacja,
- wyrównywanie momentów lub okresów referencyjnych rejestrów – w celu zapewnienia możliwości porównań wyników (przykładem rozwiązania tego problemu na potrzeby NSP jest praca Bijaka, 2009),

- tworzenie jednostek pochodnych (np. gospodarstw domowych złożonych z osób mieszkających pod jednym adresem),
- tworzenie zmiennych pochodnych (np. utworzenie zmiennej „dochody całkowite” będącej sumą dochodów z różnych źródeł),
- porównywanie zmiennych z różnych źródeł w celu korekty błędów.

Etap przetwarzania danych zintegrowanych umożliwia utworzenie operacyjnej bazy mikrodanych<sup>10</sup>, w której znajdują się cechy o ujednoliconym charakterze, zdefiniowane według norm przyjętych w statystyce publicznej oraz o zadowalającej jakości. Przetwarzanie danych zintegrowanych jest stałym procesem, który zapewnia spójność i wysoką jakość informacji statystycznej.

Bakker (2010) wymienia cztery podstawowe zalety przetworzonej, operacyjnej bazy danych:

- rzetelność i wiarygodność komunikatów statystycznych sporządzonych na podstawie zintegrowanych źródeł jest poprawiona (w porównaniu do bazy „surowej”),
- możliwa jest publikacja oszacowań na niskim, niedostępnym dla badań reprezentacyjnych, poziomie agregacji przestrzennej i merytorycznej,
- zmienne z różnych źródeł są połączone i możliwa jest ich łączna obserwacja,
- możliwe jest przeprowadzenie badań panelowych.

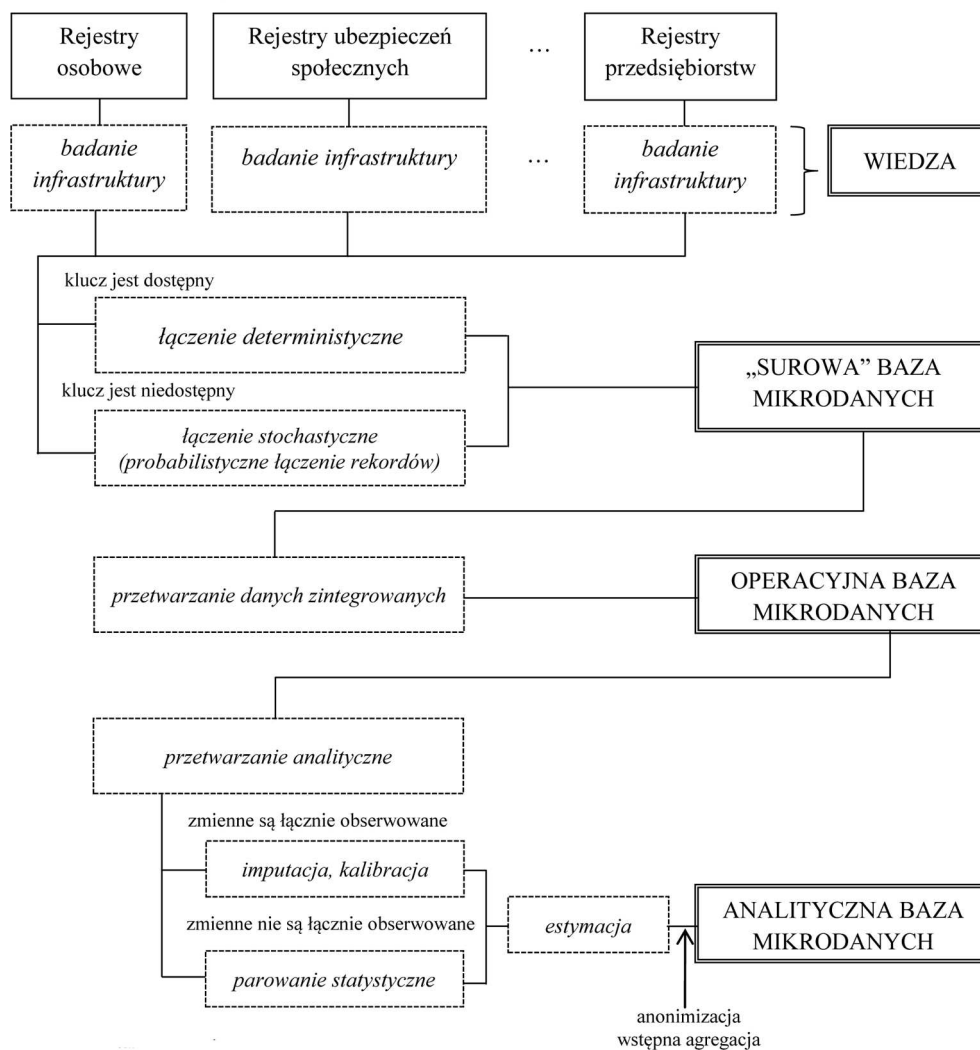
Integrowane zbiory danych charakteryzują się różnym stopniem pokrycia. Rejestry administracyjne zawierają informacje o bardzo dużej liczbie jednostek, natomiast badania reprezentacyjne charakteryzują się niewielkim pokryciem. Stąd też zintegrowana, operacyjna baza danych będzie zawierać pełną informację wyłącznie dla jednostek, które wystąpiły w każdym z integrowanych źródeł. Rekordy, które wystąpiły tylko w pojedynczych źródłach, dla cech dołączonych z innych baz będą charakteryzować się brakami danych.

Ostatnim etapem integracji danych jest proces przetwarzania analitycznego mikrodanych. Polega on na imputacji braków danych, kalibracji wag analitycznych oraz estymacji finalnych komunikatów statystycznych. Celem procesu jest zachowanie spójności numerycznej danych w sensie uzyskania takich samych wyników dla wszystkich oszacowań bez względu na źródło pochodzenia zmiennych w zintegrowanym zbiorze.

W zależności od źródła pochodzenia zmiennej, szacunki mogą charakteryzować rozbieżnościami nawet w skali całego kraju. Konsekwencje dywersyfikacji źródeł informacji mogą zostać zniwelowane poprzez różne metody analityczne, wśród których najczęściej wymienia się metody ważenia (wielokrotnego ważenia) oraz masowej imputacji (Kroese et al., 2000).

W metodzie wielokrotnego ważenia, w pierwszym etapie następuje wyodrębnienie podzbioru bazy operacyjnej bazy mikrodanych pochodzących z różnych źródeł. Następnie przyporządkowane są wagi początkowe. Później dostosowuje się wagi kalibracyjne uzyskane dla wszystkich podzbiorów danych w taki sposób, by uzyskane rezultaty były zgodne dla cech znajdujących w każdym z nich. Odbywa się to poprzez rekalkulację

<sup>10</sup> Noszącej również nazwę rejestru statystycznego (Wallgren, Wallgren, 2007).



Rysunek 2. Integracja administracyjnych źródeł danych

Źródło: opracowanie własne

wag w odniesieniu do rozkładów brzegowych zmiennych występujących w każdym podzbiorze. Zapewnia się w ten sposób spójność oszacowań przy jednoczesnym zachowaniu rzetelności oszacowań w sensie ich nieobciążoności (Zhang, 2012).

Alternatywą dla ważenia danych jest masowa imputacja. Zastosowanie metod imputacji w celu uzupełniania braków danych wartościami syntetycznymi zapewnia łączną obserwację wszystkich zmiennych w zintegrowanej bazie danych jednostkowych, jak również numeryczną spójność szacunków wszystkich cech (wszystkie wartości będą się sumować do liczebności ujednoliconej populacji generalnej). Wadą takiego podejścia jest fakt, że podstawione wartości są w dużej mierze sztuczne, nawet jeżeli

wynikają z dobrze dobranego modelu uzupełniania braków. W wyniku imputacji powstają jednostki nierzeczywiste (o nierzeczywistych wartościach cech), co z jednej strony powoduje, że maleje niebezpieczeństwo identyfikacji prawdziwych jednostek, ale z drugiej strony może prowadzić do obniżenia zgodności danych (np. 15-letni emeryt lub 40-letni przedszkolak). Dodatkowo, nawet jeżeli do imputacji stosuje się model skonstruowany w oparciu o wartości empiryczne (np. imputacja regresyjna), bardzo często na podstawie stosunkowo niewielkiej liczby wartości empirycznych, podstawia się wielką liczbę wartości teoretycznych (np. imputując aktywność ekonomiczną z BAEL – ok. 100 tys. rekordów – do rejestru PESEL – ok. 30 mln rekordów).

Może się zdarzyć, że nie wszystkie cechy są łącznie obserwowane lub łączna obserwacja cech zachodzi tylko dla niewielkiego podzbioru obserwacji uniemożliwiającego dokonywanie szacunków dla odpowiedniego przekroju danych. Proces estymacji może zostać w takim przypadku wsparty metodami parowania statystycznego (*statistical matching*). Parowanie statystyczne to grupa metod służących do integracji dwóch (lub więcej) źródeł danych (zwykle pochodzących z badań próbkowych) odnoszących się do tej samej populacji generalnej. Ponieważ prawdopodobieństwo wylosowania tej samej jednostki do dwóch różnych badań reprezentacyjnych jest bardzo małe (zbliżone do zera), zakłada się, że integrowane zbiory są rozłączne w sensie pokrycia. W każdym zbiorze (oznaczonymi jako A i B) znajduje się zwykle pewien wspólny wektor zmiennych (np. w badaniach dotyczących osób mogą być to zmienne demograficzne) o tych samych lub zbliżonych definicjach i wariantach. Nazywa się je zmiennymi wspólnymi (oznaczonymi jako X). Zbiór A zawiera wektor zmiennych obserwowanych wyłącznie w nim, oznaczy jako Y, natomiast zbiór B zawiera analogiczny wektor – Z (por. tabela 3). Celem parowania statystycznego jest analiza związków pomiędzy zmiennymi nieobserwowanymi łącznie w pojedynczym źródle.

Produktem integracji baz danych metodą parowania statystycznego są jednostki syntetyczne. Oznacza to, że w zintegrowanym zbiorze obserwacji podlegają podmioty niewystępujące w rzeczywistości. U źródeł koncepcji tworzenia zbiorowości utworzonych w sposób sztuczny leży założenie, że jednostki, które są do siebie podobne pod względem określonych cech (np. demograficznych, takich jak: wiek, płeć, miejsce zamieszkania, czy wykształcenie lub ekonomicznych, takich jak: aktywność ekonomiczna, źródło utrzymania, czy dochody) będą również podobne pod względem innych cech będących przedmiotem parowania.

Dzięki zastosowaniu metody parowania statystycznego możliwa jest łączna obserwacja cech nieobserwowanych wspólnie w żadnym ze źródeł. Umożliwia to dokonywanie analiz wielowymiarowych, jak np. badanie współzależności (wyznaczenie współczynników korelacji, czy też stworzenie tabel kontyngencji).

Metody parowania statystycznego stosuje się w sposób podobny do metody masowej imputacji, jednak kładzie się w nich nacisk, by, o ile to możliwe, dołączane

Tabela 3.

Dane wejściowe w parowaniu statystycznym

|            |                |     |                |                |     |                |
|------------|----------------|-----|----------------|----------------|-----|----------------|
| Zbiór<br>A | $\mathbf{Y}_1$ | ... | $\mathbf{Y}_Q$ | $\mathbf{X}_1$ | ... | $\mathbf{X}_P$ |
|            | $y_{11}^A$     | ... | $y_{1Q}^A$     | $x_{11}^A$     | ... | $x_{1P}^A$     |
|            | ...            | ... | ...            | ...            | ... | ...            |
|            | $y_{a1}^A$     | ... | $y_{aQ}^A$     | $x_{a1}^A$     | ... | $x_{aP}^A$     |
|            | ...            | ... | ...            | ...            | ... | ...            |
|            | $y_{n_A1}^A$   | ... | $y_{n_AQ}^A$   | $x_{n_A1}^A$   | ... | $x_{n_AP}^A$   |

Zbiór  
B

|                |     |                |                |     |                |
|----------------|-----|----------------|----------------|-----|----------------|
| $\mathbf{X}_1$ | ... | $\mathbf{X}_P$ | $\mathbf{Z}_1$ | ... | $\mathbf{Z}_R$ |
| $x_{11}^B$     | ... | $x_{1P}^B$     | $z_{11}^B$     | ... | $z_{1R}^B$     |
| ...            | ... | ...            | ...            | ... | ...            |
| $x_{b1}^B$     | ... | $x_{bP}^B$     | $z_{b1}^B$     | ... | $z_{bR}^B$     |
| ...            | ... | ...            | ...            | ... | ...            |
| $x_{n_B1}^B$   | ... | $x_{n_BP}^B$   | $z_{n_B1}^B$   | ... | $z_{n_BR}^B$   |

Źródło: opracowanie własne

wartości były wartościami empirycznymi<sup>11</sup>. Dołącza się również jedną zmienną jednocześnie dokładając starań, by model integracji zapewniał zgodność wielowymiarowych rozkładów w integrowanych zbiorach (Raessler, 2002).

Zintegrowane, w sposób deterministyczny lub (i) stochastyczny, dane z różnych źródeł, zapewniające wysokie pokrycie i łączną obserwację cech z różnych obszarów funkcjonowania społeczeństwa, gospodarki i państwa są również punktem wyjścia do tworzenia kompleksowych systemów statystycznych zapewniających wsparcie w tworzeniu symulacji i prognoz skuteczności działań organizacji państwowych (polityki podatkowej, opieki społecznej etc.) i prywatnych (inwestowanie w określone rejony, lokowanie produktów, zdobywanie klientów poprzez precyzyjnie przeprowadzane kampanie marketingowe). Takie wykorzystanie zintegrowanych źródeł nosi nazwę mikrosymulacji (*microsimulation*) i jest wykorzystywane m.in. w Europie (Atkinson *et al.* 1999), Kanadzie (Morrison 1998) oraz Australii (Kelly, 2003; Hardling *et al.*, 2009).

<sup>11</sup> Np. poszukiwanie tzw. „statystycznych bliźniąt” (Bacher, 2002), czyli wyszukiwanie w jednym zbiorze rekordów najbardziej podobnych do tych w drugim zbiorze i ich łączenie (Raessler, 2002; Di Zio *et al.*, 2006). Inną metodą jest tzw. wielokrotna imputacja (Raessler, 2002), gdzie dla jednej dołączanej wartości wyznacza się kilka wynikających z modelu lub najbardziej podobnych w innym zbiorze i albo dołącza się rekord wylosowany z tego zbioru „bliźniąt”, albo np. imputuje się wartość średnią.



## 4. ZAKOŃCZENIE

Integracja danych administracyjnych i badań reprezentacyjnych, przy zachowaniu odpowiednich procedur i metod prowadzi do utworzenia zbioru o szerokim spektrum informacyjnym i wysokim pokryciu. Koszt pozyskania informacji ze zbiorów zintegrowanych jest dużo niższy niż w przypadku klasycznych badań statystycznych, nie występuje obciążenie respondentów, a zbiory charakteryzują się wysoką jakością.

Dokonywanie szacunków na podstawie zintegrowanych źródeł wiąże się z zastosowaniem nowatorskich metod statystycznych, przekonstruowania systemu statystycznego oraz zachowania tajemnicy statystycznej. Zachowanie określonych procedur i norm integracji zapewnia możliwość tworzenia wielowymiarowych analiz na niskim poziomie agregacji przestrzennej, co może przyczyniać się do podejmowania właściwych decyzji dotyczących rozwoju społeczno-gospodarczego małych jednostek terytorialnych poprzez precyzyjne i rzetelne zasilanie informacyjne.

Zastosowanie metod statystycznej integracji danych wspomaga system statystyczny oparty o administracyjne źródła danych poprzez łączenie rekordów pozbawionych klucza identyfikacyjnego oraz umożliwia łączną obserwację cech nieobserwowanych zarówno w żadnym pojedynczym źródle, jak i w zbiorze zintegrowanym.

Łączna obserwacja cech opisujących różne zagadnienia społeczno-gospodarcze umożliwia zastosowanie mikrosymulacji wspomagającej organy administracyjne w procesach decyzyjnych dotyczących prowadzenia określonych polityk w ujęciu przestrzennym.

Uniwersytet Ekonomiczny w Poznaniu

## LITERATURA

- [1] Adamczewski P., (2005), *Słownik informatyczny*, Wydawnictwo Helion, Gliwice.
- [2] Atkinson A.B., Bourguignon F., O'Donoghue C., Sutherland H., Utili F., (1999), *Microsimulation and the formulation of policy: a case study of targeting in the European Union*, EUROMOD, Working Papers Series, Working Paper No. EM2/99.
- [3] Bacher J., (2002), *Statistisches Matching - Anwendungsmöglichkeiten, Verfahren und ihre praktische Umsetzung in SPSS*, ZA-Informationen, 51. Jg.
- [4] Bakker B., (2010), *Micro-Integration: State of the art (w:) Draft Report of WP1. State of the art on statistical methodologies for data integration*, ESSnet on Data Integration, WP1/D1.32/2010JUN
- [5] Ballano C., (2009), *A Census of Population Based on an Administrative Register*, Proceedings of Statistics Canada Symposium 2008, Data Collection: Challenges, Achievements and New Directions.
- [6] Borchsenius L., (2000), *From a conventional to a register-based census of population*, INSEE-Eurostat seminar on the censuses after 2001.
- [7] Bruhn A., (2001), *The next Population and Housing Census in Sweden is planned for 2005 – it will be totally register-based*, Symposium on Global Review of 200 Round of Population and Housing Censuses: Mid-Decade Assessment and Future Prospects, Statistic Division, Department of Economic and Social Affairs, United Nations, New York.
- [8] Bycroft Ch., (2011), *A register-based census: what is the potential for New Zealand*, Statistics New Zealand, Tatauranga Aotearoa, Wellington, New Zealand.

- [9] Cichomski B. (kierownik programu), Jerzyński T., Zieliński M., (2009), *Polskie Generalne Sondaże Społeczne: skumulowany komputerowy zbiór danych 1992-2008*, Instytut Studiów Społecznych, Uniwersytet Warszawski, Warszawa.
- [10] *Data Integration Manual*, (2006), praca zbiorowa Statistics New Zealand, Wellington.
- [11] D'Orazio M., Di Zio M., Scanu M., (2006), *Statistical Matching. Theory and Practice*, John Wiley & Sons Ltd., England.
- [12] Dygaszewicz J., (2010), *Integracja rejestrów publicznych*, Główny Urząd Statystyczny, Warszawa.
- [13] Dygaszewicz J., (2012), *Spisy powszechne jako źródło danych do analiz geoprzestrzennych*, Archiwum Fotogrametrii, Kartografii i Teledetekcji, Vol. 23.
- [14] Gill L., (2001), *Methods for Automatic Record Matching and Linkage and their use in National Statistics*, National Statistics Methodological Series No 25, National Statistics, United Kingdom.
- [15] Hardling A., Kelly S., Percival R., Keegan M., (2009), *Population Ageing and Government Age Pension Outlays*, ESRI International Collaboration Project, NATSEM, University of Canberras.
- [16] Janczur-Knapk M., (2012), *Spisy powszechne PSR 2010 i NSP 2011 oraz systemy informacji geograficznej w statystyce publicznej*, referat wygłoszony na Kongresie Statystyki Polskiej, Poznań.
- [17] Kamen C.S., (2005), *The 2008 Israel Integrated Census of Population and Housing, Basic conception and procedure*, State of Israel, Central Bureau of Statistics.
- [18] Kelly S., (2003), *Australia's Microsimulation Model – Dynamod*.
- [19] Kroese B., Renssen R.H., Trijssenaar M., (2000), *Weighting or imputation: constructing a consistent set of estimates based on data from different sources*, „Netherlands Official Statistics”, vol. 15, Summer 2000, Special issue: *Integrating administrative registers and household surveys*, Statistics Netherlands, Voorburg/Heerlen.
- [20] Linder F., (2004), *The use of administrative registers and sample surveys in the Dutch Census of 2001* (w:) *The Dutch Virtual Census of 2001. Analysis and Methodology*, Statistics Netherlands, Voorburg/Heerlen
- [21] Morrison R., (1998), *Overview of DYNACAN: a full-fledged Canadian actuarial stochastic model designed for the fiscal and policy analysis of social security schemes*, <http://www.actuaries.org/CTTEES.SOCSEC/Documents/dynacan.pdf>
- [22] Nordholdt E.S., (2004), *Introduction to the Dutch Virtual Census of 2001* (w:) *The Dutch Virtual Census of 2001. Analysis and Methodology*, Statistics Netherlands, Voorburg/Heerlen.
- [23] Penneck S., (2007), *Using administrative data for statistical purposes*, Economic & Labour Market Review.
- [24] Prevost R., Leggieri Ch., (1999), *Expansion of Administrative Records Uses at the Census Bureau: A Long-Range Research Plan*, U.S. Bureau of the Census, Washington D.C.
- [25] Program badań statystycznych statystyki publicznej na 2012 rok, załącznik do rozporządzenia Rady Ministrów z dnia 22.07.2011 r. w sprawie programu badań statystycznych statystyki publicznej na rok 2012 (Dz. U. Nr 173, poz. 1030).
- [26] Raessler S., (2002), *Statistical Matching. A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*, Springer, New York, USA.
- [27] Raessler S., (2004), *Data fusion: identification problems, validity, and multiple imputation*, Austrian Journal of Statistics 33(1-2).
- [28] Ralphs M., Tutton P., (2011), *Beyond 2011: International models for census taking: current processes and future developments*, Beyond 2011 Project, Office for National Statistics.
- [29] Statistics Austria, (2008), *Register-based census 2010 and census test 2006*, Joint UNECE/Eurostat meeting on population and housing censuses, Geneva.
- [30] Statistics Finland, (2004), *Use of Registers and Administrative Data Sources for Statistical Purposes. Best Practices of Statistics Finland*, Tilastokeskus, Statistikcentralen, Statistics Finland, Helsinki.
- [31] Swiss Federal Statistical Office, (2008), *The Swiss Census 2010: Moving towards a comprehensive system of household and person statistics*, Federal Statistical Office.

- [32] Tonder J-K., (2008), *The register-based statistical system – preconditions and processes*, IAOS Conference, Shanghai.
- [33] Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne (Dz.U. Nr 64, poz.565, z późn. zm.).
- [34] Ustawa z dnia 4 marca 2010 roku o narodowym spisie powszechnym ludności i mieszkań w 2011 r. (Dz.U. z 2010 nr 47 poz. 277).
- [35] Ustawa z dnia 29 czerwca 1995 roku o statystyce publicznej, tekst jednolity (Dz. U. 2012.591).
- [36] van der Laan P., (2000), *Integrating administrative registers and household surveys*, „Netherlands Official Statistics”, vol. 15, Summer 2000, Special issue: *Integrating administrative registers and household surveys*, Statistics Netherlands, Voorburg/Heerlen.
- [37] Wallgren A., Wallgren B., (2007), *Register-based Statistics. Administrative Data for Statistical Purposes*, John Wiley and Sons Ltd.
- [38] Zhang L-C., (2012), *Micro calibration for data integration*, referat wygłoszony na Kongresie Statystyki Polskiej, Poznań.

## SYSTEM STATYSTYKI PUBLICZNEJ OPARTY NA ZINTEGROWANYCH ŹRÓDŁACH DANYCH

### Streszczenie

Zwiększające się zapotrzebowanie na rzetelną i aktualną informację na możliwie niskim poziomie agregacji jest rosnącym wyzwaniem dla polskiej statystyki publicznej. Zastosowanie integracji danych z różnych, w tym administracyjnych, źródeł umożliwia wykorzystanie informacji pełnej w sensie pokrycia oraz bogatej merytorycznie. Łączna obserwacja cech obserwowanych w oddzielnych zbiorach generuje efekt synergii zwiększający zasób wiedzy pochodzący z badań społeczno-ekonomicznych.

**Słowa kluczowe:** integracja danych, rejestry administracyjne, spisy powszechne

## THE SYSTEM OF PUBLIC STATISTICS BASED ON INTEGRATED DATA SOURCES

### Abstract

Increasing demand for reliable and current information at the lowest possible level of aggregation is a growing challenge for the Polish public statistics. Application of data integration of different, including administrative, sources enables the use of information in terms of full coverage and rich in substance. The joint observation of variables observed in separate data collections generate a synergy effect increasing amount of knowledge derived from socio-economic research.

**Key words:** data integration, administrative registers, censuses

JACEK KOWALEWSKI, MONIKA NATKOWSKA

## WYKORZYSTANIE MAPY STATYSTYKI KRÓTKOOKRESOWEJ W PROCESIE ORGANIZACJI BADAŃ PRZEDSIĘBIORSTW PROWADZONYCH PRZEZ STATYSTYKĘ PUBLICZNĄ

### 1. WPROWADZENIE

Postęp techniczny, którego jesteśmy świadkami oraz towarzyszące mu, szybko zmieniające się technologie prowadzą do istotnych zmian w szeregu dziedzinach życia. Dotyczy to także statystyki publicznej, która ze swej natury pełni rolę służebną wobec społeczeństwa i wnosi istotny wkład w kreowanie społecznego ładu informacyjnego.

Jednym z podstawowych dylematów, z którym służby statystyczne muszą się zmierzyć jest narastająca dysproporcja pomiędzy, i tak znaczącym, zakresem udostępnianych informacji, a coraz większym popytem na dane, od których wymaga się coraz wyższej jakości, większej szczegółowości i szybszego ich dostarczenia. Wymusza to poszukiwanie nowych metod uzyskiwania informacji, takich jak wykorzystanie rejestrów administracyjnych i nowych technik statystycznych. Pociąga to za sobą także wdrażanie nowych technologii informatycznych oraz związanych z nimi poszukiwań nowych form organizacji pracy służb statystycznych.

Istotna zmiana w pracach realizowanych przez Główny Urząd Statystyczny nastąpiła w roku 2009. Począwszy od tego roku, prawie cała sprawozdawczość pozytywna od podmiotów gospodarczych realizowana jest drogą elektroniczną, poprzez portal sprawozdawczy. Obok ewidentnych korzyści takich jak ograniczenie kosztów papieru, potencjalne skrócenie czasu realizacji badań oraz wzrost wydajności pracy, przyniosła ona także konieczność reorganizacji statystyki w województwach. Jednym z jej elementów było wprowadzenie specjalizacji urzędów w określonych wiązkach tematycznych. Na przykład dla Urzędu Statystycznego w Poznaniu zostały przydzielone zadania obejmujące: statystykę krótkookresową, statystykę miast, statystykę małych obszarów oraz tworzenie banku danych makroekonomicznych. Ze specjalizacją wiąże się szereg działań takich jak programowanie, organizacja baz danych, prace metodologiczne czy rozwijanie współpracy międzynarodowej. Właściwa realizacja tych prac wymusiła potrzebę pogłębienia wiedzy o systemie badań statystycznych. Stąd też narodziła się idea utworzenia mapy badań statystycznych, jako podstawy do precyzyjnego określenia odbiorców informacji, możliwych źródeł danych oraz rozszerzonej analizy sposobu organizacji badania.

## 2. MAPA SYSTEMU BADAŃ STATYSTYCZNYCH

Przez mapę systemu badań statystycznych należy rozumieć uogólniony, zintegrowany i uporządkowany opis organizacji badań statystycznych, zawierający takie elementy jak:

– *Podstawowe informacje o badaniach*, obejmujące: zakres podmiotowy i przedmiotowy, częstotliwość oraz termin realizacji badania, metody pozyskiwania danych, organizację badania.

– *Charakterystyka danych wejściowych*, prezentująca: źródła i formę danych, zmienne wejściowe, elementy pozwalające na ocenę jakości (np. liczbę odmów, braków, szacunki błędów).

– *Charakterystyka wyników badań*<sup>1</sup>, tj. postać (dane surowe, wyrównane dniami roboczymi lub sezonowo, trend), formę prezentacji (wskaźnik, wartość bezwzględna), termin i miejsce publikacji lub zastosowanie innego sposobu udostępnienia.

– *Opis relacji transformacji* danych wejściowych do wyników. Jest on elementem kluczowym do dalszych pogłębionych analiz.

Do opisu tej relacji przyjęto dalej następujące założenia:

1. Możliwe jest precyzyjne zdefiniowanie wektora  $Y$ , zawierającego zmienne, których wartości są końcowym efektem całego procesu badań statystycznych, gromadzone i udostępniane odbiorcom zewnętrznym. Wektor ten będzie nazywany wektorem wyjściowym<sup>2</sup>. Tym samym zasadniczym celem procesu badań statystycznych jest oszacowanie wartości  $n$  różnych docelowych zmiennych:

$$Y^T = [y_1, y_2, y_3, \dots, y_n]. \quad (1)$$

2. Wartości zmiennych wyjściowych  $Y$ , są bezpośrednio oszacowywane w oparciu o zmienne wejściowe (pierwotne), uzyskiwane jako wynik badania statystycznego, które będą reprezentowane przez  $m$  elementowy wektor  $X$ :

$$X^T = [x_1, x_2, x_3, \dots, x_m]. \quad (2)$$

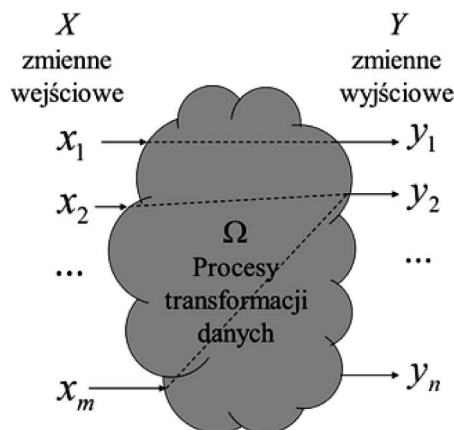
3. Proces przekształcania zmiennych wejściowych  $X$  w zmienne wyjściowe  $Y$  będzie nazywany procesem transformacji danych i oznaczany jako:

$$\Omega : X \times \dots \times X \rightarrow Y. \quad (3)$$

Możemy to przedstawić jak na rys. 1.

<sup>1</sup> Nie jest intencją autorów podejmowanie w artykule dyskusji na temat (bardzo obszerny sam w sobie) definicji badań statystycznych i ich wyników. Interesujące rozważania w tej kwestii zawierają prace np.: Sundgren (2003) oraz Stefanowicz (2007).

<sup>2</sup> Pod pojęciem *zmienna wyjściowa* rozumiane są tutaj różne obiekty. Może to być pojedyncza liczba (np. wskaźnik), wektor, macierz. Można także w ogólnym modelu przyjąć inne postaci wyników (np. wykresy mapowe).



Rysunek 1. Proces transformacji danych

Źródło: opracowanie własne

Transformacje mogą mieć różny charakter, a oszacowanie wartości określonej zmiennej wyjściowej  $y_i$  uzyskiwane jest drogą przekształcenia jednej lub kilku zmiennych wejściowych:

$$y_i = f(X). \quad (4)$$

Mogą one mieć charakter wieloetapowy, jak również w krańcowym przypadku oszacowanie zmiennej wyjściowej  $y_i$  może być tożsame z oszacowaniem pewnej zmiennej wejściowej  $x_j$ .

Jednym ze sposobów opisu wzajemnych zależności zmiennych, użytecznym w zakresie analizy całego systemu jest przedstawienie ich w postaci macierzy transformacji  $M$ :

$$M = \begin{bmatrix} m_{11} & \dots & m_{1j} & \dots & m_{1n} \\ \dots & & m_{ij} & & \dots \\ & & \dots & & \\ m_{m1} & \dots & m_{mj} & \dots & m_{mn} \end{bmatrix}, \quad (5)$$

gdzie  $m_{ij} = \begin{cases} 1, & \text{gdy zmienna wejściowa } x_i \text{ jest używana do oszacowania wartości} \\ & \text{zmiennej } y_j \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$

### 3. ISTOTA STATYSTYKI KRÓTKOOKRESOWEJ

Statystyka krótkookresowa (ang. Short-Term Statistics – STS) jest częścią statystyki przedsiębiorstw zajmującą się dostarczaniem bieżących informacji o sytuacji przedsiębiorstw prowadzących działalność gospodarczą w czterech głównych dziedzinach gospodarki: przemyśle, budownictwie, handlu detalicznym i naprawach oraz pozostałych

usługach<sup>3</sup>. Informacje te, prezentowane dla okresów miesięcznych i kwartalnych, służą między innymi do celów:

- monitorowania polityki ekonomicznej i fiskalnej Unii Europejskiej (UE) oraz strefy euro,
- prowadzenia oceny bieżącego rozwoju UE oraz każdego z krajów członkowskich,
- określania fazy cyklu koniunkturalnego poszczególnych gospodarek.

W kontraście do innych badań przedsiębiorstw, dane statystyczne krótkookresowe nie są prezentowane w postaci liczb absolutnych czy zmiennych wartościowych, np. pieniężnych. Uzyskane wielkości odnoszone są do roku bazowego (obecnie 2005) i przedstawiane za pomocą wskaźników w postaci szeregów czasowych. Przyjęta forma prezentowania informacji umożliwia analizę tendencji rozwojowej badanych zjawisk oraz obserwowania zmian np. w poziomie produkcji sprzedanej czy obrotu na przestrzeni czasu.

Szacowanie wskaźników krótkookresowych dla Unii Europejskiej opiera się o dane zbierane i opracowywane w poszczególnych krajowych urzędach statystycznych. Wyniki są następnie transmitowane do Eurostatu (urząd statystyczny UE), gdzie są publikowane oraz agregowane na poziomie UE, regionów oraz strefy euro.

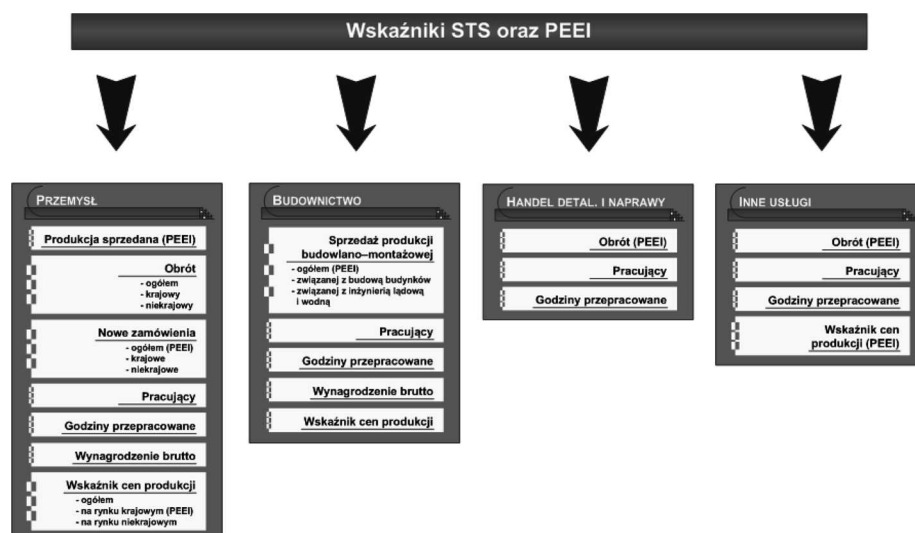
Regulacje dotyczące statystyk krótkookresowych<sup>4</sup> obligują narodowe urzędy statystyczne państw Unii Europejskiej do ścisłego stosowania przyjętych zasad przygotowania danych. Wspólne ramy prawne oraz ujednolicone wytyczne dla opracowania wskaźników mają na celu uzyskanie spójnych oraz zharmonizowanych statystyk umożliwiających porównywanie wyników gospodarczych między państwami członkowskimi. Jednocześnie mają odzwierciedlać faktyczną sytuację ekonomiczną przedsiębiorstw w Unii Europejskiej. Szczególną wartość w tym zakresie ma, wyodrębniona z 300 eurow-skaźników, grupa Głównych Europejskich Wskaźników Gospodarczych (ang. Principal European Economic Indicators – PEEI), do których należą:

- Produkcja sprzedana w przemyśle,
- Wskaźnik cen producentów w przemyśle na rynku krajowym,
- Nowe zamówienia w przemyśle,
- Ceny importu w przemyśle,
- Sprzedaż produkcji budowlano-montażowej,
- Obrót w handlu detalicznym i naprawach,
- Obrót w pozostałych usługach,
- Wskaźnik ceny produkcji usług,

Zestaw wskaźników krótkookresowych (STS) oraz głównych europejskich wskaźników gospodarczych (PEEI) prezentuje rysunek 2.

<sup>3</sup> Eurostat 2006, Methodology of short-term business statistics, Interpretation and guidelines, European Communities.

<sup>4</sup> rozporządzenie Rady (WE) Nr 1165/98 dotyczące krótkookresowych statystyk oraz późniejsze jego nowelizacje, które definiują podmiotowy i przedmiotowy zakres badań oraz zmienne, które muszą być opracowane i przesłane w wyznaczonych terminach do Eurostatu.



Rysunek 2. Wskaźniki krótkookresowe (STS) oraz główne europejskie wskaźniki gospodarcze (PEEI) wg dziedzin gospodarczych

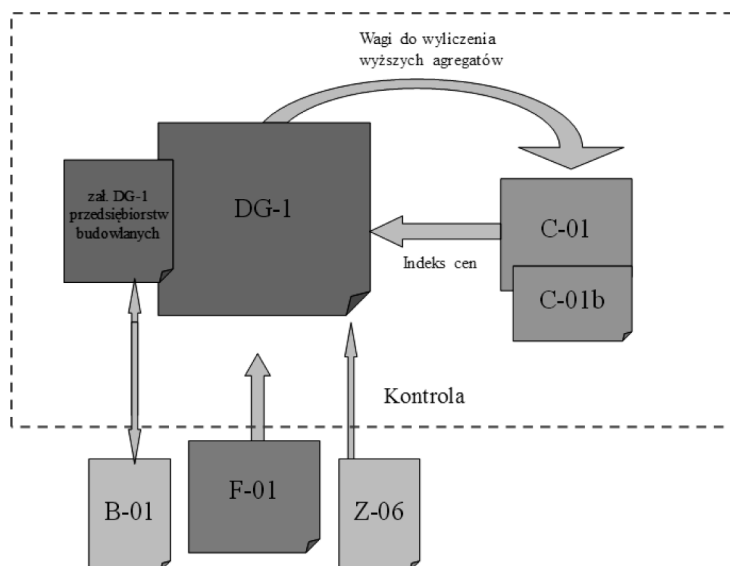
Źródło: opracowanie własne

W Polsce w badaniach prowadzonych przez Główny Urząd Statystyczny (GUS) najważniejszymi źródłami informacji do produkcji statystyk krótkookresowych są sprawozdania:

- DG-1 – Meldunek o działalności gospodarczej,
- Załącznik do DG-1 – Załącznik do meldunku DG-1 dla przedsiębiorstw budowlanych,
- C-01 – Sprawozdanie o cenach producentów wyrobów i usług,
- C-01b – Sprawozdanie o cenach robót budowlano-montażowych.

Największą liczbę zmiennych do opracowania wskaźników krótkookresowych wnosi badanie DG-1. Na jego podstawie pozyskujemy między innymi informacje o wielkości produkcji sprzedanej i obrotów, wartości nowych zamówień w przemyśle, wielkości sprzedaży detalicznej czy liczbie pracujących i poziomie wynagrodzeń. Sprawozdania C-01 i C-01b dostarczają głównie informacji o zmianach cen produkcji, odpowiednio przemysłowej i budowlano-montażowej. Należy jednak podkreślić, że badania DG-1 i C-01 wzajemnie się uzupełniają i nie można ich rozpatrywać rozdzielnie. C-01 dostarcza dla DG-1 indeksów cen do przeliczenia wartości produkcji sprzedanej, natomiast DG-1 wag do wyliczenia wskaźników cen w badaniu C-01. W celach kontrolnych wykorzystywane są także dane pozyskiwane ze sprawozdań dotyczących: produkcji budowlanej (B-01), zatrudnienia (Z-06) oraz finansów przedsiębiorstw (F-01) (rys. 3).





Rysunek 3. Schemat przetwarzania w badaniu DG-1

Źródło: opracowanie własne

#### 4. MAPA STATYSTYKI KRÓTKOOKRESOWEJ

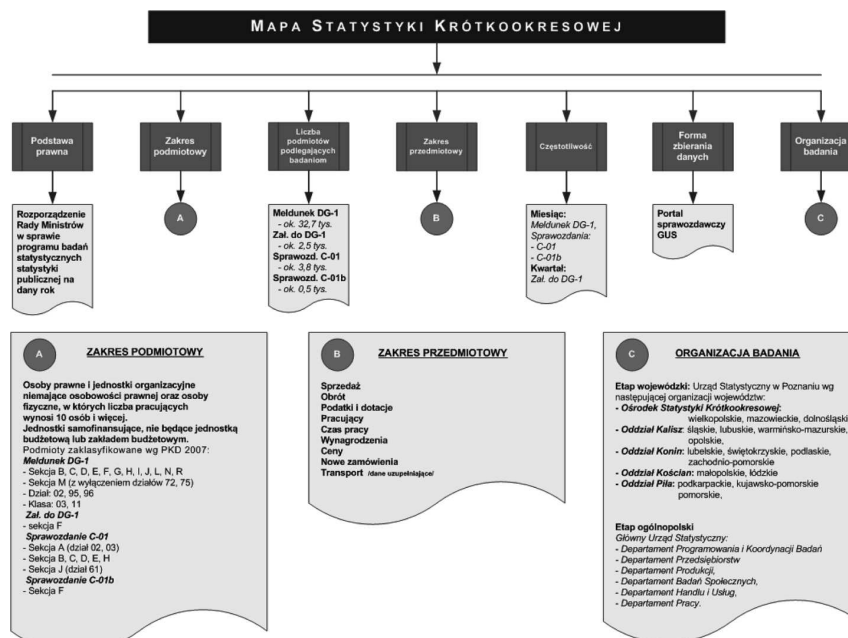
Mapa Statystyki Krótkookresowej (Mapa SK) jest systemem informacyjnym, który obejmuje badania statystyki krótkookresowej realizowane przez Urząd Statystyczny w Poznaniu.

Ogólna charakterystyka określa organizację badań (formę pozyskiwania danych, częstotliwość i tryb realizacji w strukturach GUS), kryteria doboru podmiotów, liczbę jednostek objętych obowiązkiem sprawozdawczym oraz przedmiotowy zakres badań.

Z punktu widzenia analizy badań, istotna jest drobiazgowa wiedza dotycząca zakresu niezbędnych informacji tak wejściowych, jak i wynikowych. Mapa SK porządkuje więc informacje, które stanowią przedmiot badania statystyki krótkookresowej, a także dostarcza wiedzy na temat obszarów wykorzystania danych statystycznych.

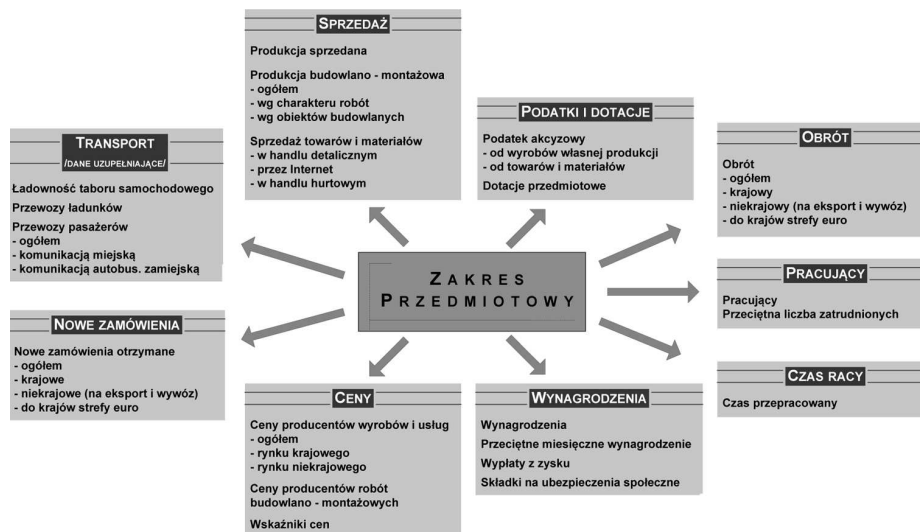
Kluczowa w tym aspekcie jest znajomość odbiorców informacji statystycznej oraz celu wykorzystania statystyk. Wiedza na temat wymaganego dla poszczególnych odbiorców poziomu szczegółowości danych oraz oczekiwanych terminów dostępności statystyk umożliwia szczegółową analizę i ocenę potrzeb użytkowników. Informacje te stanowią przesłanki do ewentualnej reorganizacji badań pod kątem wymagań odbiorców danych statystycznych.

Głównym adresatem statystyk krótkookresowych jest wspomniany już Eurostat. Jednak duża wartość informacyjna wskaźników powoduje iż mają one również innych odbiorców instytucjonalnych. Trafiają one między innymi do organów krajowych – rządu czy banku centralnego, a także na rynki finansowe, do analizy bieżącej sytuacji ekonomicznej kraju i podejmowania decyzji w zakresie polityki gospodarczej i finan-



Rysunek 4. Charakterystyka badań statystyki krótkookresowej

Źródło: opracowanie własne



Rysunek 5. Zakres przedmiotowy badań statystyki krótkookresowej wg bloków tematycznych

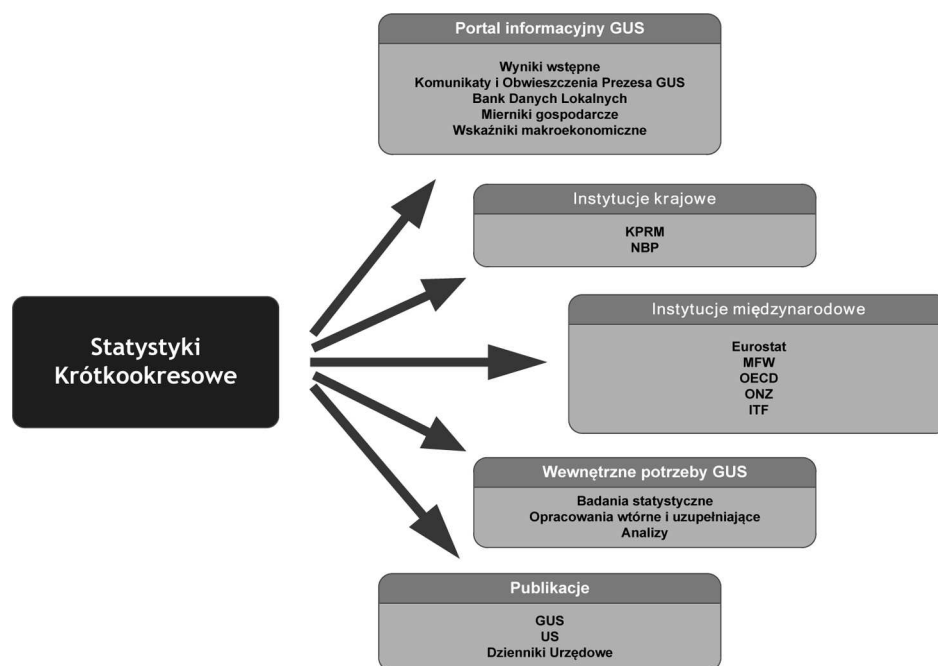
Źródło: opracowanie własne

sowej. W Polsce bezpośrednim odbiorcą informacji krótkookresowych jest Kancelaria Prezesa Rady Ministrów (KPRM) oraz Narodowy Bank Polski (NBP).

Nie mniej ważną grupą odbiorców są instytucje międzynarodowe, którym przekazujemy powyższe dane z tytułu przynależności członkowskich:

- Międzynarodowy Fundusz Walutowy (MFW),
- Organizacja Współpracy Gospodarczej i Rozwoju (OECD),
- Organizacja Narodów Zjednoczonych (ONZ),
- Międzynarodowa Federacja Pracowników Transportu (ITF).

Organizacje te są zainteresowane otrzymywaniem szerokiego wachlarza informacji na temat działalności polskich przedsiębiorstw, tj. poziomu sprzedanej produkcji przemysłowej oraz budowlano-montażowej, wskaźnika cen produkcji przemysłowej, wielkości sprzedaży detalicznej, a także czynników produkcji – liczby pracujących i zatrudnionych oraz poziomu przeciętnych wynagrodzeń.

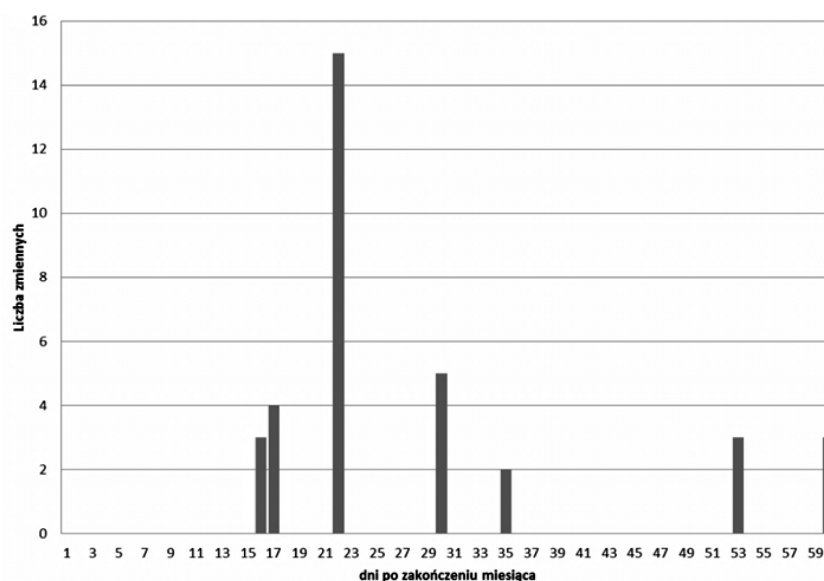


Rysunek 6. Obszary udostępniania danych statystycznych

Źródło: opracowanie własne

Krótkoterminowe statystyki cieszą się również zainteresowaniem mediów, analityków gospodarczych i finansowych oraz przedsiębiorców, dla których stanowią bazę do obserwacji koniunktury gospodarczej kraju lub wybranych dziedzin gospodarki. Stąd bogactwo opracowań z wykorzystaniem danych statystycznych krótkookresowych publikowanych przez Główny Urząd Statystyczny. Część informacji dostępna jest bezpośrednio poprzez portal informacyjny GUS w postaci zestawień tabelarycznych, a część wydawana jest w ramach cyklicznych publikacji. Warto tu wspomnieć o comiesięcznych opracowaniach, takich jak *Informacja o sytuacji społeczno-gospodarczej kraju*,

*Biuletyn statystyczny* oraz *Ceny w gospodarce narodowej*. Podstawowe informacje prezentowane są również w wynikach wstępnych oraz komunikatach i obwieszczeniach Prezesa GUS.



Rysunek 7. Terminy udostępniania danych krótkookresowych po zakończeniu miesiąca

Źródło: opracowanie własne

Mapa SK jest także pewnego rodzaju przewodnikiem po wydawnictwach GUS, w których publikuje się statystyki krótkookresowe. Dostarcza informacji o zakresie prezentowanych, w poszczególnych opracowaniach, danych biorąc pod uwagę PKD (Polska Klasyfikacja Działalności), zasięg terytorialny, formę i postać danych wynikowych oraz okres obserwacji.

Należy podkreślić, iż w procesie analizy badań nie wystarczy sama znajomość bieżącego popytu na informację. Konieczna jest również obszerna wiedza z zakresu metodologii badań, a w szczególności przebiegu transformacji zmiennych uczestniczących w badaniu.

Zmienna wejściowa, przekazywana w badaniu przez respondenta, na ogół nie jest bezpośrednio przekształcana w zmienną wynikową, lecz bierze udział w szacowaniu kilku danych statystycznych. Jako przykład można wymienić zmienną pozyskiwaną w badaniu DG-1 – *Przychody netto ze sprzedaży produktów (wyrobów i usług)*, która nie jest publikowana, ale wykorzystywana do naliczania wartości obrotów, produkcji sprzedanej oraz wydajności pracy. Z kolei obliczenie *Przeciętnego miesięcznego wynagrodzenia* wymaga pozyskania dwóch zmiennych wejściowych: *Wynagrodzeń brutto* oraz *Przeciętnej liczby zatrudnionych*. Zmienne te stanowią jednocześnie odrębne dane wynikowe.

| Formularz | Zmienne wejściowe                                                       | Zmienna wynikowa |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|-----------|-------------------------------------------------------------------------|------------------|-------------------------|----------------------------|---------------------------------------|-------|---------------|------------------|-----------------------------|---------------------------------------------|---------------------------------|---------------------------------------------|-------------------------------------------------------|-------------------------------------------------------------|--------------------------------------------------------------------------|-------------------------------------------------------------|-------------------------------------|
|           |                                                                         | Nowe zamówienia  | Nowe zamówienia krajowe | Nowe zamówienia nielokalne | Nowe zamówienia do krajów strefy euro | Obrot | Obrot krajowy | Obrot nielokalny | Obrot do krajów strefy euro | Produkcja i sprzedaż przemysłowa (c. stała) | Sprzedaż usług w cenach stałych | Produkcja w przeliczeniu w cenach bieżących | Sprzedaż produktów budowlano-montażowych (ceny stałe) | Sprzedaż produktów budowlano-montażowych w cenach bieżących | Sprzedaż produktów budowlano-montażowych w cenach bieżących (ceny stałe) | Sprzedaż produktów budowlano-montażowych w cenach bieżących | Wydatki na prace w cenach bieżących |
| DG-1      | Dotacje przedmiotowe                                                    |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Nowe zamówienia otrzymane                                               | ✓                | ✓                       |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Nowe zamówienia otrzymane do krajów strefy euro                         |                  |                         | ✓                          |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Nowe zamówienia otrzymane na eksport i wywóz                            | ✓                | ✓                       |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Podatek akcyzowy od towarów i materiałów                                |                  |                         |                            |                                       | ✓     | ✓             |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Podatek akcyzowy od wyrobów własnej produkcji                           |                  |                         |                            |                                       | ✓     | ✓             |                  |                             | ✓                                           | ✓                               | ✓                                           |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Przebiegi liczb zatrudnionych                                           |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Przychody netto ze sprzedaży produktów                                  |                  |                         |                            |                                       | ✓     | ✓             |                  |                             | ✓                                           | ✓                               | ✓                                           |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Przychody netto ze sprzedaży produktów do krajów strefy euro            |                  |                         |                            |                                       |       |               |                  |                             | ✓                                           |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Przychody netto ze sprzedaży produktów na eksport i wywóz               |                  |                         |                            |                                       | ✓     | ✓             |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Przychody netto ze sprzedaży towarów i materiałów                       |                  |                         |                            |                                       | ✓     | ✓             |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Przychody netto ze sprzedaży towarów i materiałów do krajów strefy euro |                  |                         |                            |                                       |       |               |                  |                             | ✓                                           |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Przychody netto ze sprzedaży towarów i materiałów na eksport i wywóz    |                  |                         |                            |                                       | ✓     | ✓             |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Przychody netto ze sprzedaży produkcji budowlano-montażowej             |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 | ✓                                           | ✓                                                     | ✓                                                           | ✓                                                                        | ✓                                                           |                                     |
|           | Przychody ze sprzedaży prod. budowl.-montaż. o charakt. inwestycyjnym   |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
| C-01      | Przewozy ładunków                                                       |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Przewozy pasażerów                                                      |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             | ✓                                   |
| C-01b     | Sprzedaż detaliczna                                                     |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Sprzedaż hurtowa                                                        |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             |                                                       |                                                             |                                                                          |                                                             |                                     |
| Zał. DG-1 | Wartość produktów wytworzonych niezaliczonych do sprzedaży              |                  |                         |                            |                                       |       |               |                  |                             | ✓                                           | ✓                               | ✓                                           |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Wskaźnik cen produkcji sprzedanej przy podstawie mc poprzedni-100       |                  |                         |                            |                                       |       |               |                  |                             | ✓                                           | ✓                               | ✓                                           |                                                       |                                                             |                                                                          |                                                             |                                     |
| C-01b     | Cena sprzedaży reprezentanta bez VAT i akcyzy w m-cu poprzednim         |                  |                         |                            |                                       |       |               |                  |                             | ✓                                           | ✓                               | ✓                                           |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Cena sprzedaży reprezentanta bez VAT i akcyzy w m-cu sprawozdawczym     |                  |                         |                            |                                       |       |               |                  |                             | ✓                                           | ✓                               | ✓                                           |                                                       |                                                             |                                                                          |                                                             |                                     |
| C-01b     | Wartość sprzedaży reprezentanta ogółem bez VAT i akcyzy                 |                  |                         |                            |                                       |       |               |                  |                             | ✓                                           | ✓                               | ✓                                           |                                                       |                                                             |                                                                          |                                                             |                                     |
|           | Wartość sprzedaży reprezentanta ogółem bez VAT i akcyzy                 |                  |                         |                            |                                       |       |               |                  |                             | ✓                                           | ✓                               | ✓                                           |                                                       |                                                             |                                                                          |                                                             |                                     |
| C-01b     | Cena sprzedaży reprezentanta ogółem bez VAT w m-cu poprzednim           |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             | ✓                                                     | ✓                                                           | ✓                                                                        |                                                             |                                     |
|           | Cena sprzedaży reprezentanta ogółem bez VAT w m-cu sprawozdawczym       |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             | ✓                                                     | ✓                                                           | ✓                                                                        |                                                             |                                     |
| C-01b     | Wartość sprzedaży reprezentanta ogółem bez VAT                          |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             | ✓                                                     | ✓                                                           | ✓                                                                        |                                                             |                                     |
|           | Wartość sprzedaży reprezentanta ogółem bez VAT                          |                  |                         |                            |                                       |       |               |                  |                             |                                             |                                 |                                             | ✓                                                     | ✓                                                           | ✓                                                                        |                                                             |                                     |

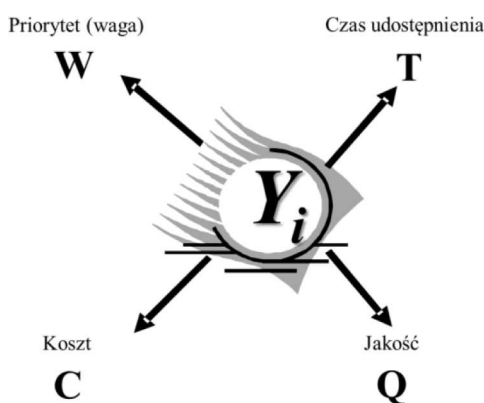
1) PROB = 111 - 113, 121 - 127, 211-215, 221-223, 230, 240-242, 2223.

Rysunek 8. Fragment macierzy transformacji statystyki krótkookresowej

Źródło: opracowanie własne

## 5. WYKORZYSTANIE MAPY SYSTEMU STATYSTYKI KRÓTKOOKRESOWEJ W ANALIZIE BADANIA

Użyteczność mapy systemu badań statystycznych, polega nie tylko na tym, że porządkuje pewną wiedzę, ale po wprowadzeniu dodatkowych parametrów, może służyć różnorodnym analizom związanym z procesem wytwarzania i dystrybucji informacji statystycznej. W tym celu pożądane jest określenie dla każdej zmiennej wyjściowej  $y_i$  wielkości, np. takich jak opisane na rysunku 9.



Rysunek 9. Przypisanie cech dla zmiennej wyjściowej

Źródło: opracowanie własne

W zakresie statystyki krótkookresowej przyjęto następujące założenia w odniesieniu do poszczególnych kryteriów<sup>5</sup>:

- *Priorytet (waga)* – wyodrębniono 3 kategorie informacji wynikowych. Najwyższą wagę przypisano danym, których oszacowanie wynika ze zobowiązań międzynarodowych bądź regulacji ustawowych. Najniższą wagę natomiast przypisano informacjom uzupełniającym.
- *Czas udostępniania* – wyjściowe wartości przyjęto w oparciu o stosowane dotychczas praktyki udostępniania zmiennych wyjściowych, które opisano w ramach mapy SK.
- *Koszt uzyskania zmiennej* – wyliczono drogą dekompozycji całkowitego kosztu badania (wynikającego z Programu Badań Statystycznych Statystyki Publicznej) przy zastosowaniu macierzy transformacji *M*.

Przy powyższych założeniach, w wyniku przeprowadzonych w oparciu o macierz transformacji wyliczeń uzyskano szereg informacji uzupełniających charakteryzujących proces realizacji badań w ramach statystyki krótkookresowej. Warto zwrócić uwagę na to, że:

- W ramach badań DG-1 i C-01, przy wykorzystaniu 62 zmiennych wejściowych uzyskuje się w skali roku ponad 450 tysięcy sprawozdań obejmujących 9 obszarów tematycznych, 72 grupy zmiennych wynikowych oraz ponad 100 tys. udostępnianych liczb.
- Największą grupą zmiennych wyjściowych jest *Przeciętne miesięczne wynagrodzenie*, na którą składa się 19,5 tysiąca liczb.
- Pierwsze wyniki badań (*Przeciętne zatrudnienie* i *Przeciętne wynagrodzenie*) są dostępne już 16 dnia roboczego po zakończeniu miesiąca. 49% wykorzystywanych zmiennych to dane miesięczne, które stanowią aż 96,5% wszystkich wartości szacowanych w rozpatrywanych badaniach. Dodatkowo 2/3 wynikowych zmiennych miesięcznych posiada najwyższy priorytet.
- Średni koszt przypadający na jedno sprawozdanie wynosi 42 zł, ale przeciętna kwota potrzebna na oszacowanie jednej wartości to już 185 zł, przy czym rozbieżności pomiędzy zmiennymi wynikowymi są znaczne („najdroższa” jest *Kwota bazowa*, dla której koszt ustalenia jednej wartości oszacowano na 400 tys. zł, a „najtańsze” *Przeciętne zatrudnienie*, dla którego koszt oszacowania jednej wartości wynosi zaledwie 3 zł). Maksymalny koszt związany z oszacowaniem pojedynczej zmiennej wyjściowej, rozumianej jako grupa, dotyczy *Wskaźnika cen producentów na rynku krajowym*, zamyka się w kwocie 1,6 mln złotych.
- Do wyliczenia *Sprzedaży produkcji budowlano-montażowej wg rodzajów obiektów budowlanych* potrzeba aż 14 zmiennych wejściowych. Tyle samo, 14 zmiennych wejściowych wykorzystywanych jest tylko raz do wyliczenia zmiennych wyniko-

<sup>5</sup> Kwestia jakości danych sama w sobie jest niezwykle istotnym elementem. Jednakże ze względu na ograniczenia techniczne i brak wypracowanych jednoznacznych metod porównywania jakości danych zagadnienie to w artykule zostało pominięte.

wych, a 6 zmiennych wejściowych służy jedynie do oszacowania danych kwartalnych lub rocznych.

## 6. PODSUMOWANIE

Współcześnie, w dobie społeczeństwa informacyjnego, od statystyki publicznej wymaga się nie tylko rzetelnych i bieżących informacji o sytuacji społeczno-gospodarczej kraju, ale również umiejętności szybkiego reagowania na zmieniające się potrzeby użytkowników danych statystycznych. Konieczność zmian wymuszają również nowoczesne technologie teleinformatyczne oraz dostępność nowych źródeł danych. Stąd potrzeba dostosowywania prowadzonych badań, która będzie odpowiedzią na zmieniające się warunki otoczenia.

Mapa badań statystycznych jest jednym z narzędzi, które może być użyteczne w tych procesach. Dostarcza bowiem niezbędnej, zintegrowanej informacji o przeprowadzanych badaniach, która pozwala na identyfikację końcowych odbiorców, terminów udostępniania danych, a także opisuje wzajemne relacje pomiędzy poszczególnymi wielkościami. Dostarcza tym samym szybkiej odpowiedzi na pytania Co? Gdzie? Kiedy?

Oprócz waloru bieżącej użyteczności może mieć również wymiar motywacyjny. Współcześnie realizowane badania statystyczne niejednokrotnie są procesami wysoce złożonymi, wymagającymi zaangażowania dużych grup pracowników i precyzyjnego podziału zadań. Powoduje to atomizację działań i czasami brak świadomości ostatecznego celu wykonywanych zadań. Mapa badań statystycznych daje szersze spojrzenie na zagadnienia analizy i udostępniania danych, podnosi świadomość znaczenia wyników i celu prowadzonych badań. Ostatecznie daje odpowiedź na pytania dla kogo i do jakich celów realizowane są wspomniane badania, które są o tyle istotne, iż uzasadniają ogromną pracochłonność oraz koszty finansowe związane z pozyskaniem i przetworzeniem informacji.

Kolejnym ważnym elementem zastosowania mapy badań statystycznych jest możliwość wykorzystania jej do przeprowadzenia pogłębionych analiz. Przy pewnych założeniach pozwala ona na precyzyjne przyporządkowanie kosztów pozyskania poszczególnych zmiennych, szybkiego określenia zmiennych wejściowych niezbędnych do uzyskania priorytetowych informacji, możliwych do uzyskania przesunięć terminów udostępniania danych.

Odrębnym zagadnieniem, wymagającym dalszego rozwinięcia jest jakość danych. Jest ona kluczowa dla ustalenia realnej użyteczności danych statystycznych. Niejednoznaczność pojęcia i brak zintegrowanych modeli jej pomiaru utrudnia porównywanie badań przeprowadzanych różnymi metodami.

Znajomość skomplikowanych procesów przekształcania zmiennych wejściowych w wynikowe jest kluczowa w procesie optymalizacji badań. Dużą pomocą w tym zakresie jest prezentowanie relacji zachodzących między zmiennymi w postaci macierzy transformacji. Posiadanie macierzy transformacji oraz precyzyjnie opisanych i skwan-

tyfikowanych kryteriów jest warunkiem koniecznym do ustalenia optymalnego sposobu przetwarzania danych w badaniach statystycznych, które można uzyskać rozwiązując stosowane zadanie w postaci modelu matematycznego.

Dodatkowym atutem, wynikającym z zastosowania mapy badań, może być również możliwość dokładnego ustalenia możliwych zmian (np. terminów) w ścisłym powiązaniu z metodologią badania i zachowaniem priorytetów.

Innym zastosowaniem mapy badań statystycznych, związanym z optymalizacją, może być jej użycie w procesie integracji formularzy statystycznych.

Ważnym aspektem optymalizacji badań jest również redukcja obciążeń respondentów z tytułu uczestnictwa w badaniach statystycznych, zarówno tych obiektywnych odnoszonych do ponoszonych kosztów jak i subiektywnych – odczuwanych przez respondentów. Te ostatnie mierzone są relacją pomiędzy użytecznością dostępnej informacji statystycznej a wysiłkiem poniesionym na przygotowanie danych na potrzeby badań. Ocena przydatności informacji statystycznej determinuje chęć respondentów do współpracy, a więc dostarczania do statystyki wymaganych informacji. Trzeba pamiętać, iż w realiach gospodarki rynkowej, informacja statystyczna powinna być traktowana jak produkt, a przez to wypełniać kryteria kosztowe i jakościowe.

Aby mapa badań statystycznych pozostawała użytecznym narzędziem wspomagającym prace statystyków musi być nieustannie zasilana bieżącymi informacjami o procesach statystycznych, odbiorcach informacji oraz żądanym poziomie szczegółowości i terminowości danych. Pociąga to za sobą konieczność ciągłego śledzenia zmian nie tylko w regulacjach będących podstawą prawną badań, ale także w popycie na informację statystyczną. Odpowiedzią powinny być zmiany w zakresie publikowanych informacji, technikach upowszechniania i prezentacji wyników, a także, co bardzo ważne, w metodach pozyskiwania i wytwarzania danych.

*Urząd Statystyczny w Poznaniu*

#### LITERATURA

- [1] Eurostat (2006), *Methodology of short-term business statistics*. Interpretation and guidelines, European Communities, Luxembourg: Office for Official Publications of the European Communities.
- [2] Eurostat (2012), Short-term business statistics, [http://epp.eurostat.ec.europa.eu/portal/page/portal/short\\_term\\_business\\_statistics/introduction](http://epp.eurostat.ec.europa.eu/portal/page/portal/short_term_business_statistics/introduction)
- [3] GUS (2010), Założenia i wytyczne do system informacji o jednostkach prowadzących działalność gospodarczą na podstawie danych zawartych w sprawozdaniu miesięcznym DG-1 w 2010 r.
- [4] Kowalewski J., (2010), *Model optymalizacji badań statystycznych*, [w:] Prace Statystyczne i Demograficzne, red. I. Roeske-Słomka, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Zeszyt naukowy 133/2010.
- [5] Regulation (EC) No 1158/2005 of the European Parliament and of the Council amending Council Regulation (EC) No 1165/98 concerning short-term statistics.



- [6] Sundgren B., (2003), Developing and Implementing Statistical Metadata Systems. A Network of Excellence for Harmonising and Synthesising the Development of Statistical Metadata, EPROS Project Number IST-1999-29093, dostępny w Internecie: [www.epros.ed.ac.uk/metanet/deliverables/D6/IST-1999-29093-D6](http://www.epros.ed.ac.uk/metanet/deliverables/D6/IST-1999-29093-D6).
- [7] Stefanowicz B., (2007), *Informacyjne systemy zarządzania*. Przewodnik. Szkoła Główna Handlowa, Warszawa.
- [8] Willeboordse A. (ed.), (1997), *Handbook on Design and Implementation of Business Surveys*, Eurostat, Luxembourg.

#### WYKORZYSTANIE MAPY STATYSTYKI KRÓTKOOKRESOWEJ W PROCESIE ORGANIZACJI BADAŃ PRZEDSIĘBIORSTW PROWADZONYCH PRZEZ STATYSTYKĘ PUBLICZNĄ

##### Streszczenie

Statystyka krótkookresowa obejmuje zbieranie, przetwarzanie i dostarczanie szybkiej i wielodzinowej informacji w sferze gospodarczej. W ramach organizacji pracy w statystyce publicznej została ona przydzielona jako specjalizacja Urzędowi Statystycznemu w Poznaniu. W Polsce statystyka krótkookresowa realizowana jest głównie poprzez badanie DG-1, które charakteryzuje się tym, że jest dużym badaniem (ponad 30 tysięcy podmiotów) przeprowadzanym co miesiąc w bardzo krótkim okresie czasu, z rozbudowanym systemem przetwarzania (łączenie danych z innych badań, uogólnianie oraz przeliczenia na ceny stałe). Konieczność redukcji obciążeń respondentów oraz rozwój nowych technik informatycznych i statystycznych wymusza podejmowanie prób usprawnienia organizacji przeprowadzenia badania. Jednym z takich działań jest opracowanie mapy statystyki krótkookresowej, która pozwala na określenie w jaki sposób pozyskiwane dane są przekształcane w informację wynikową. Pozwala ona również na identyfikację odbiorców oraz czasu i kanałów dystrybucji poszczególnych informacji. Artykuł stanowi próbę ukazania możliwości wykorzystania mapy w ramach optymalizacji badań statystycznych.

**Słowa kluczowe:** statystyka krótkookresowa, statystyka przedsiębiorstw, badanie działalności gospodarczej, optymalizacja badań, termin publikacji danych statystycznych

#### USING A MAP OF SHORT TERM STATISTICS IN THE PROCESS OF ORGANIZING BUSINESS SURVEYS CONDUCTED BY PUBLIC STATISTICS

##### Abstract

Short-Term Statistics (STS) deals with collecting, compiling and providing quick and multi-domain information about economic activity. Within the organizational system of surveys in Polish official statistics, conducting STS surveys is the task of the Statistical Office in Poznań, which specializes in this field. The main source of STS data is the survey of economic activity, code-named DG-1, which covers over 30 thousands reporting units every month. The system of data processing in DG-1 is very complex and involves combining data from other surveys, generalizing and converting data to constant prices. Faced

with the necessity to reduce the response burden and given modern advances in computer and statistical techniques, we are obliged to continue our efforts at improving the organization of surveys. One step in this process is compiling “*A Map of Short- Term Statistics*”, which enables us to determine how input data are processed into output information and to identify users of statistical information, distribution channels and data release deadlines. This article provides an example of how the Map can be used in the process of optimizing statistical surveys.

**Key words:** short-term statistics, business statistics, survey of economic activity, users of statistical information, survey optimization, data release deadlines

MAGDALENA OKUPNIAK

## ZASTOSOWANIE ANALIZY BLOKOWEJ NA PRZYKŁADZIE BADANIA ZRÓŻNICOWANIA DOCHODÓW GOSPODARSTW DOMOWYCH

### 1. WSTĘP

Analiza blokowa należy do rodziny metod wielowymiarowej analizy danych statystycznych, często stosowanych w badaniach związanych z rolnictwem. Przeprowadza się ją głównie w celu oceny wpływu na wzrost i rozwój roślin różnorodnych czynników takich jak nasłonecznienie, wilgotność, nawozy mineralne, czy środki ochrony. Znalazła ona zastosowanie również w biologii, medycynie, inżynierii, fizyce, chemii, agrokulturze czy w eksperymentach przemysłowych. Zebrane dane jakie poddajemy badaniu podlegają podwójnej klasyfikacji. Z jednej strony jednostki podzielone są według poziomu czynnika (bądź rodzaju zastosowanego obiektu), z drugiej strony podzielone na możliwie jednorodne grupy (bloki). Stąd można wyróżnić analizę wewnątrz i między-blokową, które odpowiadają badaniu wpływu czynnika zastosowanego na jednostkach oraz badaniu wpływu podziału jednostek na grupy. Twórcą analizy blokowej jest brytyjski statystyk Ronald Aylmer Fisher, który w początkach XX wieku zdefiniował schemat układu bloków kompletnie zrandomizowanych. W tym szczególnym przypadku można również badać obiekty zastosowane na losowo wybranych jednostkach eksperymentalnych podzielonych na możliwie jednorodne grupy. Różnica w tym przypadku polega na tym, że grupy te są równoliczne (Oktaba, 1974).

### 2. CHARAKTERYSTYKA METODY ANALIZY BLOKOWEJ

#### ZAŁOŻENIA OGÓLNE

Zakładamy, że w doświadczeniu dysponujemy  $n$  jednostkami eksperymentalnymi, które podzielone są na  $b$  bloków (grup) tak, by możliwie najmniej różniły się między sobą w obrębie danego bloku. Ponadto każdy blok ma wielkość (pojemność)  $k_j$  dla  $j = 1, \dots, b$ . Na tak rozdysponowanych jednostkach stosujemy  $v$  obiektów, losowo przypisanych do danej jednostki eksperymentalnej. Tak by każdy z nich był użyty (replikowany)  $r_i$  razy ( $i = 1, \dots, v$ ). (Raghavarao, Padget, 2005) Rolę obiektów mogą przyjąć również poziomy stosowanego czynnika. Ponadto warto zwrócić uwagę na to, iż w tej metodzie nie jest wymagane założenie o normalności rozkładu badanej cechy.

Model doświadczenia jest następujący:

$$Y = \mu \mathbb{I}_n + \Delta' \underline{\tau} + \mathbf{D}' \underline{\beta} + \varepsilon, \quad (1)$$

gdzie:  $Y$  – wartość badanej zmiennej,  
 $\mu$  – średnia ogólna,  
 $\underline{\tau}$  – wektor efektów obiektowych,  
 $\underline{\beta}$  – wektor efektów blokowych,  
 $\varepsilon$  – wektor błędów losowych,  
 $\Delta = (\delta_{ws})$ , gdzie:

$$\delta_{ws} = \begin{cases} 0 & s\text{-ta obserwacja nie jest związana z obiektem } w, \\ 1 & s\text{-ta obserwacja dotyczy } w\text{-tego obiektu.} \end{cases}$$

$\mathbf{D} = (\gamma_{zs})$ , gdzie:

$$\gamma_{zs} = \begin{cases} 0 & s\text{-ta obserwacja nie pochodzi z } z\text{-tego bloku,} \\ 1 & s\text{-ta obserwacja pochodzi z } z\text{-tego bloku.} \end{cases}$$

Warto zwrócić uwagę na to, że w doświadczeniu prowadzonym według metody układu bloków nie ma interakcji między efektem blokowym, a efektem obiektowym. Macierz  $\Delta$  jest wymiaru  $v \times n$ , natomiast macierz  $\mathbf{D}$  wymiaru  $b \times n$ .

O błędach losowych  $\varepsilon$  zakładamy, że

- 1)  $\varepsilon$  jest wektorem nieskorelowanych błędów losowych,
- 2)  $\varepsilon \sim N(0_n, \sigma^2 I_n)$ .

Ponadto należy zdefiniować pojęcie kontrastu:

Liniową funkcję parametryczną  $\ell' \underline{\rho}$  wektora  $\underline{\rho} = [\rho_1, \rho_2, \dots, \rho_d]$  nazywamy kontrastem, jeśli  $\ell' \mathbb{I}_d = 0$ . Kontrast nazywamy kontrastem elementarnym, jeśli  $\ell$  ma tylko dwa niezerowe elementy i są one równe -1 i 1 (Raghavarao, Padget, 2005).

#### ANALIZA WEWNĄTRZBLOKOWA

Badanie prowadzone według analizy wewnątrzblokowej polega na tym, że eliminujemy efekt podziału jednostek eksperymentalnych na bloki i rozważamy problem, czy efekty pochodzące od zastosowanych obiektów mają wpływ na wyniki doświadczenia. Ze względu na nieestymowalność każdego parametru z osobna, analizujemy następujący układ hipotez:

$$\begin{aligned} H_0 : \tau_1 - \tau_v &= \tau_2 - \tau_v = \dots = \tau_{v-1} - \tau_v = 0 \\ H_1 : &\sim H_0 \end{aligned} \quad (2)$$

co możemy zrobić powołując się na następujące twierdzenie (Raghavarao, Padget, 2005):

Wszystkie elementarne kontrasty efektów obiektowych są estymowalne wtedy i tylko wtedy, gdy  $\text{rz}(C) = v - 1$ .

Powyższy układ hipotez możemy weryfikować za pomocą statystyki Walda (Raghavarao, Padget, 2005):

$$F = \frac{Q' C^- Q /_{v-1}}{R_0^2 /_{n-b-v+1}}, \quad (3)$$

gdzie:

$$C = R - \mathbf{N} \mathbf{K}^{-1} \mathbf{N}',$$

$C^-$  – g-odwrotność macierzy  $C$ ,

$$Q = T - \mathbf{N} \mathbf{K}^{-1} B,$$

$$R_0^2 = SS_T - SS_B - SS_{T|B}.$$

Macierz  $R$  jest macierzą diagonalną,  $R = \text{diag}(r_1, r_2, \dots, r_v)$ . Podobnie macierz  $K$ , gdzie  $K = \text{diag}(k_1, k_2, \dots, k_b)$ . Macierz incydencji  $\mathbf{N}$  jest wymiaru  $v \times b$ . Dalsze oznaczenia są następujące:

$$B = (B_1, \dots, B_b)', \text{ gdzie } B_j = \sum_{i=1}^v Y_{ij}$$

$$T = (T_1, \dots, T_v)', \text{ gdzie } T_i = \sum_{j=1}^b Y_{ij}$$

#### ANALIZA MIĘDZYBLOKOWA

Badanie prowadzone według analizy międzyblokowej polega na tym, że eliminujemy efekty pochodzące od zastosowanych obiektów i rozważamy problem, czy efekty wywodzące się z podziału jednostek eksperymentalnych na bloki mają wpływ na wyniki doświadczenia. Również w tym wypadku, ze względu na nieestymowalność każdego parametru z osobna, analizujemy poniższy układ hipotez:

$$\begin{aligned} H_0 : \beta_1 - \beta_b &= \beta_2 - \beta_b = \dots = \beta_{b-1} - \beta_b = 0 \\ H_1 : &\sim H_0 \end{aligned} \quad (4)$$

Wykorzystując poniższe twierdzenie (Raghavarao, Padget, 2005), które rozwiązuje problem nieestymowalności parametrów:

Wszystkie elementarne kontrasty efektów blokowych są estymowalne wtedy i tylko wtedy, gdy  $\text{rz}(D^*) = b - 1$ .

Hipotezę zerową możemy weryfikować przy pomocy statystyki Walda (Raghavarao, Padget, 2005):

$$F = \frac{P(D^*)^- P /_{b-1}}{R_0^2 /_{n-b-v+1}}, \quad (5)$$

gdzie:

$$D^* = K - \mathbf{N}' \mathbf{R}^{-1} \mathbf{N},$$

$$\begin{aligned} (D^*)^- &- g\text{-odwrotność macierzy } D^*, \\ P &= B - \mathbf{N}'R^{-1}T, \\ R_0^2 &= SS_T - SS_{Tr} - SS_{B|Tr}. \end{aligned}$$

### 3. SCHEMAT PRZEPROWADZONEGO BADANIA

Dochód jest jednym z zasadniczych czynników służących do określania poziomu zamożności i sposobu życia, pozwalający zaspokoić potrzeby gospodarstwa domowego. Duży wpływ na wysokość dochodu mają przede wszystkim liczebność gospodarstwa oraz jego skład demograficzny, który różnicuje poziom i strukturę potrzeb członków rodziny. Często w międzynarodowych analizach porównawczych zamożności, bazujących na dochodach gospodarstw domowych, stosuje się dochód ekwiwalentny na osobę w gospodarstwie, a nie powszechnie stosowany dochód na osobę. Skala ekwiwalentności koryguje dochód na osobę. Uwzględnia fakt, iż wraz ze wzrostem liczby osób w gospodarstwie domowym jego dochody nie rosną liniowo. Ponadto, bierze pod uwagę nie tylko liczbę osób wchodzących w skład gospodarstwa, ale także ich wiek. W badaniu posłużono się zmodyfikowaną skalą OECD liczoną według poniższego wzoru (GUS, 2011):

$$e = 1 + 0,5(LA - 1) + 0,3LD,$$

gdzie  $LA$  to liczba osób w wieku co najmniej 14 lat, a  $LD$  to liczba osób wchodzących w skład gospodarstwa w wieku co najmniej 13 lat.

Celem analizy była weryfikacja hipotez o wpływie podziału gospodarstw domowych Polski na województwa i klasy miejscowości na wielkość średniego dochodu ekwiwalentnego i średniego dochodu na osobę w 2011 roku w oparciu o metodę analizy blokowej.

W badaniu wykorzystano dane pochodzące z Diagnozy Społecznej 2011 przeprowadzonej w Polsce przez Radę Monitoringu Społecznego. Badanie przyjmuje formę badania panelowego. Jest ono prowadzone od 2000 roku. W 2011 roku przeprowadzono je po raz szósty. Ankietyzacja respondentów miała miejsce w marcu, w celu uniknięcia efektu sezonowości. Opublikowany raport ma na celu charakterystykę polskiego społeczeństwa 22 lata po zmianie systemu, 7 lat po wstąpieniu Polski do Unii Europejskiej i 12 lat po pierwszym badaniu w ramach tego samego projektu. (Czapiński, Panek, 2012) Pytania zawarte w ankietach dotyczyły gospodarstw domowych oraz ich członków i odnosiły się między innymi do zachowań osób wchodzących w skład gospodarstwa domowego, ich postaw, a także dochodu, zasobności materialnej, kredytów, edukacji, leczenia. Baza danych zawiera informacje o 12387 gospodarstwach domowych.

Raport Diagnozy Społecznej 2011 (Czapiński, Panek, 2012) dostarcza dużo interesujących informacji dotyczących dochodów gospodarstw domowych dla poszczególnych województw. Na przykład potwierdza znany fakt, że w województwie mazowieckim obserwuje się najwyższe dochody netto na osobę zarówno w przeliczeniu na

jednostkę ekwiwalentną (czyli w przeliczeniu według zmodyfikowanej skali ekwiwalentności OECD), czy na gospodarstwo domowe. Natomiast najniższe dochody w województwie lubelskim. Interesujące wnioski można również wyciągnąć na temat dochodów netto w podziale na klasy miejscowości: wraz ze wzrostem liczby mieszkańców rosną dochody netto na gospodarstwo domowe, na osobę jak i na jednostkę ekwiwalentną.

Na potrzeby badania zostały przyjęte następujące definicje:

**Obiekty** – klasy miejscowości

W badaniach prowadzonych przez GUS, jak i w Diagnostyce Społecznej, określone w sześciu kategoriach. Warianty klas miejscowości:

1. miasta o liczbie mieszkańców 500 tys. i więcej,
2. miasta o liczbie mieszkańców 200-500 tys.,
3. miasta o liczbie mieszkańców 100-200 tys.,
4. miasta o liczbie mieszkańców 20-100 tys.,
5. miasta o liczbie mieszkańców poniżej 20 tys.,
6. wieś.

**Bloki** – Województwa

W analizie jako zmienne badane przyjęto:

1. średni miesięczny dochód ekwiwalentny netto gospodarstwa domowego
2. średni miesięczny dochód netto na osobę w gospodarstwie domowym

W tabelach 1 oraz 2 przedstawiono dane charakteryzujące przebadaną zbiorowość pod względem średniego dochodu ekwiwalentnego netto na osobę oraz średniego dochodu netto gospodarstwa domowego na osobę. Braki wartości oznaczają braki informacji, że w odniesieniu do danej kategorii nie udzielono ankietom odpowiedzi bądź też w obrębie danego województwa nie ma miejscowości należących do danej klasy.

Tabela 1.

Średni dochód ekwiwalentny netto na osobę w badanych gospodarstwach domowych według klasy miejscowości zamieszkania i województw

| Średni dochód ekwiwalentny netto na osobę (zł) |                                     |         |         |         |            |         |
|------------------------------------------------|-------------------------------------|---------|---------|---------|------------|---------|
| Województwo                                    | Klasa miejscowości zamieszkania     |         |         |         |            |         |
|                                                | miasta o liczbie mieszkańców (tys.) |         |         |         |            | wieś    |
|                                                | 500 i więcej                        | 200-500 | 100-200 | 20-100  | poniżej 20 |         |
| Dolnośląskie                                   | 2329,07                             | .       | 1706,93 | 1766,58 | 1566,61    | 1456,73 |
| Kujawsko-pomorskie                             | .                                   | 2139,47 | 1331,19 | 1304,97 | 1369,91    | 1358,93 |
| Lubelskie                                      | .                                   | 2235,49 | .       | 1709,96 | 1635,82    | 1161,71 |
| Lubuskie                                       | .                                   | .       | 2066,32 | 1765,27 | 1531,17    | 1379,55 |
| Łódzkie                                        | 2010,86                             | .       | .       | 1657,45 | 1506,43    | 1346,54 |
| Małopolskie                                    | 2259,96                             | .       | 1883,38 | 1817,58 | 1660,55    | 1473,03 |
| Mazowieckie                                    | 2952,28                             | 1538,25 | 1985,45 | 1696,79 | 2100,87    | 1441,08 |
| Opolskie                                       | .                                   | .       | 2099,01 | 1657,39 | 1558,12    | 1593,01 |

cd. Tabela 1.

|                     |         |         |         |         |         |         |
|---------------------|---------|---------|---------|---------|---------|---------|
| Podkarpackie        | .       | .       | 1868,37 | 1755,93 | 1794,21 | 1201,60 |
| Podlaskie           | .       | 635,84  | .       | 1669,73 | 1950,75 | 1289,40 |
| Pomorskie           | .       | 2177,79 | 1438,65 | 1896,71 | 1688,98 | 1553,60 |
| Śląskie             | .       | 2493,12 | 2015,51 | 1861,44 | 1666,24 | 1614,93 |
| Świętokrzyskie      | 1300,00 | 1676,83 | .       | 1578,74 | 1510,31 | 1279,96 |
| Warmińsko-mazurskie | .       | .       | 1687,19 | 1743,06 | 1427,50 | 1380,62 |
| Wielkopolskie       | 2344,70 | .       | 1568,84 | 1772,48 | 8,51    | 1473,92 |
| Zachodnio-pomorskie | .       | 2265,28 | 2053,07 | 1664,83 | 1715,49 | 1223,96 |

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Tabela 2.

Średni dochód netto gospodarstwa domowego na osobę w badanych gospodarstwach domowych według klasy miejscowości zamieszkania i województw

| Średni dochód netto gospodarstwa domowego na osobę (zł) |                                     |         |         |         |            |         |
|---------------------------------------------------------|-------------------------------------|---------|---------|---------|------------|---------|
| Województwo                                             | Klasa miejscowości zamieszkania     |         |         |         |            |         |
|                                                         | Miasta o liczbie mieszkańców (tys.) |         |         |         |            | wieś    |
|                                                         | 500 i więcej                        | 200-500 | 100-200 | 20-100  | poniżej 20 |         |
| Dolnośląskie                                            | 1734,17                             | .       | 1181,22 | 1344,15 | 1196,16    | 1029,65 |
| Kujawsko-pomorskie                                      | .                                   | 1583,93 | 931,66  | 980,80  | 1012,19    | 942,14  |
| Lubelskie                                               | .                                   | 1776,31 | .       | 1229,77 | 1217,47    | 828,38  |
| Lubuskie                                                | .                                   | .       | 1536,46 | 1343,17 | 1120,37    | 944,33  |
| Łódzkie                                                 | 1572,74                             | .       | .       | 1231,04 | 1071,29    | 985,22  |
| Małopolskie                                             | 1755,71                             | .       | 1198,33 | 1367,53 | 1180,89    | 1014,59 |
| Mazowieckie                                             | 2250,20                             | 1185,55 | 1492,99 | 1261,27 | 1458,72    | 1004,65 |
| Opolskie                                                | .                                   | .       | 1492,95 | 1202,81 | 1213,74    | 1113,85 |
| Podkarpackie                                            | .                                   | .       | 1504,76 | 1275,48 | 1184,66    | 804,24  |
| Podlaskie                                               | .                                   | 1242,11 | .       | 1182,15 | 1366,78    | 880,61  |
| Pomorskie                                               | .                                   | 1587,28 | 1115,18 | 1439,30 | 1231,47    | 1085,55 |
| Śląskie                                                 | .                                   | 1883,86 | 1576,45 | 1374,06 | 1248,17    | 1141,05 |
| Świętokrzyskie                                          | 1300,00                             | 1246,24 | .       | 1129,28 | 1074,24    | 893,40  |
| Warmińsko-mazurskie                                     | .                                   | .       | 1263,93 | 1304,95 | 974,25     | 965,61  |
| Wielkopolskie                                           | 1814,90                             | .       | 1086,57 | 1310,09 | 1130,66    | 1006,39 |
| Zachodnio-pomorskie                                     | .                                   | 1703,05 | 1500,31 | 1228,13 | 1270,35    | 852,36  |

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Przed przeprowadzeniem badania sprawdzono jak kształtuje się rozkład gospodarstw domowych ze względu na średni miesięczny dochód netto na osobę. Na podstawie podstawowych charakterystyk (tab. 3) jak i histogramu (rys. 1) można wnioskować



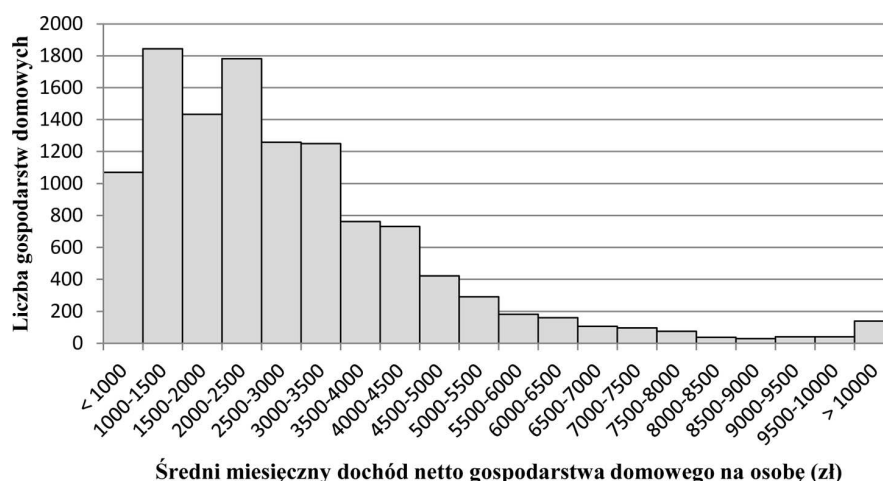
o silnej prawostronnej asymetrii rozkładu, a więc zdecydowanej przewadze gospodarstw domowych o zarobkach do 3613 zł. Największą różnicę w średnim dochodzie netto na osobę między największą i najmniejszą klasą miejscowości zaobserwować można w województwie mazowieckim, gdzie wynosi ona ponad 1200 zł. Natomiast najmniejszą w województwie świętokrzyskim, równą blisko 400 zł.

Tabela 3.

Charakterystyki rozkładu średniego miesięcznego dochodu netto gospodarstw domowych na osobę (zł)

| Statystyki opisowe        |         |
|---------------------------|---------|
| Dominanta                 | 1326,71 |
| Błąd standardowy          | 23,22   |
| Mediana                   | 2428,27 |
| Kwartyl 1                 | 1507,94 |
| Kwartyl 3                 | 3613,02 |
| Odchylenie ćwiartkowe     | 1052,54 |
| Pozycyjny wsp. zmienności | 43%     |
| Pozycyjny wsp. asymetrii  | 1,12    |
| Zakres                    | 52220   |
| Minimum                   | 100     |
| Maksimum                  | 52320   |

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011



Rysunek 1. Rozkład gospodarstw domowych ze względu na średni miesięczny dochód netto na osobę (zł)

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Tabela 4.

Wyniki analizy wewnątrzblokowej dla średniego miesięcznego dochodu ekwiwalentnego netto oraz średniego miesięcznego dochodu netto gospodarstwa domowego na osobę

|                          | Zmienna i postać hipotezy zerowej                                                                                              | Wyniki analizy                                          |
|--------------------------|--------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------|
| Analiza wewnątrz-blokowa | Średni miesięczny dochód ekwiwalentny netto na osobę<br>$H_0 : \alpha_1 - \alpha_6 = \alpha_2 - \alpha_6 = \dots = 0$          | F = 15,1458<br>p-value = 0,0000<br>df1 = 5<br>df2 = 53  |
|                          | Średni miesięczny dochód netto gospodarstwa domowego na osobę<br>$H_0 : \alpha_1 - \alpha_6 = \alpha_2 - \alpha_6 = \dots = 0$ | F = 29,1003<br>p-value = 0,00000<br>df1 = 5<br>df2 = 53 |

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Opierając się na przedstawionych wyżej wynikach można powiedzieć, że podział gospodarstw domowych pod względem klas miejscowości ma wpływ zarówno na wysokość średniego miesięcznego dochodu ekwiwalentnego netto na osobę jak i na wysokość średniego miesięcznego dochodu netto gospodarstwa domowego na osobę.

Tabela 5.

Wyniki analizy międzyblokowej dla średniego miesięcznego dochodu ekwiwalentnego netto oraz średniego miesięcznego dochodu netto gospodarstwa domowego na osobę

|                        | Zmienna i postać hipotezy zerowej                                                                                                | Wyniki analizy                                           |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|
| Analiza między-blokowa | Średni miesięczny dochód ekwiwalentny netto na osobę<br>$H_0 : \beta_1 - \beta_{16} = \beta_2 - \beta_{16} = \dots = 0$          | F = 1,68843<br>p-value = 0,08210<br>df1 = 15<br>df2 = 53 |
|                        | Średni miesięczny dochód netto gospodarstwa domowego na osobę<br>$H_0 : \beta_1 - \beta_{16} = \beta_2 - \beta_{16} = \dots = 0$ | F = 1,47680<br>p-value = 0,14810<br>df1 = 15<br>df2 = 53 |

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Na podstawie powyższych wyników można stwierdzić, iż podział gospodarstw domowych pod względem województw nie ma wpływu na wysokość średniego miesięcznego dochodu ekwiwalentnego netto na osobę. Podobne wnioski można sformułować dla średniego miesięcznego dochodu netto gospodarstwa domowego na osobę, na którego wysokość nie ma wpływu podział gospodarstw domowych pod względem województw. (p-value oznacza poziom istotności, df1 oraz df2 liczbę stopni swobody dla statystyki F)

Dla porównania otrzymanych wyników przeprowadzono test jednoczynnikowej analizy wariancji (ANOVA) (por. tabele 6 i 7).

W rezultacie można stwierdzić, że zarówno wyniki analizy blokowej jak i jednoczynnikowej analizy wariancji pokrywają się. Wartości statystyk empirycznych, graniczny poziom istotności jak i stopnie swobody są takie same dla analizy wewnątrzblo-

Tabela 6.

Wyniki testu ANOVA dla średniego dochodu netto gospodarstwa domowego na osobę

| Źródło             | Suma kwadratów | Stopnie swobody | Średnia suma kwadratów | F       | Istotność |
|--------------------|----------------|-----------------|------------------------|---------|-----------|
| Model skorygowany  | 3851910,084    | 20              | 192595,504             | 8,524   | 0,000     |
| Stała              | 105112197,197  | 1               | 105112197,19           | 4652,02 | 0,000     |
| Województwo        | 500524,599     | 15              | 33368,307              | 1,477   | 0,148     |
| Klasa miejscowości | 3287599,430    | 5               | 657519,886             | 29,100  | 0,000     |
| błąd               | 1197532,814    | 53              | 22594,959              |         |           |
| Ogółem             | 118475919,276  | 74              |                        |         |           |
| Ogółem skorygowane | 5049442,898    | 73              |                        |         |           |

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Tabela 7.

Wyniki testu ANOVA dla średniego dochodu ekwiwalentnego netto na osobę

| Źródło             | Suma kwadratów | Stopnie swobody | Średnia suma kwadratów | F       | Istotność |
|--------------------|----------------|-----------------|------------------------|---------|-----------|
| Model skorygowany  | 5374872,916    | 20              | 268743,646             | 5,093   | 0,000     |
| Stała              | 197338923,890  | 1               | 197338923,89           | 3739,64 | 0,000     |
| Województwo        | 1336459,968    | 15              | 89097,331              | 1,688   | 0,082     |
| Klasa miejscowości | 3996168,106    | 5               | 799233,621             | 15,146  | 0,000     |
| błąd               | 2796779,892    | 53              | 52769,432              |         |           |
| Ogółem             | 225828476,852  | 74              |                        |         |           |
| Ogółem skorygowane | 8171652,808    | 73              |                        |         |           |

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

kowej (zaznaczone kursywą) oraz analizy międzyblokowej. Należałoby zatem sprawdzić, czy w ogólności obie postacie statystyk prowadzą do tych samych wyników.

#### 4. WNIOSKI

- Podział gospodarstw domowych według województw nie ma wpływu na średni dochód netto gospodarstwa domowego na osobę oraz średni miesięczny dochód ekwiwalentny netto na osobę.
- W odniesieniu do podziału jednostek według klas miejscowości, podział ten ma wpływ na wartości badanych zmiennych.
- Wyniki przeprowadzonych badań analizy blokowej i jednoczynnikowej analizy wariancji prowadzą do takich samych wartości statystyk testowych.

- Metodę analizy blokowej można stosować w przypadku, gdy nie jest spełnione założenie o rozkładzie normalnym badanej zmiennej wymagane dla jednoczynnikowej analizy wariancji.
- Różnica między metodą układu bloków, a jednoczynnikową analizą wariancji tkwi w tym, że w analizie blokowej jedna obserwacja może dotyczyć kilku jednostek. Oznacza to otrzymanie macierzowej postaci wyjściowej ogólnego modelu doświadczenia. Parametry nie są estymowalne z osobna, zatem zmienia się postać hipotezy zerowej, w której analizujemy kontrasty (różnice parametrów).

*Uniwersytet Ekonomiczny w Poznaniu*

#### LITERATURA

- [1] Czapiński J., Panek T. (red.), (2012), *Diagnoza społeczna 2011*, Warszawa, <http://www.diagnoza.com> [dostęp: 15.02.2012].
- [2] GUS, (2011), *Budżety gospodarstw domowych w 2010r.*, Warszawa, s. 19.
- [3] Krzyśko M., (2000), *Wielowymiarowa analiza statystyczna*, UAM, Poznań.
- [4] Oktaba W., (1974), *Elementy statystyki matematycznej i metodyka doświadczałnictwa*, Państwowe Wydawnictwo Naukowe PWN, Warszawa.
- [5] Raghavarao D., Padgett L.V., (2005), *Block designs*, World Scientific, Singapore.
- [6] Rao R.C., (1982), *Modele liniowe statystyki matematycznej*, Państwowe Wydawnictwo Naukowe PWN, Warszawa.

#### ZASTOSOWANIE ANALIZY BLOKOWEJ NA PRZYKŁADZIE BADANIA ZRÓŻNICOWANIA DOCHODÓW GOSPODARSTW DOMOWYCH

##### Streszczenie

Analiza blokowa to rodzaj wielowymiarowej analizy danych często stosowany w badaniach związanych z rolnictwem. Badania te zwykle przeprowadza się w celu oceny wpływu na wzrost i rozwój roślin różnych czynników takich jak nawozy mineralne, czy środki ochrony roślin. Głównym jej prekursorem był brytyjski statystyk R. A. Fisher, który na początku XX wieku sformułował schemat układu bloków kompletnie zrandomizowanych. Jest to szczególnie przypadek układu bloków, w którym badamy wpływ danego obiektu zastosowanego na losowo wybranych jednostkach eksperymentalnych podzielonych na równoliczne i możliwie jednorodne grupy.

Celem pracy jest zaprezentowanie wyników zastosowania analizy blokowej w badaniu dochodu ekwiwalentnego gospodarstw domowych. Jako źródło informacji wykorzystano dane zgromadzone przez Radę Monitoringu Społecznego w ramach Diagnozy Społecznej 2011. Baza zawierała informacje o 12387 gospodarstwach domowych. Każde z nich zostało opisane za pomocą 259 zmiennych. W przeprowadzonym badaniu, w którym dokonano analizy wewnątrz- i międzyblokowej, uwzględnione zostały zmienne takie jak: położenie geograficzne gospodarstwa domowego, liczba osób wchodzących w skład gospodarstwa oraz klasa miejscowości.

**Słowa kluczowe:** analiza wewnątrzblokowa, analiza międzyblokowa, dochód ekwiwalentny

## APPLICATION OF THE BLOCK ANALYSIS TO THE INVESTIGATION OF HOUSEHOLD INCOME DIVERSITY

### A b s t r a c t

Analysis of the block designs is one of multivariate data analysis methods, which is often used in a agricultural. This type of research is usually carried out to assess the impact of various factors on the growth and development of plants, such as fertilizers pesticides and others. The main precursor of block designs was a British statistician R. A. Fisher, who in early twentieth century formulated a randomized block designs. It is a particular kind of the block designs, in which we examine the impact of an object used to randomly selected experimental units divided into equinumerous and possibly homogenous groups.

The aim of this study is to present results of applying the block analysis in a equivalent household income research. Data was obtained from the Council for Social Monitoring – The Social Diagnosis 2011. The base contained of 12387 households, which were described by 259 variables. The variables which were used in a intra-and inter-block analysis research were: administrative location of household, number of people included into the household and class of the village.

**Key words:** intrablock analysis, interblock analysis, equivalent household

ŁUKASZ WAWROWSKI

## ANALIZA UBÓSTWA W PRZEKROJU POWIATÓW W WOJEWÓDZTWIE WIELKOPOLSKIM Z WYKORZYSTANIEM METOD STATYSTYKI MAŁYCH OBSZARÓW

### 1. WSTĘP

Zjawisko ubóstwa<sup>1</sup> kojarzone jest zwykle z problemem ludzi bezdomnych czy też żebrzących na ulicy. Tymczasem jest to problem dużo bardziej złożony i nie do końca widoczny gołym okiem. Szacuje się, że 13,6% populacji Ziemi cierpi głód, a 1,35 miliarda ludzi żyje za mniej niż 1,25\$ dziennie (Hungry Notes, 2012). W Polsce wybrane statystyki na temat ubóstwa kształtują się następująco: w skrajnym ubóstwie<sup>2</sup> żyje 5,7% osób w gospodarstwach domowych, wskaźnik zagrożenia ubóstwem wynosi 17,3%, a aż 59% Polaków żyje poniżej minimum socjalnego<sup>3</sup> (GUS, 2010). W związku z niepokojącymi statystykami eliminacja skrajnego ubóstwa i głodu do 2015 roku jest głównym celem Banku Światowego oraz pierwszym zadaniem na liście Milenijnych Celów Rozwoju ONZ (UN, 2000). Niestety już teraz wiadomo, że tego celu nie uda się zrealizować.

Pierwszy krok w kierunku ograniczenia tak niekorzystnego zjawiska jakim jest niedostatek stanowi identyfikacja gospodarstw znajdujących się w strefie ubóstwa. Celem artykułu jest próba wykorzystania metod statystyki małych obszarów do szacowania wskaźnika zagrożenia ubóstwem na przykładzie powiatów województwa wielkopolskiego. W badaniu zostały wykorzystane estymatory: bezpośredni oraz pośredni. Przeprowadzona analiza pozwoliła na wskazanie ich własności i ocenę co zostało opisane szerzej w niniejszej pracy.

### 2. ISTOTA UBÓSTWA

Pojęcie ubóstwa rozumiane jest bardzo szeroko. W literaturze naukowej można znaleźć wiele definicji tego zjawiska, które dodatkowo je bardzo różnicują. Według jednej z definicji, sformułowanej przez Jana Drewnowskiego (1977), ubóstwo występuje wówczas, kiedy ludzie muszą zaspokajać swoje potrzeby na niższym poziomie

<sup>1</sup> Synonimami tego słowa są także bieda oraz niedostatek. W pracy będą stosowane wymiennie.

<sup>2</sup> Granica ubóstwa skrajnego ustalana jest przez Instytut Pracy i Spraw Socjalnych jako kwota pozwalająca na zaspokojenie tylko podstawowych potrzeb, a osoby poniżej tej granicy narażone są na biologiczne wyniszczenie.

<sup>3</sup> Jest to kwota, która uprawnia do ubiegania się o przyznanie świadczenia pieniężnego z systemu pomocy społecznej.

niż uznają to za wystarczające, natomiast indyjski ekonomista i laureat Nagrody Banku Szwecji im. Alfreda Nobla w dziedzinie ekonomii, Amartya Sen (1997) uważa, że „ubóstwo jest niezdolnością do skutecznego funkcjonowania w społeczeństwie”. Są to definicje bardzo ogólne i należy zauważyć, że postrzeganie tego zjawiska zależy od miejsca i czasu. Ubogi europejczyk zdecydowanie różni się od ubogiego mieszkańca Indii. Ponadto, biorąc pod uwagę czas można zauważyć, że istnieją gospodarstwa domowe, które aktualnie należy uznać za znajdujące się w strefie ubóstwa – mimo że kilka lat temu były poza nią (Panek, 2011). Przez lata starano się zdefiniować ubóstwo w taki sposób, aby to zjawisko stało się mierzalne. Krok w tym kierunku uczyniła Rada Ministrów Wspólnoty Europejskiej ustanawiając definicję mówiącą, że „ubóstwo odnosi się do osób, rodzin lub grup osób, których zasoby (materialne, kulturowe i społeczne) są ograniczone w takim stopniu, że poziom ich życia obniża się poza akceptowane minimum w kraju zamieszkania” (por. ECC, 1985).

## 2.1 POMIAR UBÓSTWA

Zjawisko ubóstwa jest zwykle utożsamiane z pewnym poziomem dochodów lub wydatków, który nie pozwala na prawidłowe funkcjonowanie w społeczeństwie. Istnieje jednak wiele przesłanek wskazujących na to, że w analizie ubóstwa należy także uwzględniać czynniki pozadochodowe. Wynika z tego pierwszy podział na podejścia związane z pomiarem tego zjawiska:

- podejście klasyczne oparte na analizie dochodów i wydatków,
- podejście wielowymiarowe uwzględniające także czynniki pozadochodowe.

Zarówno w podejściu klasycznym, jak i wielowymiarowym można wyróżnić kolejny podział według sposobu mierzenia ubóstwa. W ujęciu obiektywnym ocena zaspokojenia potrzeb badanych jednostek odbywa się niezależnie od ich osobistych wartościowań i jest dokonywana przez ekspertów. Z drugiej strony, w ujęciu subiektywnym jednostki same oceniają poziom zaspokojenia swoich potrzeb (Panek, 2009). Ekonomisci najczęściej bazują w swoich pracach na badaniach przeprowadzanych w ujęciu obiektywnym, jako główny argument podając, że jednostki nie zawsze są najlepsze w orzekaniu czego naprawdę potrzebują (Kaplan i Makoka, 2005).

W celu kompleksowej analizy poza metodą pomiaru należy jeszcze zdefiniować sposób pojmowania ubóstwa. Można je postrzegać w dwojaki sposób:

- absolutny – za ubogie uważana są te jednostki, które nie mają możliwości zaspokojenia podstawowych potrzeb,
- względny – za ubogie uważane są jednostki, których poziom życia jest dużo niższy od poziomu życia pozostałych członków społeczeństwa.

W zależności od wybranego sposobu pojmowania ubóstwa zmieniają się możliwości wyeliminowania tego zjawiska. Podejście absolutne charakteryzuje się tym, że przez krótkoterminowy wzrost dochodów lub znaczący wzrost gospodarczy można doprowadzić do prawie całkowitej likwidacji ubóstwa. Natomiast w podejściu względnym

ubóstwa praktycznie nie da się wyeliminować, wzrost dochodów może nie spowodować zmiany odsetka ubogich (GUS, 2011b).

## 2.2 MIERNIK ZAGROŻENIA UBÓSTWEM

Wyznaczenie granicy ubóstwa opiera się na analizie poziomu dochodów lub wydatków gospodarstw domowych. Na tej podstawie wyznacza się linię ubóstwa będącą wartością krytyczną wyżej wymienionych miar. Jednostka uznawana jest za ubogą wówczas, gdy poziom jej dochodów lub wydatków znajduje się poniżej tej granicy. Należy jednak pamiętać o niedoskonałości tych miar (Panek, 2011). W kategorii dochodu nie są ujęte przychody pozagotówkowe, a także majątek trwały. Ponadto zauważalna jest tendencja do zaniżania tej wartości poprzez zatajanie dochodów pochodzących z tzw. szarej strefy. Z kolei w przypadku wydatków obserwuje się radykalne zaniżanie przez badane jednostki wydatków na używki takie jak: tytoń i alkohol (Panek i Szulc, 2004).

## 3. STATYSTYKA MAŁYCH OBSZARÓW

Badania reprezentacyjne z zakresu statystyki społecznej prowadzone między innymi przez GUS dostarczają informacji o prawie każdej dziedzinie życia społecznego i gospodarczego. Mimo, że badany jest wówczas tylko fragment populacji, takie badania są bardzo kosztowne, a wyniki najczęściej publikuje się na poziomie województw. Zejście na niższy poziom agregacji spowodowałoby pojawienie się zbyt wysokich błędów średniokwadratowych. Tymczasem popyt na informację dla małych jednostek terytorialnych z roku na rok wzrasta. Jednym ze sposobów rozwiązania tego problemu może być zastosowanie w badaniach metod zaproponowanych przez SMO. Statystyka małych obszarów jest narzędziem pozwalającym na otrzymanie wiarygodnych informacji na podstawie nielicznej próby, przy wykorzystaniu wszystkich dostępnych informacji dodatkowych. Wyróżnia się dwie podstawowe techniki estymacji: estymację bezpośrednią i pośrednią.

### 3.1 ESTYMACJA BEZPOŚREDNIA

Estymator bezpośredni dla domen<sup>4</sup> wykorzystuje jedynie informacje pochodzące z badania reprezentacyjnego dla danej domeny. Cechą zachęcającą do jego stosowania jest fakt, że przy odpowiedniej liczebności próby jest on nieobciążony. Za jego wadę można uznać to, że mała liczebność próby implikuje wysokie wartości wariancji w przypadku szacowania wartości parametrów dla domen mniejszych od tych, dla któ-

<sup>4</sup> Określenie *domena* odnosi się do przekrojów terytorialnych lub rzeczowych np. powiat, płeć, grupy wieku.



rych zaprojektowano badanie. Najczęściej stosowanym estymatorem bezpośrednim jest estymator Horvitz-Thompsona:

$$\hat{Y}_d = \sum y_j w_j, \quad (1)$$

gdzie:  $\hat{Y}_d$  – wartość globalna cechy Y dla  $d$ -tej domeny,  $y_j$  – wartość cechy Y dla  $j$ -tej jednostki w badaniu,  $w_j$  – wartość wagi wynikającej ze schematu losowania dla  $j$ -tej jednostki w badaniu (Dehnel, 2003).

### 3.2 ESTYMACJA POŚREDNIA

Estymatory bezpośrednie z racji tego, że są nieobciążone wykorzystuje się do tworzenia estymatorów pośrednich. Estymatory pośrednie – mimo, że są obciążone – cechują się niską wartością wariancji. Idea ich stosowania polega na tzw. „pożyczaniu mocy”<sup>5</sup> spoza badanej domeny. Jednakże, aby móc je zastosować musi zachodzić warunek podobieństwa badanej domeny do większej podpopulacji. Jednym z estymatorów pośrednich jest syntetyczny estymator ilorazowy, który wykorzystuje ocenę parametru estymatora bezpośredniego dla podpopulacji większej niż domena do uzyskania oceny parametru estymatora pośredniego w małym obszarze. Opisany jest wzorem:

$$\hat{Y}_d^{synth} = \frac{X_d}{X_\Omega} \hat{Y}_\Omega^{direct}, \quad (2)$$

gdzie:  $X_d$  – zmienna dodatkowa dla badanej domeny,  $X_\Omega$  – zmienna dodatkowa dla podpopulacji,  $\hat{Y}_\Omega^{direct}$  – estymator bezpośredni dla podpopulacji (Żądło, 2008).

Estymator pośredni opisany wzorem (2) nazywany jest także estymatorem syntetycznym ilorazowym, ponieważ wykorzystuje nieobciążony i efektywny estymator bezpośredni, którego ocena mnożona jest przez iloraz zmiennej dodatkowej dla domeny (wchodzącej w skład populacji) do zmiennej dodatkowej dla populacji. Ocena estymatora pośredniego tworzona jest sztucznie jako część oceny estymatora bezpośredniego.

## 4. BADANIE BUDŻETÓW GOSPODARSTW DOMOWYCH

Badanie Budżetów Gospodarstw Domowych (BBGD) jest badaniem prowadzonym rokrocznie przez Główny Urząd Statystyczny i stanowi podstawowe źródło informacji o warunkach życia ludności. Jednostką badania jest gospodarstwo domowe, które przez miesiąc prowadzi specjalną książeczkę budżetową, notując w niej wszystkie przychody, rozchody oraz ilościowe spożycie dóbr. Informacje zebrane w badaniu wykorzystywane są między innymi do określenia zasięgu i głębokości ubóstwa (GUS, 2011a).

<sup>5</sup> „Pożyczanie mocy” polega na wykorzystywaniu danych z innych źródeł w celu zwiększenia precyzji szacunku.

W niniejszej pracy analizę przeprowadzono na podstawie zbioru zawierającego jednostkowe dane nieidentyfikowalne<sup>6</sup> pochodzące z Badania Budżetów Gospodarstw Domowych z 2005 roku<sup>7</sup>. W województwie wielkopolskim zostało zbadanych 1078 gospodarstw, natomiast wielkość próby w powiatach zawierała się w przedziale od 0 (powiat rawicki) do 137 (powiat miasto Poznań) gospodarstw. Badaną zmienną był dochód ekwiwalentny<sup>8</sup> w gospodarstwie, na podstawie którego została wyznaczona granica ubóstwa. Za zagrożone ubóstwem zostały uznane gospodarstwa, w których ekwiwalentny dochód do dyspozycji nie przekraczał 60% mediany dochodu w kraju. Taki sposób wyznaczania granicy ubóstwa jest rekomendowany i stosowany przez Eurostat (por. Panek, 2011).

## 5. ANALIZA ZAGROŻENIA UBÓSTWEM W POWIATACH WOJEWÓDZTWA WIELKOPOLSKIEGO

### 5.1 ESTYMACJA BEZPOŚREDNIA STOPY ZAGROŻENIA UBÓSTWEM NA PODSTAWIE BBGD

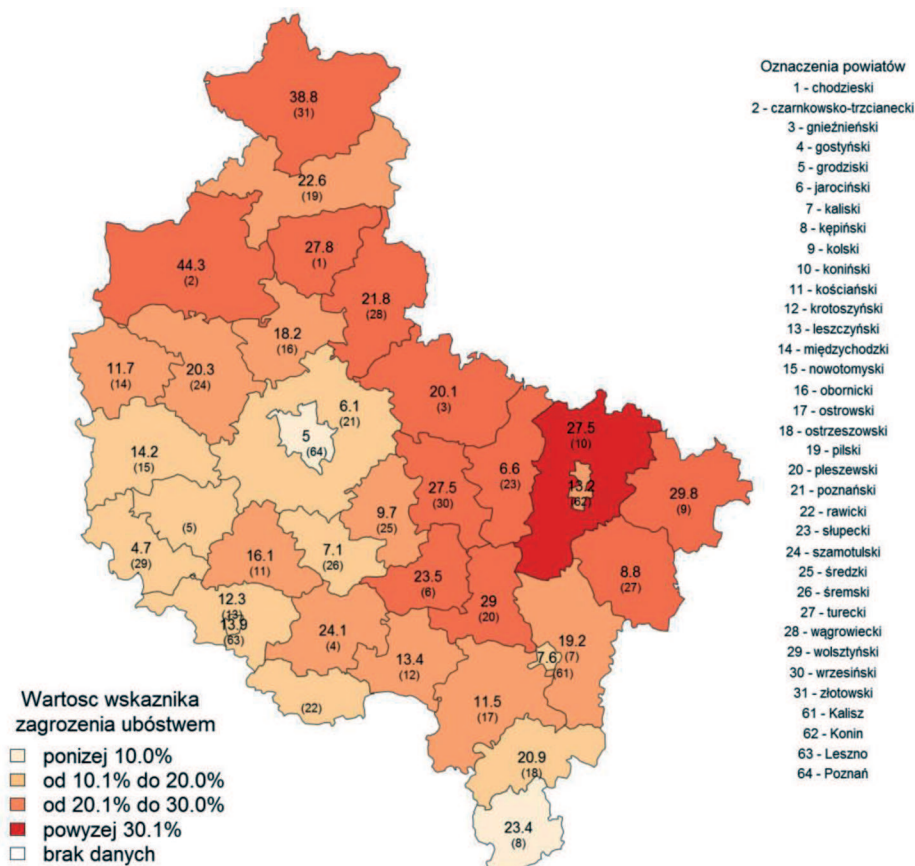
Wykorzystując estymator Horvitz-Thompsona oszacowano stopę zagrożenia ubóstwem w powiatach województwa wielkopolskiego. Jako miarę precyzji szacunków wykorzystano odchylenie standardowe oraz obliczany na jego podstawie współczynnik zmienności (por. Żądło, 2008). Narodowy Instytut Statystyczny Włoch (ISTAT) wskazuje, że współczynnik zmienności nie powinien przekraczać 18% w przypadku małych domen (D'Aló i in., 2012). Na rysunku 1 oraz w tabeli 1 zostały przedstawione wyniki estymacji bezpośredniej w powiatach województwa wielkopolskiego.

Estymacja bezpośrednia nie wskazała oceny estymatora w dwóch powiatach, rawickim oraz grodziskim. Niemożność obliczenia wartości w pierwszym wynikała z faktu, że obszar ten nie znalazł się w próbie, natomiast w powiecie grodziskim w badaniu nie znalazło się ani jedno gospodarstwo, o którym można by powiedzieć, że jest ubogie. W pozostałych powiatach oceny estymatora osiągają wartości od 5% (powiat grodzki Poznań) do ponad 44% (powiat czarnkowsko-trzcianecki) gospodarstw zagrożonych ubóstwem. Współczynnik zmienności zawiera się w przedziale od 18% (powiat kolski) do prawie 98% (powiat wolsztyński) czyniąc te szacunki mało wiarygodnymi. Ograniczenie tej zmienności było możliwe dzięki zastosowaniu estymacji pośredniej.

<sup>6</sup> Dane jednostkowe osób pozbawione cech pozwalających na jednoznaczną identyfikację danej osoby takich jak: imię, nazwisko, pesel itp.

<sup>7</sup> Na potrzeby analizy wejściowy zbiór ograniczono tylko do gospodarstw domowych, w których było co najmniej jedno dziecko poniżej 13. roku życia.

<sup>8</sup> Dochód ekwiwalentny jest to dochód przypadający na jedną osobę w gospodarstwie po uwzględnieniu jego składu demograficznego.



Rysunek 1. Odsetek ubogich gospodarstw domowych – ocena estymatora bezpośredniego

Źródło: opracowanie własne z wykorzystaniem programu R

## 5.2 ESTYMACJA POŚREDNIA STOPY ZAGROŻENIA UBÓSTWEM

Wybór zmiennych pomocniczych do estymacji pośredniej jest zadaniem trudnym oraz znacznie wpływającym na wyniki estymacji. Na podstawie publikacji Ministerstwa Pracy i Polityki Społecznej (MPiPS, 2006) wybrano zmienne związane z niedostatkiem kierując się wartością współczynnika korelacji liniowej Pearsona. Najsilniej z odsetkiem ubogich gospodarstw korelowały się następujące zmienne: odsetek gospodarstw, których głównym źródłem utrzymania są świadczenia ( $r_{xy} = 0,64$ ), stopa bezrobocia długookresowego ( $r_{xy} = 0,53$ ) oraz odsetek osób z wykształceniem podstawowym lub bez wykształcenia w wieku 20-29 lat ( $r_{xy} = 0,52$ ). Trudno jednoznacznie ocenić, czy fakt utrzymywania gospodarstwa ze świadczeń jest determinantą czy skutkiem ubóstwa dlatego do analizy wybrano wskaźnik jakim jest stopa bezrobocia długookresowego pochodzący z Narodowego Spisu Powszechnego 2002. Otrzymana wartość współczynnika korelacji sugeruje, że istnieje zależność pomiędzy liczbą ubogich gospodarstw

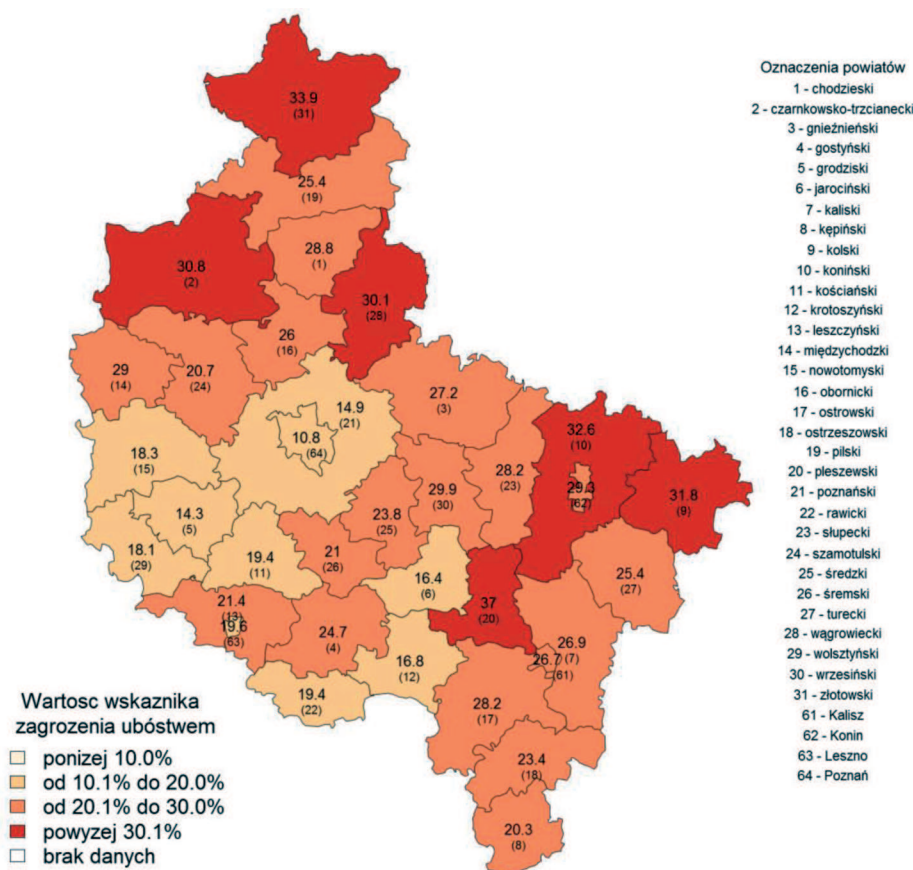
Tabela 1.

Szacunek ubóstwa w województwie wielkopolskim w przekroju powiatów – estymacja bezpośrednia

| Powiat                  | Liczebność | Odsetek ubogich | Odchylenie standardowe | Współczynnik zmienności |
|-------------------------|------------|-----------------|------------------------|-------------------------|
| rawicki                 | 0          | 0,00            | 0,00                   | –                       |
| gnieźnieński            | 5          | 20,12           | 17,98                  | 89,36%                  |
| grodziski               | 7          | 0,00            | 0,00                   | –                       |
| nowotomyski             | 7          | 14,17           | 13,14                  | 92,76%                  |
| leszczyński             | 8          | 12,29           | 11,53                  | 93,82%                  |
| szamotulski             | 10         | 20,28           | 12,79                  | 63,09%                  |
| czarnkowsko-trzcianecki | 11         | 44,35           | 14,96                  | 33,74%                  |
| obornicki               | 11         | 18,19           | 11,64                  | 64,02%                  |
| kościański              | 12         | 16,08           | 10,46                  | 65,04%                  |
| turecki                 | 13         | 8,76            | 8,32                   | 95,03%                  |
| chodzieski              | 14         | 27,78           | 11,88                  | 42,76%                  |
| śremski                 | 14         | 7,07            | 6,82                   | 96,49%                  |
| wągrowiecki             | 14         | 21,85           | 11,24                  | 51,45%                  |
| śłupecki                | 15         | 6,63            | 6,41                   | 96,69%                  |
| średzki                 | 20         | 9,67            | 6,52                   | 67,39%                  |
| wrzesiński              | 22         | 27,55           | 9,26                   | 33,62%                  |
| wolsztyński             | 23         | 4,69            | 4,58                   | 97,57%                  |
| kępiński                | 24         | 23,39           | 8,58                   | 36,69%                  |
| ostrzeszowski           | 24         | 20,89           | 8,33                   | 39,85%                  |
| międzychodzki           | 25         | 11,69           | 6,36                   | 54,39%                  |
| kaliski                 | 26         | 19,24           | 7,91                   | 41,10%                  |
| krotoszyński            | 27         | 13,45           | 6,47                   | 48,11%                  |
| m.Kalisz                | 27         | 7,62            | 5,18                   | 67,99%                  |
| m.Konin                 | 30         | 13,18           | 6,15                   | 46,67%                  |
| m.Leszno                | 36         | 13,91           | 5,78                   | 41,54%                  |
| gostyński               | 37         | 24,07           | 7,14                   | 29,66%                  |
| koniński                | 41         | 27,49           | 7,15                   | 26,01%                  |
| złotowski               | 42         | 38,84           | 7,70                   | 19,81%                  |
| pleszewski              | 44         | 28,96           | 6,81                   | 23,50%                  |
| pilski                  | 46         | 22,59           | 6,34                   | 28,06%                  |
| jarociński              | 60         | 23,49           | 5,37                   | 22,87%                  |
| kolski                  | 74         | 29,83           | 5,39                   | 18,06%                  |
| ostrowski               | 80         | 11,50           | 3,66                   | 31,81%                  |
| poznański               | 92         | 6,10            | 2,47                   | 40,57%                  |
| m.Poznań                | 137        | 5,00            | 1,84                   | 36,89%                  |

Źródło: obliczenia własne z wykorzystaniem programu SAS

a liczbą bezrobotnych. Należy także pamiętać, że syntetyczny estymator ilorazowy wykorzystuje się przy założeniu, że relacja zachodząca pomiędzy odsetkiem ubogich gospodarstw, a stopą bezrobocia w poszczególnych powiatach województwa wielkopolskiego jest podobna do relacji pomiędzy odsetkiem ubogich gospodarstw a stopą bezrobocia w całym województwie. W związku z tym współczynnik zmienności w poszczególnych powiatach będzie równy temu wskaźnikowi dla całego województwa. Oceny estymatora pośredniego dla powiatów województwa wielkopolskiego przedstawione są na rysunku 2 oraz w tabeli 2.



Rysunek 2. Odsetek ubogich gospodarstw domowych – ocena estymatora pośredniego

Źródło: opracowanie własne z wykorzystaniem programu R

Zastosowanie estymatora pośredniego pozwoliło na oszacowanie stopy gospodarstw zagrożonych ubóstwem nawet w tych powiatach, które nie były reprezentowane w próbie. W powiecie rawickim obserwuje się ponad 19% ubogich gospodarstw, a w powiecie grodziskim 14% gospodarstw cierpi niedostatek. Najmniej gospodarstw zagrożonych ubóstwem znajduje się w Poznaniu, natomiast najwięcej w powiatach: ple-

Tabela 2.

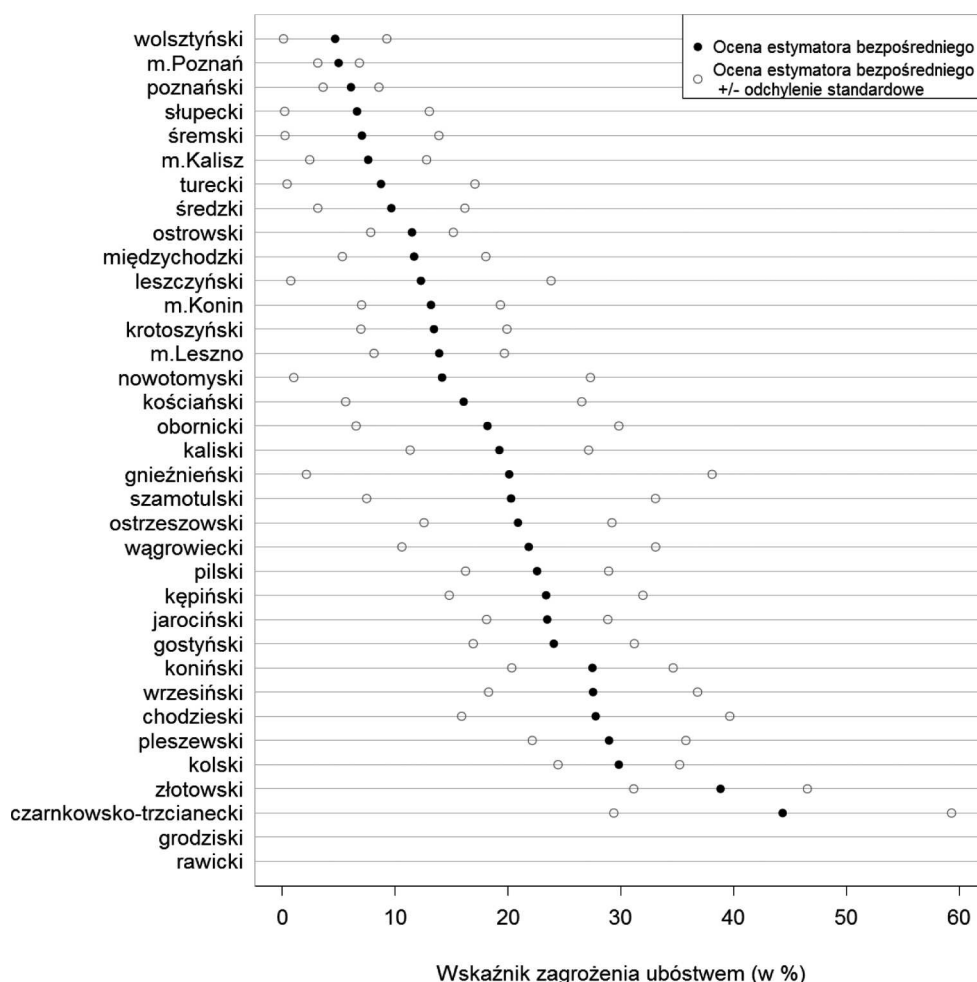
Szacunek ubóstwa w województwie wielkopolskim w przekroju powiatów – estymacja pośrednia

| Powiat                  | Liczebność | Odsetek ubogich | Odchylenie standardowe | Współczynnik zmienności |
|-------------------------|------------|-----------------|------------------------|-------------------------|
| rawicki                 | 0          | 19,43           | 1,32                   | 14,71%                  |
| gnieźnieński            | 5          | 27,17           | 1,85                   | 14,71%                  |
| grodziski               | 7          | 14,31           | 0,97                   | 14,71%                  |
| nowotomyski             | 7          | 18,31           | 1,25                   | 14,71%                  |
| leszczyński             | 8          | 21,44           | 1,46                   | 14,71%                  |
| szamotulski             | 10         | 20,69           | 1,41                   | 14,71%                  |
| czarnkowsko-trzcianecki | 11         | 30,84           | 2,10                   | 14,71%                  |
| obornicki               | 11         | 25,96           | 1,77                   | 14,71%                  |
| kościański              | 12         | 19,42           | 1,32                   | 14,71%                  |
| turecki                 | 13         | 25,37           | 1,73                   | 14,71%                  |
| chodzieski              | 14         | 28,78           | 1,96                   | 14,71%                  |
| śremski                 | 14         | 20,98           | 1,43                   | 14,71%                  |
| wągrowiecki             | 14         | 30,05           | 2,04                   | 14,71%                  |
| śłupecki                | 15         | 28,19           | 1,92                   | 14,71%                  |
| średzki                 | 20         | 23,80           | 1,62                   | 14,71%                  |
| wrzesiński              | 22         | 29,90           | 2,03                   | 14,71%                  |
| wolsztyński             | 23         | 18,14           | 1,23                   | 14,71%                  |
| kępiński                | 24         | 20,34           | 1,38                   | 14,71%                  |
| ostrzeszowski           | 24         | 23,35           | 1,59                   | 14,71%                  |
| międzychodzki           | 25         | 28,98           | 1,97                   | 14,71%                  |
| kaliski                 | 26         | 26,86           | 1,83                   | 14,71%                  |
| krotoszyński            | 27         | 16,79           | 1,14                   | 14,71%                  |
| m.Kalisz                | 27         | 26,68           | 1,81                   | 14,71%                  |
| m.Konin                 | 30         | 29,27           | 1,99                   | 14,71%                  |
| m.Leszno                | 36         | 19,59           | 1,33                   | 14,71%                  |
| gostyński               | 37         | 24,71           | 1,68                   | 14,71%                  |
| koniński                | 41         | 32,60           | 2,22                   | 14,71%                  |
| złotowski               | 42         | 33,91           | 2,31                   | 14,71%                  |
| pleszewski              | 44         | 36,99           | 2,52                   | 14,71%                  |
| pilski                  | 46         | 25,39           | 1,73                   | 14,71%                  |
| jarociński              | 60         | 16,43           | 1,12                   | 14,71%                  |
| kolski                  | 74         | 31,77           | 2,16                   | 14,71%                  |
| ostrowski               | 80         | 28,21           | 1,92                   | 14,71%                  |
| poznański               | 92         | 14,87           | 1,01                   | 14,71%                  |
| m.Poznań                | 137        | 10,79           | 0,73                   | 14,71%                  |

Źródło: obliczenia własne z wykorzystaniem programu SAS

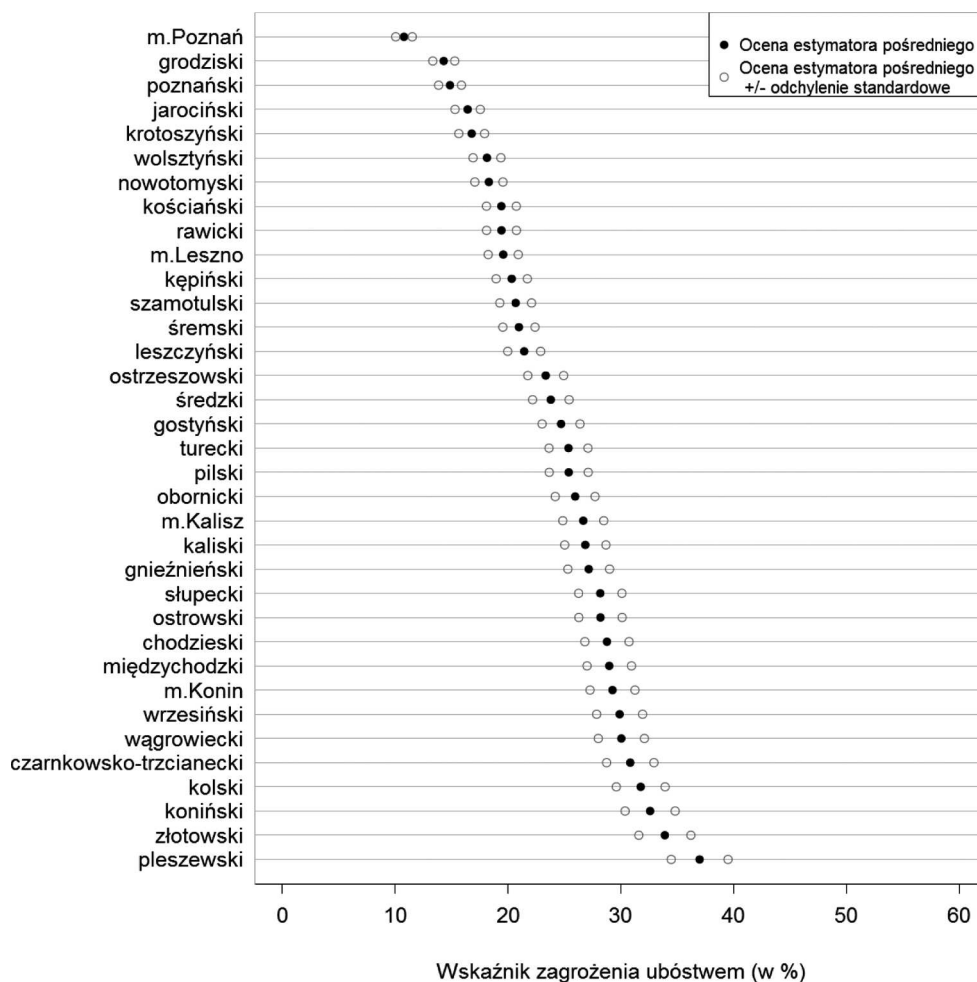
szewskim, złotowskim i konińskim. Takie oceny znajdują swoje uzasadnienie w wartościach stopy bezrobocia długookresowego w tych powiatach. W porównaniu do estymatora bezpośredniego, estymator pośredni dał wyższe oceny w przypadku 31 powiatów, a niższe tylko w przypadku czterech. Niższą ocenę otrzymano między innymi w powiecie czarnkowsko-trzcianeckim, w którym estymacja bezpośrednia wskazała zagrożenie ubóstwem na bardzo wysokim poziomie 44%, a estymacja pośrednia zredukowała tę stopę do poziomu około 31%.

Estymacja pośrednia wpłynęła także na ograniczenie zmienności badanej cechy. Na rysunkach 3 i 4 przedstawiony jest rozkład ocen estymatorów z przedziałem ufności wyznaczonym jako ocena uwzględniająca błąd standardowy.



Rysunek 3. Rozkład ocen estymatora bezpośredniego dla powiatów z uwzględnieniem błędów standardowych odpowiednich szacunków

Źródło: opracowanie własne z wykorzystaniem programu R



Rysunek 4. Rozkład ocen estymatora pośredniego dla powiatów z uwzględnieniem błędów standardowych odpowiednich szacunków

Źródło: opracowanie własne z wykorzystaniem programu R

Estymator bezpośredni – mimo, że jest nieobciążony – cechuje się bardzo dużą wariancją, co pokazuje rysunek 3. W przypadku większości powiatów punkty obrazujące odchylenie standardowe są bardzo odsunięte od punktów reprezentujących oceny. Na szczególną uwagę zasługuje zmienność powiatu gnieźnieńskiego, gdzie przedział ufności wyznaczony przez odchylenie standardowe osiąga wartości od 2% do 38% odsetka ubogich gospodarstw. Mimo tego, że dla tego obszaru przedział jest najszerszy, największą zmiennością wśród analizowanych powiatów cechuje się powiat wolsztyński. Wskaźnik zagrożenia ubóstwem oszacowano na poziomie 4,7% i ocena ta znajduje się w przedziale od 0,1% do 9,3%. Zastosowanie estymatora syntetycznego ilorazowego pozwoliło na znaczne zmniejszenie wariancji, ale kosztem obciążenia rezultatów. War-



tości odchylenia standardowego dla poszczególnych powiatów w wyniku zastosowania estymacji pośredniej znacząco się zmniejszyły, co można zaobserwować na rysunku 4.

## 6. WNIOSKI I DALSZE KIERUNKI PRAC

Uzyskane wyniki wskazują, że estymacja pośrednia pozwala na znaczącą redukcję błędu oszacowania parametrów estymatora dla powiatów województwa wielkopolskiego w porównaniu do ocen uzyskanych przy wykorzystaniu estymatora bezpośredniego. Wysoka korelacja wskaźnika zagrożenia ubóstwem z długookresową stopą bezrobocia pozwoliła na użycie tej zmiennej jako zmiennej pomocniczej. Można wnioskować, że wyższa stopa bezrobocia długookresowego determinuje wyższy poziom wskaźnika zagrożenia ubóstwem. Ubóstwo jest zjawiskiem na tyle złożonym, że analiza tego zjawiska na podstawie tylko jednej zmiennej pomocniczej pozostawia pewien niedosyt. Stąd też w dalszej kolejności prowadzone będą prace w kierunku rozszerzenia listy zmiennych objaśniających. Wyzwaniem jest znalezienie takich wskaźników, które będą jednoznacznie determinować ubóstwo. Ponadto biorąc pod uwagę kierunki, w jakich rozwija się statystyka małych obszarów, w badaniu podjęte zostaną próby zastosowania liniowych modeli mieszanych do szacowania liczby gospodarstw zagrożonych ubóstwem.

*Urząd Statystyczny w Poznaniu*

## LITERATURA

- [1] D'Aló M., Di Consiglio L., Falorsi S., Solari F., (2012), *Small Area Estimation*, Course on Small Area Estimation, ESSnet Project on SAE.
- [2] Dehnel G., (2003), *Statystyka Małych Obszarów jako narzędzie oceny rozwoju ekonomicznego regionów*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.
- [3] Drewnowski J., (1977), *Poverty: Its Meaning and Measurement*, Development and Change, no. 8, s. 183-208.
- [4] GUS, (2010), *Ubóstwo w Polsce na tle krajów Unii Europejskiej w świetle Europejskiego Badania Dochodów i Warunków Życia – EU-SILC 2008*, Warszawa.
- [5] GUS, (2011a), *Budżety gospodarstw domowych w 2010 r.*, Warszawa.
- [6] GUS, (2011b), *Ubóstwo w Polsce w 2010 r. (na podstawie badania budżetów gospodarstw domowych)*, Warszawa.
- [7] Hungry Notes, (2012), 2012 World Hunger and Poverty Facts and Statistics, <http://www.worldhunger.org/articles/Learn/world%20hunger%20facts%202002.htm> [dostęp: 9.04.2012].
- [8] Kaplan, M., Makoka, D., (2005), *Poverty and Vulnerability*, Uniwersytet w Bonn, [http://www.zef.de/fileadmin/downloads/forum/docprog/Termpapers/2005\\_2b\\_Kaplan\\_Makoka.pdf](http://www.zef.de/fileadmin/downloads/forum/docprog/Termpapers/2005_2b_Kaplan_Makoka.pdf) [dostęp: 22.06.2012].
- [9] MPiPS, (2006), *Wykluczenie i integracja społeczna w Polsce. Ujęcie wskaźnikowe*, Warszawa.
- [10] Panek T., Szulc A. (red.), (2004), *Statystyka społeczna: wybrane zagadnienia*, Oficyna Wydawnicza Szkoła Główna Handlowa, Warszawa.
- [11] Panek T., (2009), *Ubóstwo i nierówności: dylematy pomiaru*, Szkoła Główna Handlowa, Warszawa, [http://www.stat.gov.pl/cps/rde/xbcr/gus/POZ\\_Ubostwo\\_i\\_nierownosci-dylematy\\_pomiaru.pdf](http://www.stat.gov.pl/cps/rde/xbcr/gus/POZ_Ubostwo_i_nierownosci-dylematy_pomiaru.pdf) [dostęp: 08.03.2012].

- [12] Panek T., (2011), *Ubóstwo, wykluczenie społeczne i nierówność. Teoria i praktyka pomiaru*, Oficyna Wydawnicza Szkoła Główna Handlowa, Warszawa.
- [13] Sen A., (1997), *On Economic Inequality*, Oxford University Press, Oxford.
- [14] UN, (2000), *United Nations Millennium Declaration*, <http://www.un.org/millennium/declaration/ares552e.pdf> [dostęp: 5.09.2012].
- [15] Żądło T., (2008), *Elementy statystyki małych obszarów z programem R*, Wydawnictwo Akademii Ekonomicznej im. Karola Adamieckiego, Katowice.

#### ANALIZA UBÓSTWA W PRZEKROJU POWIATÓW W WOJEWÓDZTWIE WIELKOPOLSKIM Z WYKORZYSTANIEM METOD STATYSTYKI MAŁYCH OBSZARÓW

##### S t r e s z c z e n i e

Ubóstwo jest zjawiskiem bardzo złożonym. Ograniczanie tego niekorzystnego zjawiska jest obecnie najważniejszym celem Banku Światowego. Jednak, aby tę misję realizować potrzebne są metody identyfikacji ubogich. Wskaźniki opisujące ubóstwo są jednak dostarczane na bardzo ogólnym poziomie. Otrzymanie bardziej szczegółowych informacji jest możliwe dzięki zastosowaniu statystyki małych obszarów (SMO). Jest to zbiór metod pozwalających na estymację parametrów przy małej liczebności próby z wykorzystaniem wszelkich dostępnych źródeł informacji.

Głównym celem artykułu jest próba oszacowania wskaźnika zagrożenia ubóstwem na poziomie powiatów w województwie wielkopolskim przy użyciu wybranych metod SMO. Obliczenie tej miary odbędzie się na podstawie danych udostępnianych przez GUS charakteryzujących poziom życia gospodarstw domowych. Taka estymacja pozwoli uzyskać kompleksowej informacji na poziomie lokalnym dotyczącej sfery ubóstwa.

**Słowa kluczowe:** estymacja, mała próba, statystyka małych obszarów, ubóstwo

#### POVERTY ANALYSIS IN DISTRICTS OF WIELKOPOLSKA WITH SMALL AREA ESTIMATION METHODS

##### A b s t r a c t

Poverty is very complex phenomenon. Limitation of this disadvantageous phenomenon is main goal of World Bank now. Nevertheless, to realize this mission methods of identification the poor are necessary. Indicates describing poverty are provide at very general level. Getting more detailed information is possible thanks to applied small area estimation (SAE). It is the set of methods which allow estimation of parameters at small sample size with usage of all available information sources.

The main goal of this article is attempt to estimate at-risk-of-poverty rate at districts in Great Poland using chosen methods of SAE. These measure will calculation based on statistics data describing life level of households in Poland provided by Central Statistical Office. That estimation give comprehensive information at local level about poverty.

**Key words:** estimation, small sample, small area estimation, poverty

ROMA RYŚ-JUREK

## ZASTOSOWANIE HIERARCHICZNEJ KLASYFIKACJI AGLOMERACYJNEJ DO GRUPOWANIA KRAJÓW UNII EUROPEJSKIEJ ZE WZGLĘDU NA STRUKTURĘ I SKALĘ PRODUKCJI GOSPODARSTW ROLNYCH

### 1. WSTĘP

Obszary wiejskie w Unii Europejskiej (UE-27) są zróżnicowane pod względem wewnętrznych struktur i rozwoju. Wpływ na to mają czynniki geograficzne, demograficzne, a także poziom rozwoju infrastrukturalnego (Chmielewska, Mierosławska, 2007). Ponadto przestrzenne zróżnicowanie warunków przyrodniczych, gospodarczych i społecznych oddziałuje w znacznym stopniu na kierunki, skalę i intensywność produkcji rolniczej gospodarstw (Stuczyński, Jadczyżyn, Kukuła, 2006). Warto również zaznaczyć, że znaczenie mają także czynniki historyczne i polityczne (Fierla, 2011). Określanie podobieństw i różnic pomiędzy obszarami wiejskimi poszczególnych krajów UE-27 staje się koniecznością. Stosowane w praktyce delimitacje regionów, czyli wyodrębnianie (wydzielanie) ma charakter administracyjny (np. podziałna województwa), planistyczny (plany regionalne rozwoju społeczno-gospodarczego) lub autorski, który to oparty jest na wybranych przesłankach i wskaźnikach analitycznych (Harasim, 2006).

Odpowiednim podmiotem badań wydają się być opublikowane wektory średnich ważonych przypadające na jedno gospodarstwo rolne<sup>1</sup>, a reprezentujące cechy szczególne rolnictwa danego kraju. Celem opracowania jest próba delimitacji krajów w UE-27 o podobnej strukturze i skali produkcji pochodzącej z gospodarstw rolnych. W badaniu położono nacisk na uwypuklenie podobieństwa struktur produkcji typowych dla obszarów klimatycznych UE-27 w 2008 roku oraz na wskazanie i opisanie klasy, do której zaliczono Polskę.

### 2. MATERIAŁ ŹRÓDŁOWY I METODY BADAWCZE

Do przeprowadzenia badań wykorzystano dane z sieci danych rachunkowości rolnej gospodarstw rolniczych FADN (ang. *Farm Accountancy Data Network*). Zawiera

---

<sup>1</sup> W bazie FADN zawarto wartości średnie obliczone dla grup gospodarstw rolnych z pola obserwacji FADN, które wyliczane są według typów rolniczych i wielkości ekonomicznej. Wylicza się je z co najmniej 15 gospodarstw rolnych. Ponadto by precyzyjnie odzwierciedlić zróżnicowanie próby badawczej, dokonuje się podziału pola obserwacji na warstwy według: położenia regionalnego, wielkości ekonomicznej i typu rolniczego (IERiGŻ, 2010).

ona wektory średnich ważonych przeliczone na jedno gospodarstwo rolne z 27 krajów członkowskich UE w 2008 roku. Baza FADN aktualizowana jest ze znacznym opóźnieniem, więc na początku 2012 roku najbardziej aktualnymi danymi były wielkości dla 2008 roku, a dane dla roku 2009 były niepełne (FADN, 2012).

Przedmiotem badań była delimitacja krajów UE-27 o podobnej strukturze i skali produkcji pochodzącej z gospodarstw rolnych. W badaniu wykorzystano 27 cech, z których ostatecznie dokonano merytorycznego wyboru analizowanych zmiennych kierując się przeprowadzonymi badaniami eksperckimi<sup>2</sup>. Starano się zaznaczyć, że obecnie wyniki gospodarstw rolnych, oprócz uwarunkowań przyrodniczych, zależą również od uwarunkowań pozaprzyrodniczych, np.: od poziomu rozwoju społecznego, gospodarczego, przeszłości historycznej, polityki rolnej, nowoczesnych technologii, itp. (Bański, 2007). Dlatego w badaniu posłużono się strukturą produkcji oraz cechami reprezentującymi możliwości gospodarstwa do rozwijania produkcji (wielkość ekonomiczna i powierzchnia UR) i wyników z niej uzyskanych (wartość produkcji i dochodu z gospodarstwa rolnego). Do badania trafiły cechy dotyczące:

– struktury produkcji (%), gdyż odwzorowuje ona cechy charakterystyczne rolnictwa w zależności od położenia geograficznego danego kraju. Reprezentowały ją 23 cechy, a mianowicie były to udziały poszczególnych rodzajów produkcji w produkcji ogółem gospodarstwa rolnego uzyskanej w ramach działalności operacyjnej gospodarstwa rolnego. Pogrupowano je w kategoriach takich jak: zboża, rośliny białkowe, uprawy energetyczne, ziemniaki, buraki cukrowe, rośliny oleiste, rośliny przemysłowe, warzywa i kwiaty, owoce, owoce cytrusowe, wina i winogrona, oliwki i olej z oliwek, uprawy pastewne, inna produkcja roślinna<sup>3</sup>, mleko i przetwory z mleka krowiego, ży-

<sup>2</sup> Dokonując wyboru merytorycznego cech do badania oparto się na opiniach eksperckich, a nie na weryfikacji zmiennościowo-korelacyjnej, gdyż usunięcie kilku cech o niskim zróżnicowaniu ze struktury produkcji spowodowałoby niepełne odzwierciedlenie prowadzonej produkcji rolniczej na terenie UE przez gospodarstwa rolne. Ponadto spośród analizowanych cech opisujących strukturę produkcji, najwyższym zróżnicowaniem (powyżej 150%) w UE-27 charakteryzowało się 6 cech, którymi były udziały w strukturze produkcji dotyczące: upraw energetycznych, roślin przemysłowych, owoców cytrusowych, żywca baraniego i koziego, jaj, mleka i przetworów z mleka owczego i koziego. Najmniejsze zróżnicowanie (od 53% do 75%) w UE-27 wystąpiło w przypadku siedmiu cech, którymi były udziały produkcji zbóż, ziemniaków, buraków cukrowych, owoców, innej produkcji roślinnej, mleka i przetworów z mleka krowiego oraz żywca wieprzowego w strukturze produkcji ogółem. Natomiast pozostałe 4 cechy zastosowane w badaniu, czyli wielkość ekonomiczna, powierzchnia UR, wartość produkcji ogółem i dochodu z gospodarstwa rolnego wykazywały bardzo wysokie zróżnicowanie w UE-27 (powyżej 100%).

<sup>3</sup> Obejmuje inne produkty pochodzenia roślinnego, do których zaliczono m. in.: nasiona i sadzonki traw, roślin polowych i ogrodniczych (bez zbóż, roślin białkowych, ziemniaków i roślin oleistych), inne uprawy polowe, uprawy trwałe rosnące pod osłonami nieuwzględnione wcześniej, a także przyrost wartości młodych plantacji wieloletnich (do wejścia w plonowanie) (IERiGŻ, 2010).

wiec wołowy, żywiec wieprzowy, żywiec barani i kozi, żywiec drobiowy, jaja, mleko i przetwory z mleka owczego i koziego, inna produkcja zwierzęca<sup>4</sup> oraz inna produkcja<sup>5</sup>;

- wielkości ekonomicznej wyrażonej w ESU<sup>6</sup>;
- powierzchni użytków rolnych (UR w ha);
- wartości produkcji ogółem (euro);
- wartości dochodu z gospodarstwa rolnego (euro).

W celu zniwelowania wpływu niejednakowej reprezentacji cech w badaniu, nadano im wagi wyznaczone po badaniach eksperckich. W ich wyniku uznano, że skoro podstawą badania jest struktura i skala produkcji gospodarstw rolnych, to powinno się przyjąć założenie, że oba te kryteria są jednakowo ważne. Wobec tego nadano im tę samą wagę równą 0,5. W ten sposób strukturze produkcji, reprezentowanej przez 23 cechy przypadła do podziału waga 0,5, a dla pozostałych 4 cech po 0,125. Po wykonaniu ważenia cech, przeprowadzono analizę skupień z wykorzystaniem metody Warda i odległości euklidesowej.

Badanie wykonano za pomocą metody hierarchicznej klasyfikacji aglomeracyjnej należącej do analizy skupień. Hierarchiczna klasyfikacja aglomeracyjna opiera się na centralnej procedurze aglomeracyjnej, która polega na znalezieniu pary klas najmniej odległych od siebie (najbardziej podobnych) w macierzy odległości. Kolejnym etapem jest redukcja liczby klas o jeden – łączy się wcześniej uzyskane klasy w nową klasę. Następnie przekształca się odległości pomiędzy połączonymi klasami oraz pozostałymi klasami. Etapy te powtarza się tak długo, aż wszystkie obiekty nie znajdą się w jednej klasie. Różnice w procedurach metod aglomeracyjnych wynikają z odmienności definiowania odległości międzyklasowej (Walesiak, 2004, s. 324).

W przeprowadzanym badaniu dla UE-27 zastosowano metodę Warda i odległość euklidesową<sup>7</sup>. Metoda Warda, zmierza do minimalizacji sumy kwadratów odchyleń wewnątrz skupień. W tej metodzie na każdym etapie spośród wszystkich możliwych do łączenia par skupień wybiera się tę, która w rezultacie łączenia daje skupienie o minimalnym zróżnicowaniu (por. Everitt i in., 2011, s. 77-78; Stanisław, 2007, s. 122)<sup>8</sup>.

<sup>4</sup> Obejmuje inne zwierzęta i produkty zwierzęce, do których zalicza się m. in.: żywiec koński i inne zwierzęta niezaliczone gdzie indziej (np.: króliki, roje pszczele, zwierzęta futerkowe, ryby, strusie – z różnicą wartości zwierząt stada obrotowego i podstawowego), a także wełnę i pozostałe produkty pochodzenia zwierzęcego (np. obornik), przychody z krycia i z usługowego odchowu zwierząt oraz różnicę wartości stada podstawowego koni z tytułu zmiany cen (IERiGŻ, 2010).

<sup>5</sup> Należą do niej m. in.: czynsz za wydzierżawioną ziemię w stanie gotowym do siewu, przychody z okazjonalnego przekazania powierzchni paszowej, produkty z lasu, świadczenie usług, wynajem sprzętu, odsetki od aktywów obrotowych niezbędnych do bieżącego funkcjonowania gospodarstwa rolnego, przychody z agroturystyki, przychody dotyczące wcześniejszych lat obrachunkowych, itp. (IERiGŻ, 2010).

<sup>6</sup> Europejska Jednostka Wielkości (ESU) jest parametrem służącym do określania wielkości ekonomicznej gospodarstwa rolnego ustalonej na podstawie standardowych nadwyżek bezpośrednich gospodarstwa. Jedno ESU odpowiada równowartości 1200 euro (IERiGŻ, 2010).

<sup>7</sup> Przed wykonaniem hierarchicznej klasyfikacji aglomeracyjnej dane wejściowe wystandaryzowano.

<sup>8</sup> Jeśli dla metody Warda, centroidalnej i medianowej odległości między obiektami zostały wyznaczone za pomocą kwadratu odległości euklidesowej, to mają one wtedy interpretację geometryczną. Metody te mogą być stosowane, gdy macierz odległości jest liczona na podstawie innych miar odległości. Ale należy

Według Stanisza (2007, s. 122) metoda Warda „jest traktowana jako bardzo efektywna, chociaż zmierza do tworzenia skupień o małej wielkości”. Metoda Warda prowadzi do otrzymania skupień o zbliżonych liczebnościach, charakteryzujących się minimalną wariancją, dlatego też często jest wykorzystywana do klasyfikacji jednostek przestrzennych (Filipiak, 2006, s. 57). Należy podkreślić, że metoda ta prowadzi do tworzenia klas sferycznych z małą liczbą obiektów, w przybliżeniu o takim samym kształcie (Wysocki, 2010, s. 92). Jest wrażliwa na wartości odstające cech (Everitt i in., 2011, s. 79).

Wyniki klasyfikacji aglomeracyjnej przedstawia się graficznie w formie drzewka połączeń (dendrogramu). Umożliwia to dokładne określenie wzajemnego usytuowania poszczególnych klas oraz obiektów w nich zawartych (Walesiak, 2004). Właściwą liczbę klas krajów o podobnej strukturze i skali produkcji wyznaczono na podstawie analizy wielkości przyrostów wariancji wewnątrzklasowej w procesie łączenia klas między poszczególnymi poziomami łączy<sup>9</sup>. Relatywnie duże zmiany wartości tych przyrostów, rozumiane jako różnice kolejnych odległości aglomeracyjnych, sugerują najlepszą liczbę klas (Wysocki, 2010).

W celu zinterpretowania wyników klasyfikacji i aby wyodrębnić charakterystyczne typy struktury i skali produkcji w UE-27, zastosowano identyfikację na podstawie wyników klasyfikacji rozłącznej i cech metrycznych polegającą na porównaniu średnich wewnątrz klas ze średnimi ogólnymi otrzymanymi z całej zbiorowości obiektów (Wysocki, 2010). Wykonano normalizację wartości cech w  $c$ -tej klasie za pomocą miernika różnic średnich:

$$z_{ck(d)} = \frac{\bar{x}_{ck} - \bar{x}_k}{s_{k(w)}} (c = 1, \dots, C; k = 1, \dots, K),$$

gdzie:  $C$  to liczba klas,  $K$  zaś oznacza liczbę cech, z kolei  $\bar{x}_{ck}$  jest średnią  $k$ -tej cechy w  $c$ -tej klasie,  $\bar{x}_k$  jest średnią ogólną  $k$ -tej cechy w zbiorowości składającej się z  $N$  obiektów;  $s_{k(w)}$  jest przeciętnym zróżnicowaniem wewnątrzklasowym wartości  $k$ -tej cechy, wyznaczoną ze wzoru:

$$s_{k(w)} = \left[ \frac{1}{N - C} \sum_{c=1}^C (N_c - 1) \cdot s_{ck}^2 \right]^{1/2},$$

gdzie:  $s_{ck}^2$  jest wariancją wewnątrzklasową w  $c$ -tej klasie ( $c = 1, \dots, C$ ) obliczoną względem  $k$ -tej cechy:

$$s_{ck}^2 = \frac{N_c}{N_c - 1} \sum_{i \in I_c} v_{ic} (x_{ik} - \bar{x}_{ck})^2 = \frac{N_c}{N_c - 1} \left\{ \sum_{i \in I_c} v_{ic} x_{ik}^2 - \bar{x}_{ck}^2 \right\}$$

pamiętać, że wtedy odległość międzyklasowa w centralnej procedurze aglomeracyjnej nie ma interpretacji geometrycznej (por.: Anderberg, 1973, s. 141; Walesiak, Dudek, 2009, s. 42).

<sup>9</sup> Interesujący przegląd zastosowania bezwzorcowej klasyfikacji obiektów w ekonomice rolnictwa przedstawia publikacja Kisielińskiej (2009).

przy czym:  $\bar{x}_{ck} = \sum_{i \in I_c} v_{ic} x_{ik}$  jest średnią  $k$ -tej cechy w  $c$ -tej klasie;  $v_{ic}$  jest wagą  $i$ -tego obiektu w  $c$ -tej klasie taką, że  $\sum_{i \in I_c} v_{ic} = 1$ , która może być definiowana jako jednakowa  $v_{ic} = 1/N_c$ ;  $N_c$  jest liczbą obiektów w  $c$ -tej klasie;  $I_c$  oznacza zbiór  $N_c$  numerów obiektów należących do  $c$ -tej klasy (Wysocki, 2010; Wysocki, 1996).

Wartości  $z_{ck(d)}$  były podstawą do wyróżnienia cech charakterystycznych w klasach z wykorzystaniem skali wartości (Wysocki, 2010):

1.  $z_{ck(d)} \in <3; +\infty)$  – występuje bardzo duże natężenie  $k$ -tej cechy w  $c$ -tej klasie, cecha jest wysoce pozytywnie charakterystyczna,
2.  $z_{ck(d)} \in <2; 3)$  – występuje duże natężenie  $k$ -tej cechy w  $c$ -tej klasie, cecha jest średnio pozytywnie charakterystyczna,
3.  $z_{ck(d)} \in (-2; 2)$  – występuje przeciętne natężenie  $k$ -tej cechy w  $c$ -tej klasie, cecha nie wyróżnia się i nie jest charakterystyczna,
4.  $z_{ck(d)} \in (-3; -2)$  – występuje małe natężenie  $k$ -tej cechy w  $c$ -tej klasie, cecha jest średnio negatywnie charakterystyczna,
5.  $z_{ck(d)} \in (-\infty; -3)$  – występuje bardzo małe natężenie  $k$ -tej cechy w  $c$ -tej klasie, cecha jest wysoce negatywnie charakterystyczna.

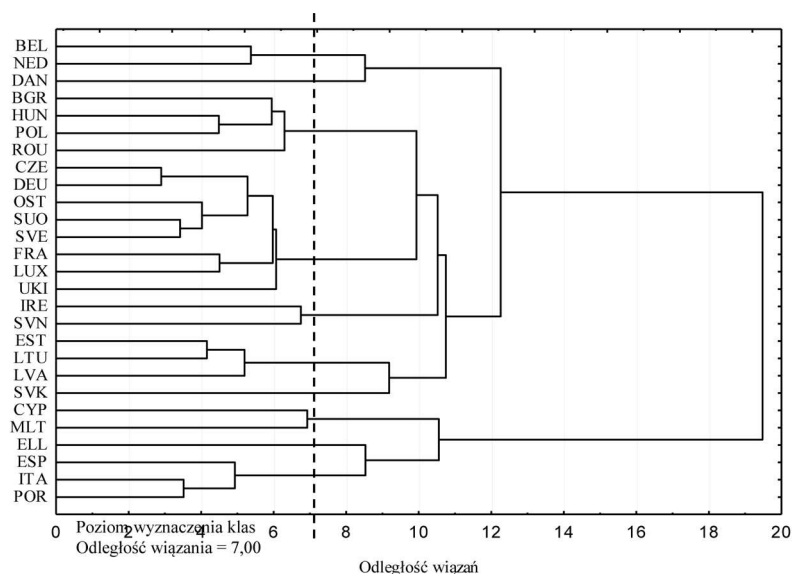
### 3. WYNIKI BADAŃ

Zanim wykonano badanie, poddano porównaniu kształtowanie się wybranych 27 cech w Polsce ze średnią obliczoną dla UE-27 w 2008 roku. Jak już zaznaczono, 23 cechy reprezentowały strukturę produkcji, która składała się z produkcji roślinnej (14 cech), zwierzęcej (8 cech) i innej produkcji (1 cecha). Warto zauważyć, że w strukturze produkcji roślinnej polskiego gospodarstwa rolnego w 2008 roku przeciętnie przeważały zboża (20,41%) oraz warzywa i kwiaty (12,84%), przy średnim poziomie w UE-27 odpowiednio: 15,50% i 11,22%. W Polsce ze względów klimatycznych nie występowała uprawa cytrusów, winorośli i oliwek (tab. 1). Z kolei w strukturze produkcji zwierzęcej gospodarstwa rolnego w Polsce w 2008 roku największy udział miały: mleko i przetwory z mleka krowiego (15,56%), żywiec wieprzowy (15,15%) i żywiec drobiowy (7,40%). Natomiast w strukturze produkcji zwierzęcej gospodarstwa rolnego z krajów UE-27 w 2008 roku przeciętnie dominowało mleko i przetwory z mleka krowiego (16,68%), żywiec wieprzowy (8,89%) oraz żywiec wołowy (7,12%). Warto dodać, że w Polsce prawie nie prowadzono produkcji mleka owczego i koziego oraz jego przetworów, a także żywca baraniego i koziego. Warto też podkreślić, że poziomy większości cech obserwowane w Polsce w 2008 roku okazały się znacznie niższe od średniej w UE-27.

Następny etap badania stanowiło wykonanie hierarchicznej klasyfikacji aglomeracyjnej do klasyfikacji obiektów, czyli 27 krajów UE na podstawie przeciętnych wyników gospodarstwa rolnego według 27 cech dotyczących 2008 roku, które zostały znormalizowane. Wyniki badań przedstawiono na rys. 1-2. Warto dodać, że aby ustalona liczba klas do interpretacji była najwłaściwsza, należy dokonać wyboru między

małą liczbą klas zawierających więcej zróżnicowanych obiektów a dużą liczbą klas o mniejszej liczebności obiektów. Przeprowadzając badanie, starano się uwzględnić fakt, że produkcja rolna w krajach UE-27 jest wysoce zindywidualizowana, gdyż w każdym z krajów występują swoiste cechy rolnictwa, determinujące warunki prowadzenia działalności przez gospodarstwa rolne (m.in.: klimat, tradycja, dostępność środków finansowych).

Na podstawie wykresu liniowego odległości wiązań względem kolejnych etapów procesu wiązania ustalono właściwą liczbę klas (rys. 2). Jest on przydatny do identyfikowania miejsc, w których formuje się wiele skupień w przybliżeniu w takiej samej odległości wiązania (Stanisz, 2007). Jak można zauważyć, największy przyrost wariancji wewnątrzklasowej wystąpił przy połączeniu dwóch rozłącznych klas w jedną (27-mio elementową), dlatego najlepszy byłby podział zbioru na dwie klasy (klasę 1 z 21-oma elementami i klasę 2 z 6-cioma), ale taki podział nie przyniósłby znaczących informacji. Dlatego wybrano drugi najwyższy przyrost na poziomie odległości wiązań równej 7,00. W ten sposób powstało 10 klas krajów UE-27.



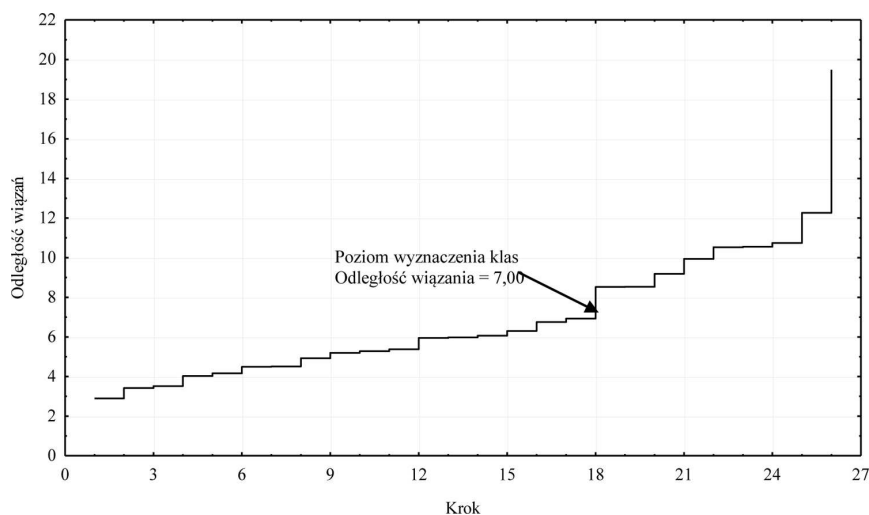
BEL – Belgia, BGR – Bułgaria, CYP – Cypr, CZE – Czechy, DAN – Dania, DEU – Niemcy, ELL – Grecja, ESP – Hiszpania, EST – Estonia, FRA – Francja, HUN – Węgry, IRE – Irlandia, ITA – Włochy, LTU – Litwa, LUX – Luksemburg, LVA – Łotwa, MLT – Malta, NED – Holandia, OST – Austria, POL – Polska, POR – Portugalia, ROU – Rumunia, SUO – Finlandia, SVE – Szwecja, SVK – Słowacja, SVN – Słowenia, UKI – Wielka Brytania.

Rysunek 1. Dendrogram struktury i skali produkcji gospodarstw rolnych według kraju UE-27 na podstawie 27 cech (odległości euklidesowe, metoda Warda)

Źródło: Opracowanie własne na podstawie danych FADN (2012) przy użyciu programu Statistica 10.

Sprawdzono również czy cechy w sposób istotny różnicują otrzymane klasy, wykonując analizę wariancji w zakładce Testy ANOVA dotyczących opcji „Statystyk w grupach” programu Statistica 10. Okazało się, że aż 19 na 27 cech posiadało poziom istotności *ex post* dla odpowiedniego testu F, tj. *p*, niższy od 0,05, a 8 cech, które





Rysunek 2. Przebieg aglomeracji na podstawie wykresu odległości wiązania względem etapów wiązania

Źródło: Jak w przypadku Rys. 1

przekroczyły ten poziom to składowe struktury produkcji i ich wykluczenie byłoby niemerytoryczne.

Opis wyznaczonych klas krajów UE-27 opisanych przeciętnymi danymi dla gospodarstw rolnych na podstawie hierarchicznej klasyfikacji aglomeracyjnej przedstawiono w tab. 1-3. W tabeli 1 zamieszczono średnie klasowe opisujące pogrupowane kraje. Dodano również informacje o odległości euklidesowej, przy której nastąpiło utworzenie danej klasy. W tabeli 2 zawarto wartości miernika różnic średnich dla analizowanych cech w zbiorowości 27 krajów UE według wyszczególnionych klas. Natomiast w tabeli 3 opisano klasy typologiczne krajów UE-27 według analizowanych cech w 2008 roku, scharakteryzowane przez obliczony miernik różnic średnich.

Klasę 1 utworzyły Belgia i Holandia o produkcji roślinno-zwierzęcej, ze znacznym udziałem warzyw i kwiatów, mleka krowiego i produktów mlecznych oraz wieprzowiny. Średnio biorąc, kraje te reprezentowały gospodarstwa większe niż średnio zaobserwowane w UE-27, które generowały produkcję ok. 5-krotnie wyższą i 2-krotnie wyższy dochód od średniej unijnej w 2008 roku (tab. 1-3).

Aż trzy klasy (2, 7 i 9) składały się z pojedynczych państw, co świadczy, że w tych krajach produkcja rolnicza zachowała cechy szczególne. W Danii (klasa 2), produkcja miała charakter głównie zwierzęcy i produkowano tam znaczne ilości wieprzowiny, ponadto jeszcze wytwarzano dość dużo mleka krowiego i produktów mlecznych przy istotnym udziale zbóż (tab. 1). Mimo bardzo dużej wielkości ekonomicznej i produkcji ogółem, dochód z gospodarstwa rolnego był ujemny (tab. 1-3) na skutek wysokich wartości zużycia pośredniego i kosztów czynników zewnętrznych (FADN 2012). Zużycie pośrednie obejmuje koszty bezpośrednie (z uwzględnieniem produktów wytworzonych w gospodarstwie rolnym) oraz koszty ogólnogospodarcze towarzyszące działalności

Tabela 1.  
Średnie klasowe charakteryzujące klasy typologiczne krajów UE-27 według struktury i skali produkcji gospodarstw rolnych w 2008 roku

|                                       | Klasa       |       |                             |                                                     |             |                     |       |             |       |                     |
|---------------------------------------|-------------|-------|-----------------------------|-----------------------------------------------------|-------------|---------------------|-------|-------------|-------|---------------------|
|                                       | 1           | 2     | 3                           | 4                                                   | 5           | 6                   | 7     | 8           | 9     | 10                  |
| Wyszczególnienie                      | BEL,<br>NED | DAN   | BGR,<br>HUN,<br>POL,<br>ROU | CZE,<br>DEU,<br>EST,<br>SVK,<br>FRA,<br>LUX,<br>UKI | IRE,<br>SVN | EST,<br>LTU,<br>LVA | SVK   | CYP,<br>MLT | ELL   | ESP,<br>ITA,<br>POR |
| 1. zboża (%)                          | 3,54        | 13,55 | 23,09                       | 17,25                                               | 6,14        | 22,63               | 25,23 | 0,25        | 13,27 | 11,09               |
| 2. rośliny białkowe (%)               | 0,03        | 0,12  | 0,17                        | 0,26                                                | 0,03        | 0,36                | 0,29  | 0,10        | 1,36  | 0,23                |
| 3. uprawy energetyczne (%)            | 0,09        | 0,29  | 0,12                        | 0,40                                                | 0,09        | 2,23                | 2,51  | 0,00        | 0,00  | 0,00                |
| 4. ziemniaki (%)                      | 3,93        | 1,66  | 2,34                        | 2,53                                                | 1,48        | 3,35                | 0,93  | 4,76        | 1,83  | 1,83                |
| 5. buraki cukrowe (%)                 | 1,73        | 0,69  | 0,56                        | 1,47                                                | 0,01        | 0,38                | 1,12  | 0,00        | 0,29  | 0,45                |
| 6. rośliny oleiste (%)                | 0,15        | 2,31  | 7,57                        | 3,79                                                | 0,29        | 6,89                | 11,41 | 0,00        | 0,16  | 0,44                |
| 7. rośliny przemysłowe (%)            | 0,38        | 0,03  | 1,02                        | 0,37                                                | 0,73        | 0,07                | 0,15  | 0,00        | 3,34  | 0,58                |
| 8. warzywa i kwiaty (%)               | 25,45       | 5,77  | 9,41                        | 5,67                                                | 1,17        | 3,76                | 1,10  | 28,72       | 12,38 | 14,04               |
| 9. owoce (%)                          | 3,46        | 0,30  | 3,36                        | 1,31                                                | 1,35        | 0,47                | 0,39  | 2,73        | 8,73  | 7,00                |
| 10. owoce cytrusowe (%)               | 0,00        | 0,00  | 0,00                        | 0,00                                                | 0,00        | 0,00                | 0,00  | 1,72        | 3,69  | 2,67                |
| 11. wina i winogrona (%)              | 0,00        | 0,00  | 2,96                        | 4,76                                                | 4,24        | 0,00                | 0,74  | 3,31        | 7,00  | 10,88               |
| 12. oliwki i olej z oliwek (%)        | 0,00        | 0,00  | 0,00                        | 0,01                                                | 0,08        | 0,00                | 0,00  | 1,65        | 13,78 | 6,88                |
| 13. uprawy pastewne (%)               | 0,51        | 6,61  | 3,63                        | 3,12                                                | 16,76       | 9,28                | 9,60  | 2,31        | 5,09  | 5,36                |
| 14. inna produkcja roślinna (%)       | 4,16        | 2,68  | 1,44                        | 1,22                                                | 0,35        | 1,34                | 2,76  | 0,32        | 0,51  | 1,98                |
| 15. mleko krowie i jego przetwory (%) | 17,88       | 18,45 | 11,28                       | 25,15                                               | 28,90       | 25,40               | 18,28 | 13,91       | 2,19  | 13,90               |
| 16. żywiec wołowy (%)                 | 7,66        | 3,55  | 2,70                        | 9,85                                                | 24,04       | 4,97                | 3,32  | 1,76        | 1,10  | 5,99                |
| 17. żywiec wieprzowy (%)              | 17,44       | 26,86 | 7,29                        | 8,91                                                | 1,93        | 6,71                | 3,45  | 12,81       | 1,74  | 6,21                |

cd. Tabela 1.

|                                                 |        |        |       |        |       |       |        |       |       |       |
|-------------------------------------------------|--------|--------|-------|--------|-------|-------|--------|-------|-------|-------|
| 18. żywiec barani i kozi (%)                    | 0,13   | 0,10   | 1,71  | 0,94   | 2,83  | 0,34  | 0,25   | 4,66  | 5,61  | 2,40  |
| 19. żywiec drobiowy (%)                         | 2,92   | 2,20   | 5,01  | 2,02   | 0,06  | 0,19  | 2,57   | 5,08  | 1,28  | 0,50  |
| 20. jaja (%)                                    | 2,11   | 2,03   | 1,69  | 1,30   | 0,26  | 4,95  | 1,82   | 8,13  | 1,24  | 1,19  |
| 21. mleko owcze i kozie oraz jego przetwory (%) | 0,87   | 0,00   | 1,30  | 0,21   | 0,01  | 0,08  | 0,45   | 6,23  | 10,66 | 2,51  |
| 22. inna produkcja zwierzęca (%)                | 1,39   | 6,03   | 1,08  | 0,93   | 2,23  | 0,95  | 0,34   | 0,48  | 1,91  | 1,33  |
| 23. inna produkcja (%)                          | 6,17   | 6,77   | 12,27 | 8,53   | 7,02  | 5,65  | 13,29  | 1,07  | 2,84  | 2,54  |
| 24. wielkość ekonomiczna (ESU)                  | 131,45 | 114,00 | 11,45 | 71,20  | 15,35 | 15,07 | 130,00 | 19,40 | 10,80 | 26,80 |
| 25. powierzchnia UR (ha)                        | 39,03  | 82,57  | 27,89 | 101,53 | 28,37 | 81,32 | 579,35 | 5,80  | 7,09  | 25,97 |
| 26. produkcja ogółem (tys. euro)                | 310,83 | 310,36 | 38,40 | 170,55 | 35,29 | 55,08 | 526,51 | 45,26 | 19,65 | 44,51 |
| 27. dochód z gospodarstwa rolnego (tys. euro)   | 36,78  | -45,82 | 9,92  | 32,83  | 12,51 | 16,29 | -2,46  | 14,56 | 11,89 | 20,61 |
| Odstęłość euklidesowa                           | 5,38   | -      | 6,30  | 6,07   | 6,75  | 5,20  | -      | 6,92  | -     | 4,93  |

BEL – Belgia, BGR – Bułgaria, CYP – Cypr, CZE – Czechy, DAN – Dania, DEU – Niemcy, ELL – Grecja, ESP – Hiszpania, EST – Estonia, FRA – Francja, HUN – Węgry, IRE – Irlandia, ITA – Włochy, LTU – Litwa, LUX – Luksemburg, LVA – Łotwa, MLT – Malta, NED – Holandia, OST – Austria, POL – Polska, POR – Portugalia, ROU – Rumunia, SLO – Słowacja, SVK – Słowacja, SVN – Słowenia, UKI – Wielka Brytania.

Źródło: Opracowanie własne w oparciu o dane FADN (2012) przy użyciu programu Statistica 10.

Tabela 2.  
Wartości miernika różnic średnich dla 27 cech struktury i skali produkcji gospodarstw rolnych z krajów UE-27 w 2008 roku według opracowanych klas typologicznych

| Wyszczególnienie                         | Klasa      |       |                          |                                               |            |                   |       |            |       |                   |
|------------------------------------------|------------|-------|--------------------------|-----------------------------------------------|------------|-------------------|-------|------------|-------|-------------------|
|                                          | 1          | 2     | 3                        | 4                                             | 5          | 6                 | 7     | 8          | 9     | 10                |
|                                          | BEL<br>NED | DAN   | BGR<br>HUN<br>POL<br>ROU | CZE<br>DEU<br>EST<br>SVE<br>FRA<br>LUX<br>UKI | IRE<br>SVN | EST<br>LTU<br>LVA | SVK   | CYP<br>MLT | ELL   | ESP<br>ITA<br>POR |
| 1. zboża                                 | -2,29      | -0,28 | 1,54                     | 0,38                                          | -1,78      | 1,86              | 2,34  | -2,93      | -0,40 | -0,82             |
| 2. rośliny białkowe                      | -1,07      | -0,66 | -0,40                    | 0,08                                          | -1,10      | 0,59              | 0,25  | -0,73      | 5,68  | -0,09             |
| 3. uprawy energetyczne                   | -1,28      | -0,68 | -1,21                    | -0,31                                         | -1,28      | 5,40              | 6,26  | -1,57      | -1,57 | -1,56             |
| 4. ziemniaki                             | 1,08       | -0,75 | -0,20                    | -0,05                                         | -0,89      | 0,61              | -1,34 | 1,75       | -0,61 | -0,61             |
| 5. buraki cukrowe                        | 1,28       | -0,18 | -0,36                    | 0,92                                          | -1,14      | -0,61             | 0,43  | -1,15      | -0,75 | -0,51             |
| 6. rośliny oleiste                       | -1,44      | -0,55 | 1,64                     | 0,07                                          | -1,39      | 1,36              | 3,40  | -1,50      | -1,44 | -1,32             |
| 7. rośliny przemysłowe                   | -0,21      | -0,65 | 0,59                     | -0,21                                         | 0,22       | -0,59             | -0,50 | -0,68      | 3,46  | 0,04              |
| 8. warzywa i kwiaty                      | 3,30       | -0,86 | -0,10                    | -0,89                                         | -1,83      | -1,29             | -1,85 | 3,97       | 0,53  | 0,88              |
| 9. owoce                                 | 0,65       | -1,80 | 0,57                     | -1,02                                         | -0,98      | -1,66             | -1,73 | 0,08       | 4,73  | 3,39              |
| 10. owoce cytrusowe                      | -0,68      | -0,68 | -0,68                    | -0,68                                         | -0,68      | -0,68             | -0,68 | 1,41       | 3,80  | 2,56              |
| 11. wina i winogrona                     | -0,84      | -0,84 | -0,20                    | 0,18                                          | 0,07       | -0,84             | -0,68 | -0,13      | 0,67  | 1,50              |
| 12. oliwki i olej z oliwek               | -1,25      | -1,25 | -1,25                    | -1,24                                         | -1,18      | -1,25             | -1,25 | 0,22       | 11,01 | 4,87              |
| 13. uprawy pastewne                      | -1,44      | 0,38  | -0,50                    | -0,66                                         | 3,40       | 1,18              | 1,27  | -0,90      | -0,07 | 0,01              |
| 14. inna produkcja roślinna              | 2,25       | 0,98  | -0,07                    | -0,26                                         | -1,00      | -0,16             | 1,06  | -1,02      | -0,86 | 0,39              |
| 15. mleko i przet. z mleka krowiego      | -0,24      | -0,16 | -1,21                    | 0,88                                          | 1,39       | 0,86              | -0,19 | -0,83      | -2,55 | -0,83             |
| 16. żywioł wołowy                        | 0,07       | -0,70 | -0,86                    | 0,47                                          | 3,13       | -0,44             | -0,75 | -1,04      | -1,16 | -0,25             |
| 17. żywioł wieprzowy                     | 2,09       | 4,36  | -0,34                    | 0,04                                          | -1,63      | -0,48             | -1,27 | 0,98       | -1,68 | -0,61             |
| 18. żywioł barani i kozi                 | -0,66      | -0,67 | 0,04                     | -0,30                                         | 0,53       | -0,57             | -0,61 | 1,34       | 1,76  | 0,34              |
| 19. żywioł drobiowy                      | 0,36       | -0,02 | 1,46                     | -0,12                                         | -1,15      | -1,08             | 0,17  | 1,50       | -0,51 | -0,91             |
| 20. jaja                                 | -0,08      | -0,12 | -0,28                    | -0,46                                         | -0,95      | 1,26              | -0,22 | 2,76       | -0,49 | -0,52             |
| 21. ml. i przet. z ml. owczego i koziego | -0,41      | -0,97 | -0,12                    | -0,83                                         | -0,97      | -0,92             | -0,68 | 3,12       | 6,03  | 0,68              |
| 22. inna produkcja zwierzęca             | 0,10       | 5,27  | -0,24                    | -0,41                                         | 1,04       | -0,39             | -1,07 | -0,91      | 0,68  | 0,03              |
| 23. inna produkcja                       | -0,14      | -0,06 | 0,74                     | 0,20                                          | -0,02      | -0,22             | 0,89  | -0,88      | -0,63 | -0,67             |

cd. Tabela 2.

|                                   |             |              |       |      |       |       |              |       |       |       |
|-----------------------------------|-------------|--------------|-------|------|-------|-------|--------------|-------|-------|-------|
| 24. wielkość ekonomiczna          | <b>4,00</b> | <b>3,15</b>  | -1,84 | 1,07 | -1,65 | -1,66 | <b>3,93</b>  | -1,45 | -1,87 | -1,09 |
| 25. powierzchnia UR               | -0,84       | 0,14         | -1,09 | 0,57 | -1,08 | 0,11  | <b>11,33</b> | -1,59 | -1,56 | -1,13 |
| 26. produkcja ogółem              | <b>2,92</b> | <b>2,92</b>  | -1,43 | 0,68 | -1,48 | -1,17 | <b>6,37</b>  | -1,32 | -1,73 | -1,34 |
| 27. dochód z gospodarstwa rolnego | <b>2,29</b> | <b>-8,14</b> | -1,11 | 1,79 | -0,78 | -0,30 | <b>-2,67</b> | -0,52 | -0,86 | 0,24  |

BEL – Belgia, BGR – Bułgaria, CYP – Cypr, CZE – Czechy, DAN – Dania, DEU – Niemcy, ELL – Grecja, ESP – Hiszpania, EST – Estonia, FRA – Francja, HUN – Węgry, IRE – Irlandia, ITA – Włochy, LTU – Litwa, LUX – Luksemburg, LVA – Łotwa, MLT – Malta, NED – Holandia, OST – Austria, POL – Polska, POR – Portugalia, ROU – Rumunia, SUO – Finlandia, SVE – Szwecja, SVK – Słowacja, SVN – Słowenia, UKI – Wielka Brytania.

Odcienie szarości oznaczają średnie i wysokie wartości bezwzględne miernika  $Z_{ek(d)}$ , które były podstawą do wyróżnienia cech charakterystycznych w klasach na podstawie wyników klasyfikacji rozłącznej i cech merycznych polegającą na porównaniu średnich wewnętrznych klas ze średnimi ogólnymi otrzymanymi z całej zbiorowości obiektów.

Jasnoszary – natężenie dość duże.

Ciemnoszary – natężenie bardzo duże.

Źródło: Opracowanie własne na podstawie danych z Tabl. 1. przy użyciu programu Statistica 10.

Tabela 3.

Charakterystyka klas typologicznych krajów UE-27 według struktury i skali produkcji w gospodarstwach rolnych w 2008 roku

| Klasa typologiczna | Kraje                                                                             | Liczba krajów | Opis typu                                                                                                                                                                                                                                                                                                                              |
|--------------------|-----------------------------------------------------------------------------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1                  | Belgia, Holandia                                                                  | 2             | produkcja roślinno-zwierzęca ze znacznym udziałem warzyw i kwiatów oraz innej produkcji roślinnej i wieprzowiny, wytwarzana przez gospodarstwa o bardzo dużej wielkości ekonomicznej, osiągające znaczącą wartość produkcji ogółem i dochodu z gospodarstwa rolnego                                                                    |
| 2                  | Dania                                                                             | 1             | produkcja zwierzęca z dużym udziałem wieprzowiny i innej produkcji zwierzęcej przez gospodarstwa o bardzo dużej wielkości ekonomicznej, osiągające znaczącą wartość produkcji ogółem, ale bardzo niski dochód z gospodarstwa rolnego                                                                                                   |
| 3                  | Bułgaria, Węgry, Polska, Rumunia                                                  | 4             | produkcja mieszana, głównie roślinna z dominującą rolą zbóż, warzyw i kwiatów, roślin oleistych, przy znacznym udziale produkcji mleka krowiego i jego przetworów oraz żywca drobiowego przez gospodarstwa o małej wielkości ekonomicznej i powierzchni UR, osiągające niską wartość produkcji ogółem i dochodu z gospodarstwa rolnego |
| 4                  | Czechy, Niemcy, Austria, Finlandia, Szwecja, Francja, Luksemburg, Wielka Brytania | 8             | produkcja roślinno-zwierzęca, ze znaczną produkcją mleka krowiego i produktów mlecznych, żywca wołowego, warzyw i kwiatów oraz win i winogron przez gospodarstwa dużej wielkości ekonomicznej, osiągające znaczącą wartość produkcji ogółem i dochodu z gospodarstwa rolnego                                                           |
| 5                  | Irlandia, Słowenia                                                                | 2             | produkcja zwierzęca z dużym udziałem żywca wołowego, mleka krowiego i innych produktów mlecznych oraz upraw pastewnych przez gospodarstwa małe o niskiej produkcji i dochodzie z gospodarstwa                                                                                                                                          |
| 6                  | Estonia, Litwa, Łotwa                                                             | 3             | produkcja roślinna z wyraźnym udziałem upraw energetycznych, zbóż, upraw pastewnych oraz produkcji mleka krowiego i produktów mlecznych przez gospodarstwa małe ekonomicznie, ale o dużej powierzchni UR, osiągające niską wartość produkcji i dochodu z gospodarstwa rolnego                                                          |
| 7                  | Słowacja                                                                          | 1             | produkcja roślinna z dużym udziałem zbóż, upraw energetycznych i roślin oleistych przez gospodarstwa o bardzo dużej wielkości ekonomicznej i powierzchni UR, osiągające znaczącą wartość produkcji ogółem, ale bardzo niski dochód z gospodarstwa rolnego                                                                              |
| 8                  | Cypr, Malta                                                                       | 2             | produkcja zwierzęca ze znacznym udziałem produkcji mleka owczego i koziego oraz jego przetworów, jaj drobiowych oraz warzyw i kwiatów przez gospodarstwa małe o niskiej produkcji i dochodzie z gospodarstwa rolnego                                                                                                                   |
| 9                  | Grecja                                                                            | 1             | produkcja roślinna z dużym udziałem oliwek i olei roślinnych, warzyw i kwiatów, owoców, owoców cytrusowych i roślin przemysłowych przy dużym udziale produkcji owczego i koziego mleka przez gospodarstwa o małej wielkości ekonomicznej i powierzchni UR, osiągające niską wartość produkcji ogółem i dochodu z gospodarstwa rolnego  |
| 10                 | Hiszpania, Włochy, Portugalia                                                     | 3             | produkcja roślinna z wyraźnym udziałem produkcji oliwek i olei roślinnych, owoców, owoców cytrusowych, win i winogron oraz warzyw i kwiatów przez gospodarstwa relatywnie małe z niską wartością produkcji, ale o znaczącym dochodzie z gospodarstwa rolnego                                                                           |

Źródło: Opracowanie własne na podstawie danych z Tabl.2

operacyjnej w roku obrachunkowym, a koszty czynników zewnętrznych to koszty zaangażowania obcych czynników wytwórczych (pracy, ziemi i kapitału) w procesie produkcyjnym, zalicza się do nich: wynagrodzenia za pracę, czynsze i odsetki (IERiGŻ, 2010). W przypadku Danii, średnia produkcja ogółem w 2008 roku wyniosła 310,36 tys. euro, należało od niej odjąć (w tys. euro): zużycie pośrednie -225,68, koszty czynników zewnętrznych -124,77 i amortyzację -35,51, a dodać saldo dopłat i podatków z działalności operacyjnej i inwestycyjnej, czyli 29,78 tys. euro. Wobec tego dochód z gospodarstwa rolnego wyniósł -46,82 tys. euro (FADN, 2012). Natomiast na Słowacji (klasa 7) dominowała produkcja roślinna skupiająca się na zbożach, roślinach oleistych, uprawach energetycznych i paszy, przy dużym udziale produkcji mleka krowiego i innych produktów mlecznych. Przeciętnie było to gospodarstwo o bardzo dużej wielkości ekonomicznej, posiadające bardzo dużo użytków rolnych i osiągające bardzo wysoką produkcję ogółem, ale wypracowywało ujemny dochód z gospodarstwa rolnego (także na skutek wysokich wartości zużycia pośredniego i kosztów czynników zewnętrznych). Dla Słowacji średnia produkcja ogółem w 2008 roku wyniosła 526,51 tys. euro, gdy od niej odjęto (w tys. euro): zużycie pośrednie -451,66, koszty czynników zewnętrznych -148,76 i amortyzację -81,67, a dodano saldo dopłat i podatków z działalności operacyjnej i inwestycyjnej, czyli 153,12 tys. euro, wykazano dochód z gospodarstwa rolnego w wysokości -2,46 tys. euro (FADN, 2012). W przypadku Grecji (klasa 9) realizowana była głównie produkcja roślinna, w której przeważały oliwki i oleje roślinne, a także warzywa i kwiaty, owoce oraz owoce cytrusowe tudzież rośliny przemysłowe przy dużym udziale produkcji owczego i koziego mleka oraz jego przetworów. Przeciętnie rzecz ujmując, to gospodarstwo było ok. 4-krotnie mniejsze i produkowało 3-krotnie mniej niż średnio w UE-27 (tab. 1-3).

W klasie 3 znalazły się 4 państwa, a mianowicie Bułgaria, Węgry, Polska i Rumunia. Produkcja gospodarstw miała charakter mieszany, głównie roślinny z dominującą rolą zbóż, warzyw i kwiatów, roślin oleistych, przy znacznym udziale produkcji mleka krowiego i jego przetworów oraz żywca drobiowego (tab. 2-3). Wyrażna – ale poniżej średniej unijnej – była produkcja warzyw i kwiatów oraz mleka i przetworów z mleka krowiego. Jednakże przeciętnie rzecz biorąc, gospodarstwa wypracowywały połowę produkcji i dochodu obserwowanego średnio w UE-27 (tab.1-3).

Do najliczniejszej klasy 4, zakwalifikowano aż 8 państw. Były nimi: Czechy, Niemcy, Austria, Finlandia, Szwecja, Francja, Luksemburg i Wielka Brytania. Ich produkcja miała charakter mieszany (roślinno-zwierzęcy) ze znacznym udziałem produkcji mleka krowiego i produktów mlecznych, z wyraźnym udziałem produkcji żywca wołowego, warzyw i kwiatów oraz win i winogron. Gospodarstwa tych krajów na tle UE-27 były relatywnie duże z ok. 3-krotnie wyższą produkcją i 2-krotnie wyższym dochodem niż średnio obserwowany w 2008 roku (tab. 1-3).

Klasa 5 objęła Irlandię i Słowenię, które pomimo różnego położenia geograficznego, realizowały produkcję zwierzęcą z dużym udziałem żywca wołowego, mleka krowiego i innych produktów mlecznych przy znacznym udziale upraw pastewnych. Były to gospodarstwa małe o niskiej produkcji i dochodzie z gospodarstwa (tab. 1-3).

W klasie 6 znalazły się: Estonia, Litwa i Łotwa, w których przeważała produkcja roślinna ze znacznym udziałem upraw energetycznych na tle UE-27 oraz z istotnym udziałem zbóż, upraw pastewnych tudzież produkcji mleka krowiego i produktów mlecznych. Były to gospodarstwa relatywnie duże, gdyż ich powierzchnia była ok. 2,5-krotnie wyższa niż średnio w UE-27, ale ich produkcja i dochód pozostały na poziomie ok. 88% poziomu unijnego (tab.1-3).

Klasę 8 utworzyły Cypr i Malta, gdzie przeciętnie gospodarstwa rolne charakteryzowała produkcja głównie zwierzęca, ale ze znacznym udziałem produkcji mleka owczego i koziego oraz jego przetworów, jaj drobiowych oraz warzyw i kwiatów. Były to przeciętnie gospodarstwa ok. 6-krotnie mniejsze od średniej unijnej, a ich produkcja osiągała tylko ok. 70% wartości średniej w UE-27, a dochód ok. 80% (tab.1-3).

W klasie 10 znalazły się: Hiszpania, Włochy i Portugalia z dominującą produkcją roślinną z wyraźnym udziałem produkcji oliwek i olei roślinnych, owoców, owoców cytrusowych, win i winogron oraz warzyw i kwiatów. Przeciętnie, gospodarstwa rolne osiągnęły w 2008 roku produkcję na poziomie 70% średniej z UE-27, ale ich dochód przekroczyło ok. 15% średni w UE-27 przy relatywnie małej wielkości ekonomicznej i powierzchni UR (tab. 1-3).

#### 4. PODSUMOWANIE

W wyniku zastosowania hierarchicznej klasyfikacji aglomeracyjnej, analizy wielkości przyrostów wariancji wewnątrzklasowej w procesie łączenia klas między poszczególnymi poziomami łączy oraz miernika różnic średnich, wyznaczono 10 klas krajów o różnej strukturze i skali produkcji realizowanej przez gospodarstwa rolne w UE-27 w 2008 roku. Aż 3 klasy wykazały cechy szczególne, gdyż składały się tylko z jednego obiektu.

Polska zaklasyfikowana została do klasy 3-ciej obok Bułgarii, Węgier i Rumunii. Produkcja gospodarstw rolnych z tej klasy miała charakter mieszany, głównie roślinny z dominującą rolą zbóż, roślin oleistych, przy znacznym udziale produkcji żywca drobiowego. Zajmowano się również produkcją warzyw i kwiatów oraz mleka i przetworów z mleka krowiego. Jednakże przeciętnie rzecz ujmując, były to gospodarstwa wypracowujące połowę produkcji i dochodu obserwowanego średnio w UE-27.

*Uniwersytet Przyrodniczy w Poznaniu*

#### LITERATURA

- [1] Anderberg M.R., (1973), *Cluster analysis for applications*, Academic Press, Nowy Jork, San Francisco, Londyn.
- [2] Bański J., (2007), *Geografia rolnictwa Polski*, PWE, Warszawa.
- [3] Chmielewska B., Mierosławska A., (2007), *Krajowa czy regionalna strategia rozwoju rolnictwa i obszarów wiejskich*, Zagadnienia Ekonomiki Rolnej, 2, 85-105.
- [4] Everitt B.S., Landau S., Leese M., Stahl D., (2011), *Cluster analysis*, 5th Edition, Wiley, Chichester.



- [5] FADN, (2012), [http://ec.europa.eu/agriculture/rica/database/database\\_en.cfm](http://ec.europa.eu/agriculture/rica/database/database_en.cfm), (dostęp luty 2012).
- [6] Fierla I., (red.), (2011), *Geografia ekonomiczna Unii Europejskiej*, PWE, Warszawa.
- [7] Filipiak K., (2006), *Metody statystyczne stosowane do oceny regionalnego zróżnicowania rolnictwa*, [w:] A. Harasim (red.), Regionalne zróżnicowanie produkcji rolniczej w Polsce, Raporty PIB nr 3, IUNG – PIB, Puławy, 53-60.
- [8] Harasim A., (2006), *Dobór wskaźników do oceny regionalnego zróżnicowania rolnictwa*, [w:] Harasim A., (red.), Regionalne zróżnicowanie produkcji rolniczej w Polsce, Raporty PIB nr 3, IUNG – PIB, Puławy, 61-69.
- [9] IERiGŻ, (2010), *Wyniki standardowe uzyskane przez gospodarstwa rolne uczestniczące w Polskim FADN w 2009 roku. Część I. Wyniki standardowe*, Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej – PIB, Warszawa.
- [10] Kisielińska J., (2009), *Bezwzorcowa klasyfikacja obiektów w ekonomice rolnictwa*, Zeszyty Naukowe SGGW w Warszawie „Problemy Rolnictwa Światowego”, t. 8 (XXIII), Wydawnictwo SGGW, Warszawa, 104-115.
- [11] Stanisław A., (2007), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, T. 3. Analizy wielowymiarowe, StatSoft, Kraków.
- [12] Stuczyński T., Jadczyński J., Kukuła S., (2006), *Wykorzystanie systemu informacji o rolniczej przestrzeni produkcyjnej do analiz regionalnych*, [w:] A. Harasim (red.), Regionalne zróżnicowanie produkcji rolniczej w Polsce, Raporty PIB nr 3, IUNG – PIB, Puławy, 33-52.
- [13] Walesiak M., (2004), *Metody klasyfikacji*, [w:] E. Gatnar, M. Walesiak (red.), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej im. O. Łangego we Wrocławiu, Wrocław, 316-350.
- [14] Walesiak M., Dudek A., (2009), *Ocena wybranych procedur analizy skupień dla danych porządkowych*, [w:] K. Jajuga, M. Walesiak (red.), *Taksonomia 16, Klasyfikacja i analiza danych – teoria i zastosowania*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 47, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, 41-49.
- [15] Wysocki F., (1996), *Metody statystycznej analizy wielowymiarowej w rozpoznawaniu typów struktury przestrzennej rolnictwa*, Rozprawy Naukowe z. 266, Roczniki Akademii Rolniczej w Poznaniu, Poznań.
- [16] Wysocki F., (2010), *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*, Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu, Poznań.

#### ZASTOSOWANIE HIERARCHICZNEJ KLASYFIKACJI AGLOMERACYJNEJ DO GRUPOWANIA KRAJÓW UNII EUROPEJSKIEJ ZE WZGLĘDU NA STRUKTURĘ I SKALĘ PRODUKCJI GOSPODARSTW ROLNYCH

##### Streszczenie

Celem opracowania była próba delimitacji krajów w UE-27 o podobnej strukturze i skali produkcji pochodzącej z gospodarstw rolnych. W badaniu położono nacisk na uwypuklenie podobieństwa struktur produkcji typowych dla obszarów klimatycznych UE-27 w 2008 roku, a także na wskazanie klasy, do której zakwalifikowano polskie rolnictwo. Wykorzystano dane z bazy FADN, którymi były średnie ważone przeliczone na gospodarstwo rolne z 27 krajów członkowskich UE w 2008 roku. Delimitację wykonano w oparciu o strukturę produkcji, wielkość ekonomiczną, powierzchnię użytków rolnych, wartość produkcji i dochodów z gospodarstwa rolnego. Badanie wykonano za pomocą hierarchicznej klasyfikacji aglomeracyjnej metodą Warda. Przy czym strukturę produkcji reprezentowały 23 cechy, a pozostałe cechy były pojedyncze. Optymalną liczbę klas krajów o podobnej strukturze i skali produkcji wyznaczono na podstawie analizy wielkości przyrostów wariancji wewnątrzklasowej w procesie łączenia klas między

poszczególnymi poziomami łączeń. W ten sposób uzyskano 10 klas krajów. Klasy opisano według obliczonego miernika różnic średnich dla analizowanych cech.

**Słowa kluczowe:** hierarchiczna klasyfikacja aglomeracyjna, metoda Warda, produkcja

#### THE USE OF THE AGGLOMERATION HIERARCHICAL CLASSIFICATION TO GROUP THE EU COUNTRIES ACCORDING TO THE STRUCTURE AND SCALE OF FARMS' PRODUCTION

##### A b s t r a c t

The aim of the article was an attempt at delimitation of countries in the EU-27 with similar structure and scale of the production of farms. In this study, the emphasis is placed on highlighting the similarity of production structures typical of the climate's areas of the EU-27 in the year 2008, and also on indicating the class in which the Polish agriculture were classified. For this research, the data from FADN database was used. This database contains weighted averages per individual farm in every of 27 UE member countries in 2008. The delimitation were made based on the structure of production, economic size, agricultural utilized area, production and income of agricultural farm. The research was conducted using the agglomeration hierarchic classification based on the Ward's method. The structure of production is represented by 23 characteristics; the remaining characters are represented by single variables. The optimal number of classes of countries of similar structure and scale of production was estimated on the basis of the analysis of the amount of the intra-class variance's increment. 10 classes of countries were obtained. These classes were described by the measure of the average differences for analyzed characters.

**Key words:** agglomeration hierarchical classification, Ward's method, production

HANNA GRUCHOCIAK

## DELIMITACJA LOKALNYCH RYNKÓW PRACY W POLSCE

### 1. WPROWADZENIE

Delimitacja lokalnych rynków pracy wydaje się zagadnieniem niezmiernie istotnym z punktu widzenia wspomagania zrównoważonego rozwoju regionów. Przed próbą podziału Polski na obszary, które nazwane mogłyby być lokalnymi rynkami pracy należy określić, co badacz rozumie pod pojęciem lokalnego rynku pracy oraz jakimi cechami powinien się on charakteryzować. W literaturze lokalny rynek pracy określa się jako ekonomicznie zintegrowany obszar geograficzny, w ramach którego zamieszkująca ludność może znaleźć zatrudnienie bądź je zmienić, bez konieczności zmiany miejsca zamieszkania (por. Gołata, 2004). Tak więc jest to taki obszar, wewnątrz którego zamieszkująca tam ludność jest skłonna dojeżdżać do pracy. Można przyjąć, że gotowość ta zdeterminowana będzie dwiema grupami charakterystyk: dotyczących dostępności i atrakcyjności pracy oraz określających łatwość codziennego przejazdu. Do pierwszej grupy zaliczyć można by liczbę przeprowadzanych rekrutacji oraz na jakie stanowiska poszukiwano kandydatów, ile osób poszukuje pracy na terenie danego obszaru oraz wszelkie charakterystyki warunków pracy, takie jak wynagrodzenie czy wymiar zatrudnienia, ale również wiele zmiennych niemierzalnych określających komfort pracy. Do drugiej grupy zaliczyć należy przede wszystkim odległość, która mierzona może być w kilometrach, zarówno w linii prostej jak i uwzględniając drogę samochodową czy kolejową lub biorąc pod uwagę koszty (cena biletu, koszt benzyny itp.). Tak więc, aby podjąć się delimitacji lokalnych rynków pracy a priori, należałoby określić grupę czynników charakteryzujących rynek pracy oraz łatwość dojazdu. Wiele z powyżej opisanych czynników jest jednak niemierzalnych a informacje na temat większości z nich są bardzo trudno dostępne. Ponadto trudno przewidzieć dokładnie, co i w jakim stopniu zmotywuje konkretną osobę o indywidualnych preferencjach do podjęcia pracy w danym miejscu. W związku z tym zdecydowanie bardziej uzasadnione wydaje się podejście a posteriori, czyli ocena skłonności ludności do podejmowania pracy w danym miejscu na podstawie informacji o miejscu zamieszkania oraz zatrudnienia poszczególnych osób.

Takie podejście stosowane jest przez badaczy w wielu krajach (por. Coombes, 2001; Andersen, 2002; Bamber, Walter, 2009; Brown, Holmes, 1971; Casado-Diaz, 2000; Coombes, Green, Openshaw, 1986; Coombes, Wymer, Charlton, Bailey, Stonehouse, Openshaw, 1997; Fusco, Cagliani, 2011; Kristensen, 1998; Lagnerö, 2003; Newell, Papps, 2001; Nystuen, Dacey, 1968; Papps, Newell, 2002; Sforzi, 1987; Tol-

bert, Sizer, 1996; Van der Haegen, Pattyn, 1980). Wymaga ono jednak informacji o dojazdach do pracy, a w szczególności znajomości macierzy przepływów związanych z zatrudnieniem, czyli informacji o miejscu zamieszkania i zatrudnienia ludności. W związku z powyższym do niedawna przeprowadzenie delimitacji a posteriori lokalnych rynków pracy w Polsce nie było możliwe ze względu na brak odpowiednich danych statystycznych.

W 2010 roku Urząd Statystyczny w Poznaniu opublikował wyniki unikatowego badania przepływów ludności związanych z zatrudnieniem za rok 2006, opartego na danych pozyskanych z zasobów rejestrów podatkowych Ministerstwa Finansów (por. Filas-Przybył, Stachowiak, 2010; Kruszką, 2010). Do szacunków włączono po raz pierwszy informacje zgromadzone w rejestrach podatkowych (POLTAX). Udostępnienie danych przez Ministerstwo Finansów poprzedzone było kilkuletnimi pracami studialnymi w Urzędzie Statystycznym w Poznaniu. Są to pierwsze od 1988 roku dane dotyczące dojazdów do pracy. Ponadto, w roku 2011, wyniki badania dojazdów do pracy zostały uzupełnione poprzez udostępnienie tabeli przepływu pracowników najemnych, którzy dojeżdżają do pracy poza gminę swojego zamieszkania. Udostępnione dane umożliwiają przeprowadzenie pierwszej od wielu lat kompleksowej oraz niezależnej od subiektywnej opinii badacza delimitacji lokalnych rynków pracy w Polsce.

Celem pracy jest podjęcie próby porównania opisanych metod delimitacji lokalnych rynków pracy oraz ocena, które podejście najlepiej sprawdza się w warunkach polskich.

## 2. OPIS DANYCH

Większość stosowanych algorytmów służących do delimitacji lokalnych rynków pracy działa na zasadzie dzielenia skończonej liczby jednostek bazowych na grupy na podstawie informacji o codziennych dojazdach do pracy pomiędzy jednostkami bazowymi. Oczywiście na im więcej jednostek bazowych podzielony jest badany obszar (kraj), tym dokładniejszej delimitacji można dokonać. Najczęściej jedyną wymaganą informacją wejściową jest macierz przepływów związanych z zatrudnieniem, nazywana również macierzą migracji związanych z zatrudnieniem. Jest to macierz kwadratowa o wymiarze równym liczbie jednostek bazowych, zawierająca na przecięciu *i-tego* wiersza oraz *j-tej* kolumny liczbę osób zamieszkałych na terenie *i-tej* jednostki bazowej i jednocześnie zatrudnionych na terenie *j-tej* jednostki bazowej.

Taką informację opublikował na swojej stronie internetowej<sup>1</sup> w 2011 Ośrodek Statystyki Miast działający w Urzędzie Statystycznym w Poznaniu w ramach prezentacji wyników badania przepływów ludności związanych z zatrudnieniem. Dane z tego badania są z dnia 31 grudnia 2006 roku. Badana zbiorowość to dojeżdżający do pracy pracownicy najemni, którzy mieszkają w gminie innej niż ta, gdzie znajduje się ich miejsce pracy, i jednocześnie mają zwiększone koszty uzyskania przychodu z tytułu dojazdów do pracy (2,3 mln osób). Jako jednostkę terytorialną przyjęto w nim gminę

<sup>1</sup> [http://www.stat.gov.pl/poznan/69\\_567.PLK.HTML.htm](http://www.stat.gov.pl/poznan/69_567.PLK.HTML.htm).

z uwzględnieniem podziału na część miejską i wiejską, w wyniku czego na terenie Polski otrzymano 3064 jednostki bazowe (por. *Ogólna charakterystyka badania przepływów ludności związanych z zatrudnieniem w 2006 r.*).

Dane udostępniono w formie tabeli przepływu pracowników najemnych, którzy dojeżdżają do pracy poza gminę swojego zamieszkania, przy czym w celu ochrony danych osobowych uwzględniano wyłącznie przepływy powyżej dziewięciu osób. W ten sposób opublikowana została informacja na temat 25375 niezerowych przepływów (por. Filas-Przybył, Stachowiak, 2010). Dla każdego przepływu w tabeli podane są takie informacje jak kod terytorialny gminy zamieszkania podatnika, kod terytorialny gminy płatnika, nazwa gminy zamieszkania podatnika czy nazwa gminy płatnika.

Tak więc w celu rekonstrukcji macierzy przepływów związanych z zatrudnieniem potrzebnej do delimitacji lokalnych rynków pracy w Polsce utworzono macierz o wymiarze  $3064 \times 3064$ , zawierającą 25375 niezerowych wartości poza diagonalą, zaś w pozostałych jej miejscach z wyjątkiem diagonali przyjęto zera (pomimo świadomości, że mogą to być dowolne liczby nie większe od 9) (por. Kruszką, 2010). Na diagonalach macierzy przepływów powinny znaleźć się liczebności osób jednocześnie mieszkających i pracujących w danej jednostce bazowej. Aby wyliczyć odpowiednie wartości wykorzystano dane z badania pracujący w głównym miejscu pracy pochodzące ze sprawozdawczości statystycznej dostępne dla poziomu gminy z uwzględnieniem podziału na miasto i wieś z roku 2006. W danych tych uwzględnieni nie są pracujący w jednostkach budżetowych działających w zakresie obrony narodowej i bezpieczeństwa publicznego, podmioty gospodarcze o liczbie pracujących do 9 osób, SOF, duchowni pełniący obowiązki duszpasterskie oraz pracujących w gospodarstwach indywidualnych. Ponadto dane te informują o tym, ile osób jest zatrudnionych w danej gminie, wliczając w to również osoby mieszkające poza daną gminą. W związku z tym na diagonalach macierzy przepływów umieszczono różnice pomiędzy liczbą osób pracujących w danej jednostce bazowej i liczbą osób przyjeżdżających tam do pracy. Dane o liczbie osób przyjeżdżających do pracy zaczerpnięto z wyżej omawianego badania Przepływów Ludności Związanych z Zatrudnieniem. Różne źródła danych, a także nieznacznie odmienne definicje osoby pracującej, użytych do obliczenia tej różnicy mogły doprowadzić do niewielkich przekłamań. Odzwierciedleniem tej sytuacji może być na przykład fakt, że dla 25 z 3064 gmin oszacowano nieznacznie ujemne liczebności mieszkańców pracujących na ich terenie. Dla dalszych rozważań wartości te zastąpiono zerami.

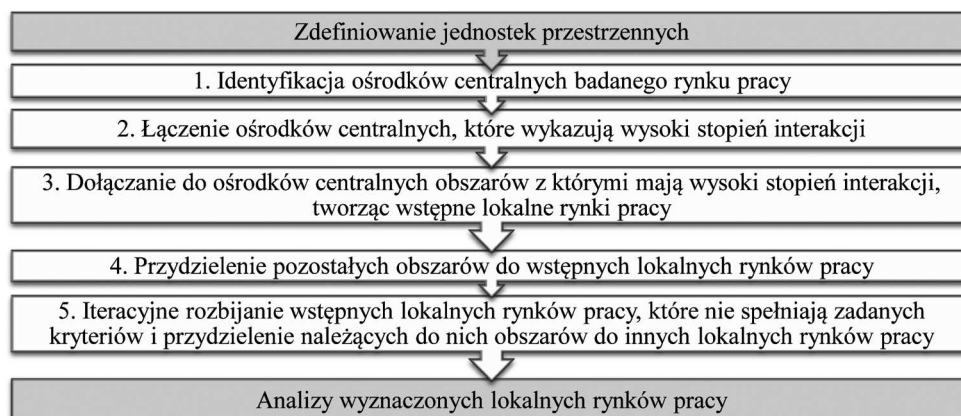
### 3. ALGORYTMY SŁUŻĄCE DO DELIMITACJI LOKALNYCH RYNKÓW PRACY

W celu delimitacji lokalnych rynków pracy w Polsce postanowiono wykorzystać dwa podejścia zaczerpnięte z literatury. Na podstawie analizy obu metod zaproponowano także samodzielnie opracowane podejście taksonomiczne.

## 3.1 EUROPEJSKI ALGORYTM REGIONALIZACJI

Pierwszym podejściem zaczerpniętym z literatury był algorytm opracowany przez Coombes'a, Greena i Openshawa do delimitacji lokalnych rynków pracy w Wielkiej Brytanii (por. Coombes, Green, Openshaw, 1986). Procedura ta znana jest jako Europejski Algorytm Regionalizacji i rekomendowana przez Eurostat jako standardowe podejście do definiowania lokalnych rynków pracy w krajach europejskich (por. Eurostat, 1992). Oprócz Wielkiej Brytanii jest ona stosowana w wielu krajach, między innymi Hiszpanii, Włoszech, Danii, Nowej Zelandii oraz Australii (por. Casado-Díaz, 2000; Sforzi, 1987; Andersen, 2002; Newell, Papps, 2001; Bamber, Walter, 2009).

Algorytm ten został szczegółowo opisany zarówno przez jego autorów (por. Coombes, Green, Openshaw, 1986) jak i przez część jego użytkowników (por. Casado-Díaz, 2000; Newell, Papps, 2001). W niniejszym opracowaniu przedstawiona zostanie idea jego działania.



Rysunek 1. Etapy Europejskiego Algorytmu Regionalizacji

Źródło: Opracowanie własne na podstawie Papps i Newell (2002)

W pierwszej fazie omawianej procedury wytypowane zostają te jednostki bazowe, które pełnią rolę tak zwanych ognisk, czyli ośrodków centralnych wokół których skupiać się będą lokalne rynki pracy. Następnie jednostki bazowe przydzielane są do ognisk, z którymi wykazują odpowiedni stopień interakcji. W trakcie działania algorytmu niektóre z wcześniej wybranych ośrodków centralnych mogą stracić ten status. Również przynależność jednostki bazowej do danego lokalnego rynku pracy może zostać negatywnie zweryfikowana. Wszystkie kryteria, zarówno służące do wyboru ognisk, jak i do przydzielania jednostek bazowych do lokalnych rynków pracy są wyliczane tylko i wyłącznie na podstawie danych zawartych w macierzy przepływów związanych z zatrudnieniem (oznaczanej dalej  $T$ ). Czynności te wykonywane są w pięciu etapach (por. rys. 1), opisanych poniżej.

1) W pierwszym etapie algorytmu, w celu zdefiniowania zbioru tych jednostek bazowych, które mogą zostać wstępnie uznane jako ośrodki centralne, sprawdza się dwa

kryteria. Pierwsze z nich określa stosunek ( $\alpha_{1i}$ ) liczby mieszkańców rozważanej gminy, którzy pracują w dowolnej gminie (w tym również gminie zamieszkania) do ludności pracującej, której miejsce pracy znajduje się na terenie rozważanej gminy (przy czym miejsce zamieszkania jest dowolne), który w przypadku ośrodków centralnych powinien przekraczać zadaną wartość  $\alpha_1$ :

$$\alpha_{1i} = \frac{\sum_{j=1}^N T_{ji}}{\sum_{j=1}^N T_{ij}} > \alpha_1, \quad (1)$$

gdzie:  $T_{ij}$  – liczba ludności zamieszkałej na terenie  $i$ -tego obszaru i pracującej na terenie  $j$ -tego obszaru (element macierzy przepływów związanych z zatrudnieniem),  $N$  – liczba jednostek bazowych.

Drugie kryterium dotyczy udziału mieszkańców rozważanej gminy, którzy podejmują zatrudnienie na jej terenie, wśród wszystkich pracujących mieszkańców (niezależnie od miejsca pracy) ( $\alpha_{2i}$ ), który powinien przekraczać zadaną wartość  $\alpha_2$  aby dana gmina mogła zostać rozważona jako ośrodek centralny:

$$\alpha_{2i} = \frac{T_{ii}}{\sum_{j=1}^N T_{ij}} > \alpha_2. \quad (2)$$

2) W drugim etapie sprawdza się stopień interakcji między wybranymi w etapie pierwszym „wstępnymi” ośrodkami centralnymi. Jeżeli dwa lub więcej z nich są ze sobą zbyt silnie powiązane, w dalszych rozważaniach traktowane są jako jeden ośrodek (status ośrodka centralnego przyznawany jest jednostce najsilniej oddziaływującej, jednak siła oddziaływania jest sumowana). Wszystkie wybrane w etapie pierwszym ośrodki centralne szeregowane są malejąco ze względu na liczbę ludności dojeżdżającej do nich do pracy. Następnie dla każdego ośrodka centralnego wybierany jest mniejszy z udziału ludności pracującej na terenie tego ośrodka wśród wszystkich jego pracujących mieszkańców (niezależnie od miejsca pracy) oraz udziału osób mieszkających na terenie tego ośrodka wśród ludności pracującej na jego terenie (por. wzór (3)). Minimum to ( $\alpha_{3j}$ ) powinno przekraczać zadaną wartość  $\alpha_3$ , co świadczy o odpowiednio dużej liczbie mieszkańców zatrudnionych na terenie badanej jednostki bazowej. Wtedy dana jednostka uznawana jest ostatecznie za ośrodek centralny

$$\alpha_{3j} = \min \left( \frac{T_{jj}}{\sum_{k=1}^N T_{jk}}; \frac{T_{jj}}{\sum_{k=1}^N T_{kj}} \right) > \alpha_3. \quad (3)$$

W przeciwnym razie spośród pozostałych ośrodków centralnych wytypowanych w etapie pierwszym wybierane są te, które spełniają dwa następujące kryteria określające nasilenie interakcji z badanym obszarem. Po pierwsze wybierane są te ośrodki centralne, które mają odpowiedni wysoki udział ( $\alpha_{4ji}$ ) zatrudnionych na terenie ośrodka

(*j-tego*) nie spełniającego poprzedniego (por. wzór (3)) kryterium wśród wszystkich swoich pracujących mieszkańców:

$$\alpha_{4ji} = \frac{T_{ij}}{\sum_{k=1}^N T_{ik}} > \alpha_4. \quad (4)$$

Spośród spełniających warunek (4) ośrodków centralnych typowane są te, do których dojeżdża odpowiedni procent mieszkańców ośrodka *j-tego* (który nie spełnia kryterium (3) jednak spełnia kryterium (4)) wśród wszystkich pracujących mieszkańców *j-tego* ośrodka:

$$\alpha_{5ji} = \frac{T_{ji}}{\sum_{k=1}^N T_{jk}} > \alpha_5. \quad (5)$$

W ten sposób, jeżeli *j-ty* ośrodek centralny nie spełniał kryterium określonego wzorem (3), wytypowano dla niego zbiór ośrodków centralnych o wysokiej z nim interakcji (spełniającej kryteria określone wzorami (4) i (5)). Jeżeli zbiór ten jest pusty, *j-ty* ośrodek centralny zachowuje status samodzielnego ośrodka centralnego. W przeciwnym razie ze zbioru tego wybierany jest ośrodek, który charakteryzuje się najwyższą wartością wskaźnika ważonej interakcji ( $I_{ji}$ ) z ośrodkiem *j-tym*:

$$I_{ji} = \frac{T_{ij}^2}{\sum_{k=1}^N T_{ik} * \sum_{k=1}^N T_{kj}} + \frac{T_{ji}^2}{\sum_{k=1}^N T_{jk} * \sum_{k=1}^N T_{ki}} \geq \alpha_{6j}. \quad (6)$$

Jeżeli dwa (lub więcej) ośrodki mają równe wartości wskaźnika ważonej interakcji z ośrodkiem *j-tym* wybierany jest ten z nich, dla którego suma przepływów bezwzględnych pomiędzy nim a ośrodkiem *j-tym* jest większa. Następnie ośrodek ten przestaje pełnić rolę samodzielnego ośrodka centralnego i przyłączony zostaje do lokalnego rynku pracy, którego rozwój stymulowany jest przez *j-ty* ośrodek centralny. W dalszych rozważaniach oddziaływanie obu tych ośrodków jest sumowane. Następnie można przystąpić do badania kolejnego pod względem liczby ludności dojeżdżającej do niego do pracy ośrodka centralnego, aż do momentu, kiedy zostaną rozważone wszystkie wytypowane w pierwszym etapie ośrodki centralne, z wyjątkiem tych, które wcześniej zostaną przyłączone do innych ośrodków, wyższych w rankingu.

3) W etapie trzecim rozpoczęte zostaje przydzielanie jednostek bazowych do lokalnych rynków pracy, których rozwój stymulowany jest przez ośrodki centralne zdefiniowane w etapie drugim. W etapie tym nie wszystkie jednostki bazowe muszą zostać rozdzielone. Procedura ta rozpoczęta zostaje uszeregowaniem dotychczas utworzonych lokalnych rynków pracy rosnąco ze względu na wartość funkcji celu ( $F_j$ ), która określa ich siłę oddziaływania (por. wzór (7)). Warto podkreślić, że indeks *j* we wzorze (7) dotyczy całego wstępnego lokalnego rynku pracy stymulowanego przez dany ośrodek centralny. Zatem  $T_{jk}$  oznacza wszystkich mieszkańców jednostek bazowych wchodzących w skład tego lokalnego rynku pracy zatrudnionych na terenie *k-tej* jednostki bazowej. Pierwszy składnik iloczynu dotyczy więc łącznej liczby pracujących



mieszkańców badanego lokalnego rynku pracy (bez względu na miejsce zatrudnienia). Drugi ze składników iloczynu zależy od liczby mieszkańców tego wstępnego lokalnego rynku pracy zatrudnionych na jego terenie, która porównywana jest z maksimum z liczby ludności pracującej oraz mieszkającej (ale jednocześnie zatrudnionej w dowolnym obszarze) na jego terenie. Zatem wartość funkcji celu ( $F_j$ ) dla wstępnych lokalnych rynków pracy których rozwój stymulowany jest danym ośrodkiem centralnym będzie zwiększać się w miarę dołączania do niego kolejnych jednostek bazowych:

$$F_j = \min \left( 1; \frac{c_1 * \sum_{k=1}^N T_{jk}}{\alpha_7}; \frac{c_2 * \sum_{k=1}^N T_{jk} + c_3}{\alpha_7} \right) * \min \left( 1; \frac{\alpha_8 * T_{jj}}{\max \left( \sum_{k=1}^N T_{kj}; \sum_{k=1}^N T_{jk} \right)} \right), \quad (7)$$

gdzie:  $\alpha_7$  – parametr ustalany ze względu na wielkość populacji,  $\alpha_8$  – parametr zależny od przeciętnej liczby ludności pracującej w swoim obszarze zamieszkania,  $c_1$ ,  $c_2$ ,  $c_3$  – parametry zależne od nasilenia wyjazdów do pracy z ośrodków centralnych.

Następnie sprawdzany zostaje warunek, czy wartość funkcji celu lokalnego rynku pracy, którego rozwój stymuluje dany ( $j$ -ty) ośrodek centralny (por. wzór (7)) przekracza zadaną wartość  $\alpha_9$ . Powyższe kryterium ma na celu zapobieganie tworzeniu lokalnych rynków pracy, na terenie których mieszka zbyt mało osób pracujących lub na terenie których pracuje zbyt mało jego mieszkańców. Jeżeli funkcja celu przekracza zadaną wartość, dany rynek nie jest dalej rozważany w tym etapie. W przeciwnym wypadku spośród pozostałych jednostek bazowych wybierane są takie, które spełniają poniżej określone kryterium. W tym etapie rozważane są zarówno jednostki bazowe nie będące i będące ośrodkami centralnymi, dla których sprawdza się, czy udział zatrudnionych na terenie  $j$ -tego ośrodka centralnego wśród wszystkich pracujących mieszkańców danej jednostki bazowej (por. wzór (4)) przekracza wartość  $\alpha_4$ . Jeżeli zbiór ten jest pusty, do lokalnego rynku pracy, którego rozwój stymulowany jest przez  $j$ -ty ośrodek centralny nie zostaje dołączona żadna jednostka bazowa. W przeciwnym razie ze zbioru jednostek bazowych spełniających ten warunek wybierana jest ta, która charakteryzuje się najwyższą wartością wskaźnika ważonej interakcji z ośrodkiem  $j$ -tym (por. wzór (6)). Jeżeli dwie (lub więcej) jednostki bazowe mają równe wartości wskaźnika ważonej interakcji z ośrodkiem  $j$ -tym wybierana jest ta z nich, dla której suma przepływów bezwzględnych pomiędzy nią a ośrodkiem  $j$ -tym jest większa. Następnie wybrana jednostka bazowa zostaje przyłączona do lokalnego rynku pracy, którego rozwój stymuluje  $j$ -ty ośrodek centralny. W dalszych rozważaniach suma ich oddziaływania traktowana jako oddziaływanie lokalnego rynku pracy.

Po dokonaniu każdej modyfikacji lokalnego rynku pracy ponownie przeprowadzany jest ranking otrzymanych lokalnych rynków pracy ze względu na funkcję celu (por. wzór (7)). Następnie cała procedura etapu trzeciego zostaje przeprowadzona od nowa. Dzieje się to do momentu, aż do każdego ośrodka centralnego dołączone zostaną tyle jednostek bazowych, aby otrzymany lokalny rynek pracy charakteryzował się odpowiednio wysoką ( $\alpha_9$ ) wartością funkcji celu (por. wzór (7)). Niektóre z otrzy-

many w tym etapie lokalnych rynków pracy mogą składać się wyłącznie z ośrodka centralnego, choć w etapie następnym zostaną rozbudowane.

4) W etapie czwartym wszystkie jednostki bazowe zostaną przydzielone do lokalnych rynków pracy. Wyjątek stanowić mogą jedynie jednostki bazowe o bardzo słabej interakcji z otoczeniem (np. gdyby na terenie takiej jednostki był rezerwat przyrody, gdzie nikt nie mieszka ani nie pracuje). Tak więc wszystkie nie przyporządkowane do tej pory jednostki bazowe szeregowane są malejąco ze względu na liczbę pracujących mieszkańców (niezależnie od miejsca pracy). Następnie, kolejno dla każdej z nich, spośród lokalnych rynków pracy wybierane są te, dla których suma przepływów związanych z zatrudnieniem w obie strony jest niezerowa. Oznacza to, że rozważane są tylko te lokalne rynki pracy, do których dojeżdża choć jeden (w praktyce polskiej 9 ze względu na charakter danych) mieszkaniowiec rozważanej jednostki bazowej lub z którego dojeżdża choć jeden pracownik (w praktyce polskiej 9).

Jeżeli nie istnieje do tej pory lokalny rynek pracy spełniający ten warunek, dana jednostka bazowa dołączona zostaje do listy rezerwowej w celu rozważenia w późniejszym terminie (nie jest to równoznaczne z tym, że pomiędzy nią a otoczeniem nie ma żadnego przepływu ludności związanego z zatrudnieniem, może być tak, że przepływy takie następują pomiędzy nią a innymi jednostkami bazowymi nie przyporządkowanymi do tej pory do żadnego z lokalnych rynków pracy). W przeciwnym razie ze zbioru ośrodków centralnych spełniających ten warunek wybierany jest ten, który charakteryzuje się najwyższą wartością wskaźnika ważonej interakcji (por. wzór (6)) z badaną jednostką bazową. Jeżeli dwa (lub więcej) lokalne rynki pracy mają równe wartości wskaźnika ważonej interakcji z badaną jednostką bazową – wybierany jest ten z nich, dla którego suma przepływów bezwzględnych pomiędzy nim a tą jednostką bazową jest większa. Następnie rozważana jednostka bazowa zostaje przyłączona do tego lokalnego rynku pracy. W dalszych rozważaniach suma ich oddziaływania jest traktowana jako oddziaływanie lokalnego rynku pracy.

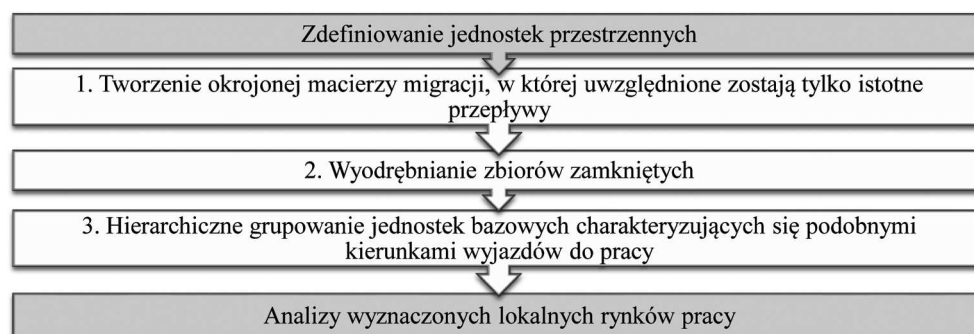
Po próbie przyłączenia wszystkich jednostek bazowych do lokalnych rynków pracy pozostają jednostki z listy rezerwowej, które w momencie próby ich przyłączenia nie wykazywały przepływów ludności związanych z zatrudnieniem z żadnym lokalnym rynkiem pracy. Ponieważ teraz lokalne rynki pracy się powiększyły, ponowiona zostaje próba ich przyporządkowania do któregoś z nich. W tym celu dla każdej jednostki bazowej będącej na liście rezerwowej przeprowadza się ponownie procedurę etapu czwartego. Ponadto próba ta ponawiana jest do momentu, aż nie uda się przyporządkować ani jednej jednostki bazowej pozostającej do danego przebiegu na liście rezerwowej. Jest to gwarancją tego, że wszystkie jednostki bazowe zostały przyporządkowane do lokalnych rynków pracy, jeśli było to tylko możliwe. Jednostki bazowe pozostające w dalszym ciągu na liście rezerwowej albo nie mają żadnych przepływów związanych z zatrudnieniem z innymi obszarami, albo przepływy te następują tylko między nimi samymi.

5) W etapie piątym sprawdzane jest, czy otrzymane lokalne rynki pracy charakteryzują się odpowiednio wysokimi wartościami funkcji celu (por. wzór (7)). W tym celu,

podobnie jak na początku etapu trzeciego, lokalne rynki pracy uszeregowane zostają rosnąco ze względu na tą funkcję. Następnie sprawdza się, czy dla każdego z nich wartość ta przekracza zadaną wartość graniczną  $\alpha_9$ , w tym wypadku większą niż w etapie 3. Jeżeli któryś z lokalnych rynków pracy nie spełnia tego warunku, zaklasyfikowane do niego jednostki bazowe znajdują się na liście rezerwowej i przydzielone zostają do pozostałych lokalnych rynków pracy zgodnie z procedurą z etapu czwartego. Następnie powraca się do etapu piątego, aby zweryfikować czy tak zmodyfikowane lokalne rynki pracy charakteryzują się odpowiednio wysoką wartością funkcji celu. Procedura kończy się w momencie, kiedy podczas etapu piątego nie zostaje rozbitý żaden lokalny rynek pracy.

### 3.2 ALGORYTM KRISTENSEN'A

Drugim podejściem zaczerpniętym z literatury był algorytm zaproponowany przez Kristensen'a do delimitacji lokalnych rynków pracy w Danii (por. Kristensen, 1998). Podejście to jest szczegółowo opisane w pracy jego autora, dlatego w niniejszym opracowaniu przedstawiono jedynie jego ideę.



Rysunek 2. Etapy algorytmu Kristensen'a

Źródło: Opracowanie własne

Algorytm podzielić można na trzy główne etapy (por. rys. 2). W pierwszym z nich następuje transformacja macierzy przepływów związanych z zatrudnieniem na macierz binarną  $TT$  o takim samym wymiarze, w której jedynki przypisywane są tym przepływom, które zostaną uznane za istotne zgodnie z kryterium ilorazu lokacji, charakteryzującym względną siłę natężenia przepływu ludności związanego z zatrudnieniem. Jedynki przypisywane są również wszystkim elementom nowej macierzy należącym do jej diagonalii:

$$TT_{ij} = \begin{cases} 1, & \text{jeżeli } \frac{\frac{T_{ij}}{\sum_{i=1}^N T_{ij}}}{\frac{\sum_{j=1}^N T_{ij}}{\sum_{i=1}^N \sum_{j=1}^N T_{ij}}} > \alpha_1 \quad \forall i = j, \\ 0, & \text{w pozostałych przypadkach,} \end{cases} \quad (8)$$

gdzie:  $TT$  – okrojona macierz przepływów związanych z zatrudnieniem, uwzględniająca wyłącznie istotne przepływy związane z zatrudnieniem,  $T_{ij}$  – liczba ludności zamieszkałej na terenie  $i$ -tej gminy i pracującej na terenie  $j$ -tej gminy (element macierzy przepływów związanych z zatrudnieniem),  $N$  – liczba jednostek bazowych.

W następnej kolejności, na podstawie okrojonej macierzy przepływów związanych z zatrudnieniem podjęta zostaje próba wyodrębnienia zbiorów zamkniętych. Jako zbiór zamknięty rozumiany jest taki podzbiór jednostek bazowych, z których nie istnieje istotny wyjazd ludności związany z zatrudnieniem do jednostki spoza obszaru. Jeżeli zidentyfikowane zostaną zbiory zamknięte uznawane zostają one za lokalne rynki pracy.

Pozostałe jednostki bazowe łączone są w etapie trzecim w lokalne rynki pracy według kryterium funkcji podobieństwa  $P_{ij}$  (por wzór (9)). Przebiega to iteracyjnie w ten sposób, że wyszukiwana jest para jednostek bazowych nie należących jeszcze do żadnych lokalnych rynków pracy charakteryzująca się możliwie największą wartością funkcji podobieństwa ( $P_{ij}$ ). Jeżeli wartość ich funkcji podobieństwa przekracza  $(1 - \alpha_2)$  jednostki te są ze sobą łączone, a następnie dołączane do nich są wszystkie jednostki bazowe nie należące jeszcze do lokalnych rynków pracy charakteryzujące się wartością funkcji podobieństwa przekraczającą  $(1 - \alpha_2)$  dla jednej z połączonych jednostek lub jednostką wcześniej przyłączoną do nich. Wszystkie połączone jednostki tworzą nowy lokalny rynek pracy:

$$P_{ij} = \frac{\overline{W_i \cap W_j}}{\overline{W_i \cup W_j}}, \quad (9)$$

gdzie:  $W_i = \{k; TT_{ik} = 1\}$  – zbiór jednostek bazowych do których następują istotne wyjazdy związane z zatrudnieniem jednostki  $i$ -tej,  $\bar{A}$  – moc zbioru  $A$ .

Jeżeli wśród jednostek bazowych nie należących do lokalnych rynków pracy nie ma żadnej pary charakteryzującej się odpowiednio wysoką wartością funkcji podobieństwa kryterium podobieństwa zostaje obniżone o  $\alpha_2$  i cała procedura rozpoczyna się od nowa. Postępowanie to powtarzane jest do momentu, kiedy wszystkie jednostki bazowe pogrupowane są w lokalne rynki pracy lub nie ma już pary jednostek bazowych nie należących do lokalnych rynków pracy, dla których wartość funkcji podobieństwa przekracza  $\alpha_2$ .

### 3.3 PORÓWNANIE OBU METOD

Obie opisane metody są do siebie podobne pod względem danych potrzebnych aby je zastosować, a także pod względem rezultatu działania: w obu przypadkach jednostki bazowe dzielone są na rozłączne i pokrywające całą zbiorowość lokalne rynki pracy. Podejścia te różnią się jednak między innymi pod względem liczby definiowanych przez badacza parametrów, wyłaniania bądź nie w trakcie postępowania ośrodków centralnych, uwzględnianych rodzajów powiązań schematu tworzenia lokalnych rynków pracy (por tab. 1). Wśród cech obu podejść znaleźć można zarówno wady jak i zalety. Przewagą europejskiego algorytmu regionalizacji jest niewątpliwie uwzględnianie powiązań pomiędzy jednostkami bazowymi a wstępnymi lokalnymi rynkami

pracy oraz dynamiczny charakter budowania lokalnych rynków pracy, pozwalający w każdym momencie na dołączenie do już istniejącego lokalnego rynku pracy kolejnej jednostki bazowej a także na rozbijanie w etapie piątym wcześniej zbudowanych lokalnych rynków pracy. Z kolei zaletą algorytmu Kristensen'a jest mniejsza liczba określanych przez badacza parametrów wstępnych. Również określanie w europejskim algorytmie regionalizacji dla każdego lokalnego rynku pracy dokładnie jednego ośrodka centralnego stymulującego jego rozwój wydaje się być dyskusyjne, stąd brak takiego podejścia można uznać za kolejną przewagę algorytmu Kristensen'a. Poważnym brakiem algorytmu Kristensen'a jest jednak dość kontrowersyjna kolejność budowania lokalnych rynków pracy. Mianowicie na określonym poziomie podobieństwa może nie być możliwości dołączenia kolejnej jednostki bazowej do już istniejącego lokalnego rynku pracy, podczas kiedy jednocześnie na tym samym poziomie podobieństwa istnieje możliwość zbudowania nowego lokalnego rynku pracy.

Tabela 1.

Porównanie europejskiego algorytmu regionalizacji oraz algorytmu Kristensen'a

| Lp | Cecha:                          | Europejski algorytm regionalizacji                                                        | Algorytm Kristensen'a                          |
|----|---------------------------------|-------------------------------------------------------------------------------------------|------------------------------------------------|
| 1  | Dane wejściowe                  | Macierz przepływów związanych z zatrudnieniem                                             |                                                |
| 2  | Liczba definiowanych parametrów | 12                                                                                        | 2                                              |
| 3  | Ośrodki centralne               | Wyłaniane w trakcie działania                                                             | brak                                           |
| 4  | Uwzględniane rodzaje powiązań   | Pomiędzy pojedynczą jednostką bazową a innymi jednostkami oraz grupami jednostek          | Wyłącznie pomiędzy jednostkami bazowymi        |
| 5  | Procedura delimitacji           | Równoległe budowanie lokalnych rynków pracy                                               | Hierarchiczne budowanie lokalnych rynków pracy |
| 6  | Rezultat                        | Podział jednostek bazowych na rozłączne i pokrywające całą zbiorowość lokalne rynki pracy |                                                |

Źródło: Opracowanie własne

### 3.4 PODEJŚCIE TAKSONOMICZNE

Jako alternatywę dla powyżej opisanych podejść postanowiono zaproponować podejście taksonomiczne łączące zalety obu z nich. Podobnie jak w opisanych podejściach jako dane wyjściowe wymagana jest macierz przepływów związanych z zatrudnieniem zaś rezultatem jest podział jednostek bazowych na rozłączne i pokrywające całą zbiorowość lokalne rynki pracy. W opracowanym algorytmie nie ma jednak konieczności określania żadnych parametrów wejściowych a w trakcie postępowania nie są wyodrębniane ośrodki centralne. Ponadto uwzględniane są zarówno przepływy pomiędzy jednostkami bazowymi oraz grupami jednostek bazowych, a procedura tworzenia lokalnych rynków pracy przebiega równoległe, czyli w trakcie trwania algorytmu do każdego z wcześniej wyodrębnionych lokalnych rynków pracy dołączona może zostać kolejna jednostka bazowa lub inny lokalny rynek pracy.

Algorytm opiera się na łączeniu jednostek bazowych, jeżeli występuje między nimi silny przepływ związany z zatrudnieniem. Rozważane są więc pary jednostek bazowych, uszeregowane malejąco według wartości wyjazdów względnych  $WT_{ij}$  z pierwszej z nich do drugiej (por. wzór (10)):

$$WT_{ij} = \frac{T_{ij}}{\sum_{j=1}^N T_{ij}}. \quad (10)$$

Dla każdej kolejno rozważanej pary jednostek bazowych możliwe są cztery sytuacje:

1. Obie jednostki bazowe nie zostały wcześniej włączona do żadnego ze wstępnych lokalnych rynków pracy. Wtedy zostają one ze sobą łączone i w ten sposób powstaje nowy lokalny rynek pracy.
2. Jednostka z której następuje wyjazd nie została jeszcze przyłączona do żadnego z wstępnych lokalnych rynków pracy a jednostka do której wyjazd ten jest skierowany należy już do wstępnego lokalnego rynku pracy. Wtedy jednostka bazowa z której następuje wyjazd jest przyłączana do tego lokalnego rynku pracy.
3. Jednostka wyjazdu należy już do wstępnego lokalnego rynku pracy a jednostka do której wyjazd ten następuje jeszcze nie. W takim wypadku jednostki bazowe nie są ze sobą łączone.
4. Obie jednostki bazowe należą już do wstępnych lokalnych rynków pracy. W takiej sytuacji lokalne rynki pracy są ze sobą łączone jeżeli spełnione są dwa poniżej określone warunki.

Po pierwsze sprawdzane jest, czy istnieją przesłanki, aby lokalny rynek pracy z którego następuje wyjazd przyłączyć do lokalnego rynku pracy, do którego jest on skierowany. W tym celu weryfikuje się, czy liczba mieszkańców lokalnego rynku pracy z którego wyjazd następuje zatrudnionych na terenie lokalnego rynku pracy do którego jest on skierowany przewyższa liczbę mieszkańców lokalnego rynku pracy z którego następuje wyjazd zatrudnionych na jego terenie (por. wzór (11)):

$$\sum_{i=1}^{\bar{A}} \sum_{j=1}^{\bar{B}} T_{A_i B_j} > \sum_{i=1}^{\bar{A}} \sum_{j=1}^{\bar{A}} T_{A_i A_j}, \quad (11)$$

gdzie:  $A$  – zbiór jednostek bazowych zaliczonych do tego wstępnego lokalnego rynku pracy co jednostka z której następuje wyjazd,  $B$  – zbiór jednostek bazowych zaliczonych do tego wstępnego lokalnego rynku pracy co jednostka do której skierowany jest wyjazd,  $A_i$  –  $i$ -ty element zbioru  $A$ .

Następnie sprawdzane jest, czy nie istnieje inny lokalny rynek pracy, do którego przyłączenie lokalnego rynku pracy z którego następuje wyjazd byłoby bardziej optymalne. W tym celu porównuje się liczebności osób wyjeżdżających do pracy z badanego lokalnego rynku pracy do pozostałych lokalnych rynków pracy (por. wzór (12)):

$$\frac{\sum_{i=1}^{\bar{A}} \sum_{j=1}^{\bar{B}} T_{A_i B_j}}{\bar{B}} = \max_c \frac{\sum_{i=1}^{\bar{A}} \sum_{j=1}^{\bar{C}} T_{A_i C_j}}{\bar{C}}, \quad (12)$$

gdzie:  $C$  – zbiór jednostek bazowych zaliczonych do  $c$ -tego wstępnego lokalnego rynku pracy.

Powyżej opisana procedura kontynuowana jest do momentu, aż rozważone zostaną wszystkie pary jednostek bazowych, pomiędzy którymi odnotowano niezerowy przepływ ludności związany z zatrudnieniem.

#### 4. ZASTOSOWANIE OPISANYCH ALGORYTMÓW W WARUNKACH POLSKICH

Delimitację przeprowadzono dla 3064 gmin polskich z podziałem na część miejską i wiejską dla 2006 roku, co zdeterminowane było dostępnością danych.

##### 4.1 EUROPEJSKI ALGORYTM REGIONALIZACJI W WARUNKACH POLSKICH

Europejski algorytm regionalizacji podczas działania porównuje kryteria każdej z jednostek bazowych, wyliczone na podstawie macierzy przepływów związanych z zatrudnieniem, z dziesięcioma wartościami granicznymi. Przy pomocy tych wartości granicznych algorytm można dostosować do warunków danego kraju. Ustala się je w zależności od liczby, wielkości i przeciętnej liczebności ludności zamieszkującej jednostki bazowe oraz łatwością migracji między nimi. W zastosowanym algorytmie postanowiono wykorzystać dla warunków polskich część parametrów z wersji pierwotnej zastosowanej dla Wielkiej Brytanii (por. Coombes, Green, Openshaw, 1986) dostosowując jedynie niektóre z nich (por. tab. 2). Podobne podejście w swoich pracach zastosowali również inni badacze (por. Casado-Diaz, 2000; Newell, Papps, 2001). Stwierdzono, że natężenie migracji między gminami w Polsce było znacznie mniejsze niż między jednostkami bazowymi w Wielkiej Brytanii. Jest to między innymi wynikiem większej powierzchni jednostki bazowej oraz mniejszej łatwości przejazdu. W związku z powyższym konieczne okazało się dostosowanie dwóch wartości granicznych stosowanych przy doborze ośrodków centralnych. Stosunek liczby mieszkańców gminy, którzy pracują w dowolnej gminie (w tym również gminie zamieszkania) do ludności pracującej, której miejsce pracy znajduje się na terenie rozważanej gminy (przy czym miejsce zamieszkania jest dowolne) powinien przekraczać 1,1, aby dana gmina rozważana była jako ośrodek centralny. W pierwotnej wersji algorytmu wartość ta wynosiła aż 1,3, jednak w warunkach polskich kryterium tego nie spełniało prawie żadne z miast (łącznie z Warszawą). Z drugiej strony udział mieszkańców rozważanej gminy, którzy podejmują zatrudnienie na jej terenie, wśród wszystkich pracujących mieszkańców (niezależnie od miejsca pracy) powinien przekraczać 75% aby dana gmina mogła zostać rozważona jako ośrodek centralny. W oryginalnej wersji algorytmu kryterium to zostało ustalone na poziomie 55%. W Polsce jednak mieszkańcy dużych miast poszukują pracy raczej na ich terenie.

Przy zastosowaniu tak dostosowanego algorytmu 3064 gminy (z uwzględnieniem podziału na część miejską i wiejską) podzielono na 222 lokalne rynki pracy. Otrzymano mniej lokalnych rynków pracy niż w Wielkiej Brytanii, jednak więcej niż w Nowej

Tabela 2.

Parametry europejskiego algorytmu regionalizacji w wersji pierwotnej opracowanej dla Wielkiej Brytanii oraz zastosowanej w warunkach polskich

| Wartość     | Parametr   |            |            |            |            |            |            |            |            |               |
|-------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|---------------|
|             | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ | $\alpha_6$ | $\alpha_7$ | $\alpha_8$ | $\alpha_9$ | $\alpha_{10}$ |
| pierwotna   | 1,3        | 0,55       | 0,5        | 0,1        | 0,01       | 0,002      | 2000       | 0,75       | 0,625      | 0,9267        |
| zastosowana | 1,1        | 0,75       | 0,5        | 0,1        | 0,01       | 0,002      | 2000       | 0,75       | 0,625      | 0,9267        |

Źródło: Coombes, Green, Openshaw (1986) oraz opracowanie własne

Tabela 3.

Liczba jednostek bazowych oraz wyznaczonych lokalnych rynków pracy na tle innych krajów

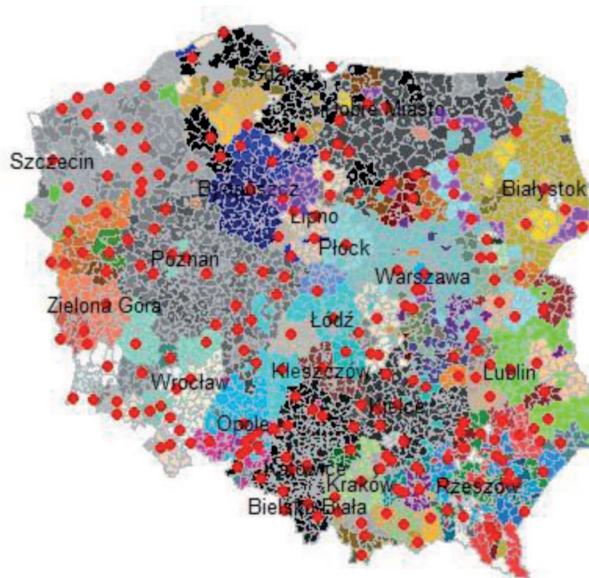
| Charakterystyka wyników delimitacji       | Kraj            |               |        |
|-------------------------------------------|-----------------|---------------|--------|
|                                           | Wielka Brytania | Nowa Zelandia | Polska |
| liczba obszarów bazowych                  | 10102           | 1716          | 3062   |
| liczba otrzymanych lokalnych rynków pracy | 322             | 140           | 222    |

Źródło: Kristensen (1998) oraz opracowanie własne

Zelandii (por. tab. 3). Ich rozkład przestrzenny wraz z ośrodkami centralnymi stymulującymi ich rozwój przedstawiono na wykresie mapowym (por. rys. 3). Największy pod względem liczby wchodzących w jego skład gmin lokalny rynek pracy składał się z 128 gmin. Był to lokalny rynek pracy, którego rozwój stymulowała gmina miasto Warszawa (por. tab. 3). Wielkość tego lokalnego rynku pracy wynika z silnego oddziaływania Warszawy na otaczające ją gminy. Drugim z kolei co do wielkości mierzonej liczbą gmin był lokalny rynek pracy Białegostoku. W tym przypadku nie należy raczej przeceniać oddziaływania ośrodka centralnego, gdyż natężenie dojazdów do pracy w województwie podlaskim, z którym otrzymany lokalny rynek pracy w dużym stopniu się pokrywa, jest najniższe w całej Polsce (por. Kruszka, 2010). Uzasadnienia dużego rozmiaru tego lokalnego rynku pracy można szukać w tym, iż w najbliższym sąsiedztwie Białegostoku nie ma innych miast, które mogłyby pełnić rolę ośrodków centralnych, co z kolei spowodowałoby podział tego rynku. W skład lokalnego rynku pracy, którego rozwój stymuluje Poznań wchodzi o jedną gminę mniej. Jego dużą wielkość, podobnie jak w przypadku Warszawy, wynika z silnego oddziaływania ośrodka centralnego. Należy przy tym uwzględnić, że wielkość lokalnego rynku pracy mierzona liczbą gmin z uwzględnieniem podziału na część miejską i wiejską nie zawsze pokrywa się z jego rzeczywistą wielkością. Jest tak na przykład w przypadku lokalnego rynku pracy skupionego wokół Wrocławia; w jego skład wchodzi tylko 30 gmin, jednak większość z nich charakteryzuje się stosunkowo dużą powierzchnią. Dyskusyjną kwestią może być również definiowanie oraz łączenie w trakcie działania algorytmu ośrodków centralnych. Na przykład w trakcie działania zastosowanej procedury jako ośrodek centralny lokalnego rynku pracy do którego należy Olsztyn uznane zostało położone



w pobliżu Dobrze Miasto. Należy mieć świadomość, że rozwój części otrzymanych lokalnych rynków pracy stymulowany jest przez więcej niż jedno miasto, jednak w takim przypadku tylko jedno z tych miast uzyskuje status ośrodka centralnego.



Rysunek 3. Lokalne rynki pracy wraz z ośrodkami centralnymi wg. algorytmu Coombesa, Polska 2006

Źródło: Opracowanie własne

Średnio lokalny rynek pracy składał się z 14 gmin. 75% lokalnych rynków pracy składało się z co najmniej 6 gmin, w skład połowy z nich wchodziło przynajmniej 9 gmin, zaś tylko 25% z nich składało się z 16 lub więcej jednostek bazowych. Świadczy to o silnej asymetrii prawostronnej: otrzymano wiele mniejszych lokalnych rynków pracy, zaś największe z nich odbiegają silnie od przeciętnej liczby jednostek bazowych wchodzących w ich skład. Wynika to z faktu, że ludność mieszkająca w pobliżu największych ośrodków miejskich jest skłonna dojeżdżać do nich nawet w przypadku dość dużych odległości. Dzieje się tak ponieważ warunki oraz dostępność pracy są tam często znacznie korzystniejsze oraz z tego, że zazwyczaj łatwość dojazdu do nich jest lepsza niż między gminami peryferyjnymi. Z kolei ludność zamieszkująca w odległości wykluczającej codzienny dojazd do największych ośrodków miejskich nie znajduje motywacji do podejmowania wysiłku codziennego dojazdu na dużej odległości do gmin nie będących dużymi miastami. Dwadzieścia największych pod względem liczby gmin lokalnych rynków pracy wymieniono w tabeli 4.

#### 4.2 ALGORYTM KRISTENSEN'A W WARUNKACH POLSKICH

W celu dokonania delimitacji 3064 gmin polskich (z podziałem na część miejską i wiejską) podjęto próby dopasowania obu parametrów algorytmu zaproponowanego

Tabela 4.

Ośrodki centralne stymulujące rozwój 20 największych pod względem liczby jednostek bazowych lokalnych rynków pracy

| Ośrodek centralny | Wielkość lokalnego rynku pracy (liczba gmin) |
|-------------------|----------------------------------------------|
| Warszawa          | 128                                          |
| Białystok         | 96                                           |
| Poznań            | 95                                           |
| Gdańsk            | 68                                           |
| Dobre Miasto      | 65                                           |
| Bydgoszcz         | 62                                           |
| Lublin            | 53                                           |
| Kraków            | 52                                           |
| Katowice          | 47                                           |
| Łódź              | 44                                           |
| Opole             | 44                                           |
| Szczecin          | 43                                           |
| Płock             | 41                                           |
| Rzeszów           | 39                                           |
| Zielona Góra      | 31                                           |
| Kleszczów         | 30                                           |
| Wrocław           | 30                                           |
| Kielce            | 29                                           |
| Bielsko-Biała     | 28                                           |
| Lipno             | 27                                           |

Źródło: Opracowanie własne

przez Kristensen'a. Niestety, przy wielu zastosowanych kombinacjach nie udało się uzyskać mniej niż 868 lokalnych rynków pracy (por. tab. 5), ponadto ponad połowa z nich składała się z dwóch jednostek bazowych (por. rys. 4).

Wyniki te uznano za niezadowolające na podstawie czego stwierdzono, że algorytm ten nie nadaje się do delimitacji lokalnych rynków pracy w Polsce. Jedną z przyczyn takiego rezultatu może być fakt, że liczebność jednostek bazowych w Polsce i w Dani różniła się w znacznym stopniu (por. tab. 6). Ponadto schemat działania algorytmu, zgodnie z którym na określonym poziomie podobieństwa może nie być możliwości dołączenia kolejnej jednostki bazowej do już istniejącego lokalnego rynku pracy, pod-

Tabela 5.

Parametry algorytmu w wersji pierwotnej opracowanej dla Danii oraz zastosowanej w warunkach polskich

| Wartość     | Parametr   |            |
|-------------|------------|------------|
|             | $\alpha_1$ | $\alpha_2$ |
| pierwotna   | 0,75       | 0,05       |
| zastosowana | 1          | 0,2        |

Źródło: Opracowanie własne



Rysunek 4. Lokalne rynki pracy wg. Algorytmu Kristensen'a, Polska 2006

Źródło: Opracowanie własne

czas kiedy jednocześnie na tym samym poziomie podobieństwa istnieje możliwość zbudowania nowego lokalnego rynku pracy wydaje się być podejściem mocno kontrowersyjnym.

Tabela 6.

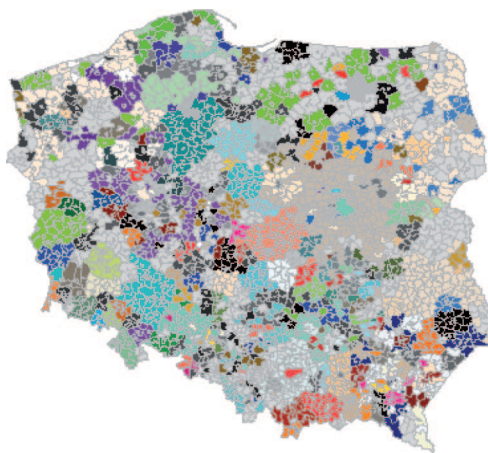
Charakterystyka wyników delimitacji według algorytmu Kristensen'a w Dani i Polsce

| Charakterystyka wyników delimitacji       | Kraj  |        |
|-------------------------------------------|-------|--------|
|                                           | Dania | Polska |
| liczba obszarów bazowych                  | 275   | 3064   |
| liczba otrzymanych lokalnych rynków pracy | 43    | 868    |

Źródło: Opracowanie własne

#### 4.3 WYNIKI PODEJŚCIA TAKSONOMICZNEGO

W trakcie działania algorytmu taksonomicznego powstało 513 wstępnych lokalnych rynków pracy, jednak w wyniku łączenia ich ze sobą, w efekcie końcowym otrzymano 239 lokalnych rynków pracy. Podobnie jak w przypadku europejskiego algorytmu regionalizacji największy lokalny rynek pracy skupiony był wokół Warszawy (por. rys. 5). Istotne różnice odnotowano z kolei na terenie województwa podlaskiego. W wyniku braku oddziaływania w tym województwie silnych ośrodków centralnych (poza Białymstokiem) w wyniku podejścia taksonomicznego otrzymano kilka częściowo przenikających się i nie zawsze spójnych terytorialnie lokalnych rynków pracy. Powoduje to wątpliwość, czy zasadnym jest wyodrębnianie w niektórych wypadkach lokalnych rynków pracy do których nie należą żadne ośrodki centralne. Kolejną problematyczną kwestią wydaje się być spójność terytorialna lokalnych rynków pracy. Należałoby odpowiedzieć bowiem na pytanie, czy posiadanie wspólnej granicy można utożsamiać z łatwością przejazdu pomiędzy dwoma obszarami. W większości przypadków obie kwestie są ze sobą mocno związane, jednak można sobie wyobrazić sytuacje kiedy gminy graniczą ze sobą a łatwość przejazdu między nimi jest mocno ograniczona (np. jeśli płynie między nimi rzeka) oraz kiedy gminy nie mają wspólnej granicy, jednak przejazd pomiędzy nimi nie stanowi zbytniego utrudnienia w codziennych przepływach pracowniczych (np. występuje pomiędzy nimi dogodne połączenie kolejowe).



Rysunek 5. Lokalne rynki pracy, podejście taksonomiczne, Polska

Źródło: Opracowanie własne

#### 4.4 PODSUMOWANIE WYNIKÓW PRZEPROWADZONYCH DELIMITACJI

Jak zauważono wcześniej zastosowanie algorytmu Kristensen'a nie dało w Polsce zadowalających rezultatów. Z kolei wyniki delimitacji przeprowadzonych przy zastosowaniu europejskiego algorytmu regionalizacji oraz podejścia taksonomicznego są do

pod wieloma względami zbliżone. W obu przypadkach otrzymano nieco ponad 200 lokalnych rynków pracy, przy czym 25% z nich składało się z nie więcej niż 6 gmin (por. tab. 7). Również pozostałe kwartale były w obu przypadkach zbliżone (por. tab. 7). Największe rozbieżności odnotowano w przypadku regionów charakteryzujących się niskim natężeniem ośrodków centralnych (np. Podlasie). Wydaje się więc, że wybór metody powinien być zdeterminowany odpowiedzią na pytanie, czy każdy lokalny rynek pracy musi być skupiony wokół ośrodka centralnego.

Tabela 7.

Liczba oraz wielkość (mierzona liczbą gmin wchodzących w ich skład) lokalnych rynków otrzymanych przy zastosowaniu różnych algorytmów, Polska 2006

| Metoda delimitacji                 | Liczba lokalnych rynków pracy | Charakterystyki wielkości lokalnych rynków pracy mierzonej liczbą gmin |       |       |       |            |
|------------------------------------|-------------------------------|------------------------------------------------------------------------|-------|-------|-------|------------|
|                                    |                               | <i>min</i>                                                             | $Q_1$ | $Q_2$ | $Q_3$ | <i>max</i> |
| Europejski algorytm regionalizacji | 222                           | 1                                                                      | 6     | 9     | 16    | 128        |
| Algorytm Kristensen'a              | 868                           | 2                                                                      | 2     | 2     | 3     | 48         |
| Podejście taksonomiczne            | 239                           | 2                                                                      | 6     | 8     | 13    | 209        |

Źródło: Opracowanie własne

## 5. WNIOSKI

Dzięki wykorzystaniu udostępnionej przez Urząd Statystyczny w Poznaniu w 2011 roku, macierzy przepływów związanych z zatrudnieniem możliwe było wyznaczenie lokalnych rynków pracy, na terenie których faktycznie mają miejsce codzienne przepływy ludności związane z zatrudnieniem. Umożliwiło to podjęcie próby przeprowadzenia pierwszej od wielu lat kompleksowej oraz niezależnej od subiektywnej opinii badacza delimitacji lokalnych rynków pracy w Polsce.

Zastosowane zostały dwa zaczerpnięte z literatury algorytmy wykorzystywane w innych krajach oraz samodzielnie opracowane podejście taksonomiczne. Algorytm opracowany przez Kristensen'a do delimitacji lokalnych rynków pracy w Danii uznano za niedostosowany do warunków polskich. Z kolei europejski algorytm regionalizacji został przystosowany do warunków polskich, ponadto lokalne rynki pracy otrzymane przy jego zastosowaniu pod wieloma względami były podobne do tych, które otrzymano w wyniku podejścia taksonomicznego. Dalszym kierunkiem prac będzie zarówno bardziej szczegółowe dostosowanie parametrów w europejskim algorytmie regionalizacji oraz modyfikowanie kryterium łączenia lokalnych rynków pracy w podejściu taksonomicznym w celu obserwacji ich wpływu na otrzymane wyniki.

Przedmiotem dalszych prac będzie również rozważenie, czy spójność terytorialna powinna być warunkiem koniecznym przy tworzeniu lokalnych rynków pracy oraz ewentualne dostosowanie pod tym kątem podejścia taksonomicznego.

*Uniwersytet Ekonomiczny w Poznaniu*

## LITERATURA

- [1] Andersen A.K., (2002), *Are Commuting Areas Relevant for the Delimitation of Administrative Regions in Denmark?*, Regional Studies, 36 (8), 833-844.
- [2] Bamber A., Walter R., (2009), *Implementing the Australian statistical geography standard (ASGS)*. In: Ostendorf B., Baldock P., Bruce D., Burdett M., Corcoran P.(eds.), Proceedings of the Surveying & Spatial Sciences Institute Biennial International Conference, Adelaide 2009, Surveying & Spatial Sciences Institute, pp. 605-616.
- [3] Brown L., Holmes J., (1971), *The delimitation of functional regions, nodal regions, and hierarchies by functional distance approaches*, Journal of Regional Science, 1, 57-72.
- [4] Casado-Diaz J.M., (2000), *Local Labour Market Areas in Spain*, Regional Studies, 34 (9), 843-856.
- [5] Coombes M.G., Green A.E., Openshaw S., (1986), *An Efficient Algorithm to Generate Official Statistical Reporting Areas*, Journal of the Operational Research Society, 37 (10), 943-53.
- [6] Coombes M.G., Wymer C., Charlton M.E., Bailey S., Stonehouse A., Openshaw S., (1997), *Review of Travel-to-Work Areas and Small Area Unemployment Rates*, Labour Market Trends, 105, 9-12.
- [7] *Dojazdy do pracy w Polsce. Terytorialna identyfikacja przepływów ludności związanych z zatrudnieniem* (2010), red. K. Kruska, Główny Urząd Statystyczny i Urząd Statystyczny w Poznaniu, Poznań 2010.
- [8] Eurostat (1992), *Study on Employment Zones*, Eurostat (E/LOC/20), Luxembourg.
- [9] Filas-Przybył S., Stachowiak D., 2010, *Badanie przepływów ludności związanych z zatrudnieniem w: Procesy metropolizacyjne w teorii naukowej i praktyce* konferencja naukowa Łódź, 12-14 października 2009, GUS, Warszawa.
- [10] Fusco G., Cagliani M., (2011), *Hierarchical Clustering through Spatial Interaction Data. The Case of Commuting Flows in South-Eastern France*, Computer Science, 2011, Volume 6782/2011, 135-151.
- [11] Gołata E., (2004), *Estymacja bezpośrednia bezrobocia na lokalnym rynku pracy*, Wydawnictwo Akademii Ekonomicznej w Poznaniu.
- [12] Kristensen K., (1998), *Functional Economic Areas of Denmark: Applying Input-Output Techniques to Commuting*, Bornholms Forskningscenter.
- [13] Lagnerö M., (2003), *Local Labour Markets*. Working Paper. Statistics, Stockholm.
- [14] Newell J., Papps K., (2001), *Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study using Travel-to-Work Data*, New Zealand Department of Labour Occasional Paper Series, 2001/6.
- [15] Nystuen J., Dacey M., (1968), *A Graph Theory Interpretation of Nodal Regions*, Papers and Proceedings of the Regional Science Association, 7, 29-42.
- [16] *Ogólna charakterystyka badania przepływów ludności związanych z zatrudnieniem w 2006 r.*, opracowanie: Ośrodek Statystyki Miast, Urząd Statystyczny w Poznaniu, [http://www.stat.gov.pl/cps/rde/xbcr/poznan/ASSETS\\_Dojazdy\\_do\\_pracy\\_metodologia\\_do\\_internetu.pdf](http://www.stat.gov.pl/cps/rde/xbcr/poznan/ASSETS_Dojazdy_do_pracy_metodologia_do_internetu.pdf).
- [17] Papps K.L., Newell J.O., (2002), *Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data*, IZA Discussion Paper 443.
- [18] Sforzi F., (1987), *L'identificazione spaziale*, in BECATTINI G. (Ed) Mercato e Forze Locali: Il Distretto Industriale. Il Mulino, Bologna.
- [19] Tolbert C., Sizer M., (1996), *US Commuting Zones and Labour Market Areas*. A 1990 Update. US Department of Agriculture, Washington DC.
- [20] Van der Haegen H., Pattyn M., (1980), *An operationalization of the concept of city region in West-European perspective: the Belgian city regions*, Tijdschrift voor Economische en Sociale Geografie, 61, 70-77.

## DELIMITACJA LOKALNYCH RYNKÓW PRACY W POLSCE

## Streszczenie

Stosowane w Europie i na świecie algorytmy służące do delimitacji lokalnych rynków pracy wymagają znajomości danych na temat dojazdów do pracy, w szczególności macierzy migracji związanej z zatrudnieniem. W 2011 roku Urząd Statystyczny w Poznaniu opublikował drugą część wyników unikatowego badania przepływów ludności związanych z zatrudnieniem za rok 2006, opartego na danych pozyskanych z zasobów rejestrów podatkowych Ministerstwa Finansów. Udostępnione dane umożliwiają przeprowadzenie pierwszej od wielu lat kompleksowej oraz niezależnej od subiektywnej opinii badacza delimitacji lokalnych rynków pracy w Polsce. Delimitacja lokalnych rynków pracy dokonana została przy pomocy dwóch metodologii zaczerpniętych z literatury. Pierwszą z nich jest algorytm opracowany dla Wielkiej Brytanii przez M. Coombes'a, A. Greena i S. Openshaw, który rekomendowany jest przez Eurostat jako standardowe podejście do definiowania lokalnych rynków pracy w krajach europejskich. Kolejnym zastosowanym podejściem był algorytm zaproponowany przez Kristensena do delimitacji lokalnych rynków pracy w Danii. W dalszej części opracowania zaproponowano podejście taksonomiczne łączące zalety obu podejść zaczerpniętych z literatury. Po dokonaniu delimitacji lokalnych rynków pracy w Polsce podjęto próbę porównania i oceny otrzymanych wyników.

**Słowa kluczowe:** delimitacja lokalnych rynków pracy, dojazdy do pracy, macierz przepływów związanych z zatrudnieniem

## DELIMITATION OF LOCAL LABOR MARKETS IN POLAND

## Abstract

Used in Europe and worldwide algorithms for the delimitation of local labor markets require knowledge of data on commuter routes, in particular the migration matrix associated with employment. In 2011, the Statistical Office in Poznań published the second part of results of a unique research concerning employment-related population flows for the year 2006, based on data obtained from tax records of the resources of the Ministry of Finance. Available data allow performing a comprehensive and independent researcher on the subjective opinion of delimitation of local labor markets in Poland for first from many years. Delimitation of local labor markets is conducted using two methodologies taken from the literature. The first is an algorithm developed for the UK by M. Coombes, A. Green and S. Openshaw, which is recommended by Eurostat as a standard approach to defining local labor markets in European countries. Another approach was used is an algorithm proposed by Kristensen for delimitation of local labor markets in Denmark. In the following part of the paper was proposed a taxonomic approach, which combines the benefits of both approaches from literature. After the delimitation of local labor markets in Poland attempts to compare and evaluate the results.

**Key words:** delimitation of local labor markets, commuter routes, the matrix of employment-related flows

ALEKSANDRA ŁUCZAK, FELIKS WYSOCKI

## ZASTOSOWANIE UOGÓLNIONEJ MIARY ODLEGŁOŚCI GDM ORAZ METODY TOPSIS DO OCENY POZIOMU ROZWOJU SPOŁECZNO-GOSPODARCZEGO POWIATÓW WOJEWÓDZTWA WIELKOPOLSKIEGO

### 1. WPROWADZENIE

Poziom rozwoju społeczno-gospodarczego<sup>1</sup> można traktować jako strukturę złożoną – wielokryterialną i hierarchiczną. Składa się ona z głównego kryterium oceny (poziom rozwoju społeczno-gospodarczego), kryteriów podrzędnych obejmujących różne aspekty rozwoju np. warunki przyrodnicze, społeczne, poziom rozwoju infrastruktury technicznej i społecznej, stopień rozwoju gospodarczego. W ramach poszczególnych kryteriów przyjmuje się cechy o charakterze metrycznym i porządkowym, opisujących oceniane obiekty (Wysocki 2010).

Celem pracy jest rozpoznanie poziomu rozwoju społeczno-gospodarczego powiatów ziemskich województwa wielkopolskiego. Do jego określenia wykorzystano cechę syntetyczną, która jest funkcją cech prostych – wyznaczników częściowych poziomu rozwoju społeczno-gospodarczego. W procedurze tworzenia cechy syntetycznej zastosowano do mierzenia odległości od wzorca i antywzorca rozwoju uogólnioną miarę odległości GDM (*Generalised Distance Measure*) (Walesiak 1993, 2011), a do obliczenia wartości syntetycznego miernika rozwoju – metodę TOPSIS (*Technique for Order Preference by Similarity to an Ideal Solution*) (Hwang, Yoon 1981, Wysocki 2010). Miara Walesiaka (1993) pozwala na obliczanie odległości pomiędzy obiektami opisywanymi na różnych typach skal tj. porządkowych, przedziałowych i ilorazowych. Natomiast metoda TOPSIS jest metodą wzorcową służącą do obliczania wartości syntetycznego miernika rozwoju na podstawie odległości obiektów od wzorca i antywzorca rozwoju. Zaproponowane podejście pozwala dokonać oceny poziomu rozwoju społeczno-gospodarczego powiatów opisywanych przez kryteria oceny oraz cechy, zarówno metryczne, jak i porządkowe.

Do oceny poziomu rozwoju społeczno-gospodarczego powiatów województwa wielkopolskiego wykorzystano dane statystyczne z ankiety przeprowadzonej w starostwach powiatowych w województwie wielkopolskim nt. *Stanu i możliwości rozwojowych powiatów województwa wielkopolskiego* (2000) oraz Banku Danych Lokalnych Głównego Urzędu Statystycznego za rok 2010 (<http://www.stat.gov.pl/bdl>).

<sup>1</sup> Przez pojęcie rozwoju społeczno-gospodarczego rozumie się korzystne zmiany w potencjale gospodarczym oraz tworzenie i utrzymanie warunków pracy i jakości życia ludności, natomiast „stan rozwoju w jakimś ustalonym okresie nazywa się poziomem rozwoju” (Heffner 2007; Wysocki 2010).



## 2. METODYKA BADAŃ

W procesie tworzenia cechy syntetycznej można wyróżnić siedem etapów postępowania (tab. 1). Pierwszym z nich jest wybór cech opisujących wybrane obiekty (etap I). Istnieją dwa podstawowe podejścia do wyboru cech – statystyczne i merytoryczne (zob. Wysocki 2010).

Tabela 1.

Etapy konstrukcji cechy syntetycznej

| Etapy postępowania                                                     | Opis etapów                                                                                                                                                                                                         |
|------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| I. Wybór cech                                                          | Dobór cech (metrycznych i porządkowych) oraz ich weryfikacja pod względem merytorycznym i/lub statystycznym                                                                                                         |
| II. Podział cech                                                       | Ustalenie kierunku preferencji cech w stosunku do rozpatrywanego kryterium ogólnego                                                                                                                                 |
| III. Przyjęcie systemu wag                                             | Ustalenie wag dla cech na podstawie ich analizy statystycznej lub/i merytorycznej                                                                                                                                   |
| IV. Normalizacja                                                       | Normalizacja cech metrycznych                                                                                                                                                                                       |
| V. Obliczenie odległości obiektu od wzorca i antywzorca rozwoju        | Obliczenie odległości każdego ocenianego obiektu od wzorca i antywzorca rozwoju za pomocą uogólnionej miary odległości GDM                                                                                          |
| VI. Obliczenie wartości syntetycznego miernika rozwoju                 | Obliczenie wartości cechy syntetycznej (syntetycznego miernika rozwoju) za pomocą metody TOPSIS                                                                                                                     |
| VII. Uporządkowanie liniowe obiektów i identyfikacja typów rozwojowych | Wyodrębnienie klas typologicznych dla całego obszaru zmienności cechy syntetycznej metodami statystycznymi (zastosowanie średniej i odchylenia standardowego z wartości cechy syntetycznej) lub w sposób arbitralny |

Źródło: Opracowanie własne.

Następnie należy ustalić kierunek preferencji cech prostych w stosunku do rozpatrywanego kryterium ogólnego dzieląc je na stymulanty, destymulanty i nominanty (etap II). W konstrukcji cechy syntetycznej można również cechom prostym przyporządkować wagi (etap III). Mogą one być ustalone na podstawie analizy statystycznej lub merytorycznej. Pierwsze podejście wykorzystuje procedury statystyczne bazując na informacjach o cechach tkwiących tylko w samej macierzy danych, a w szczególności wykorzystuje analizę zmienności cech oraz analizę korelacji między nimi albo tylko jedną z tych analiz (Wysocki 2010). Drugie podejście opiera się na opiniach ekspertów. W tym zakresie można zaproponować analityczny proces hierarchiczny (AHP) (Saaty 1980; Łuczak, Wysocki 2005). W AHP konstruowany jest hierarchiczny schemat drogą rozkładu rozważanego problemu na elementy składowe tj.: kryterium główne, kryteria podrzędne i cechy. Proces ten rozpoczyna porównanie parami ważności kryteriów podrzędnych w ramach kryterium głównego przy wykorzystaniu skali Saaty'ego (1980) (tab. 2). Wyniki porównań zestawia się w macierz:

$$\mathbf{A} = \begin{bmatrix} 1 & a_{12} & \dots & a_{1j} \\ \frac{1}{a_{12}} & 1 & \dots & a_{2j} \\ \vdots & \vdots & \dots & \vdots \\ \frac{1}{a_{1j}} & \frac{1}{a_{2j}} & \dots & 1 \end{bmatrix}, \quad (1)$$

gdzie:  $j$  – liczba kryteriów ( $j = 1, \dots, n$ ).

Zestawienie w macierz (1) dokonuje się według następujących zasad:

- jeżeli ważność  $l$ -tego kryterium nad  $p$ -tym jest  $a_{lp} = \alpha$ , wtedy  $a_{pl} = \frac{1}{\alpha}$ ,  $\alpha \neq 0$ , (zasada przechodniości ocen),
- jeżeli  $l$ -te kryterium jest równie relatywnie ważne jak  $p$ -te, wtedy  $a_{lp} = a_{pl} = 1$  (zasada równoważności ocen).

Tabela 2.

Dziewięciostopniowa skala preferencji między dwoma porównywanymi elementami według Saaty'ego

| Definicja znaczenia                                         | Objaśnienie                                                                                                                                                                                                                                                                                                                                          | Siła ważności ( $\alpha$ )      |
|-------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|
| Równoważność                                                | Oba czynniki przyczyniają się równo do osiągnięcia celu (jeden czynnik ma takie samo znaczenie jak drugi).                                                                                                                                                                                                                                           | 1                               |
| Słabe lub umiarkowane                                       | Nieprzekonywujące znaczenie lub słaba preferencja jednego czynnika nad drugim (jeden czynnik ma nieco większe znaczenie niż drugi).                                                                                                                                                                                                                  | 3                               |
| Istotne, zasadnicze, mocne                                  | Zasadnicze lub mocne znaczenie lub mocna preferencja jednego czynnika nad innymi (jeden czynnik ma wyraźnie większe znaczenie niż drugi).                                                                                                                                                                                                            | 5                               |
| Zdecydowane lub bardzo mocne                                | Zdecydowane znaczenie lub bardzo mocna preferencja jednego czynnika nad innym (jeden czynnik ma bezwzględnie większe znaczenie niż drugi).                                                                                                                                                                                                           | 7                               |
| Absolutne                                                   | Absolutne znaczenie lub absolutna preferencja jednego czynnika nad innym.                                                                                                                                                                                                                                                                            | 9                               |
| Dla porównań kompromisowych pomiędzy powyższymi wartościami | Czasami istnieje potrzeba interpolacji numerycznej kompromisowych opinii, ponieważ nie ma odpowiedniego słownictwa do ich opisanie, wtedy stosujemy pośrednie wartości między dwoma sąsiednimi ocenami.                                                                                                                                              | 2, 4, 6 i 8                     |
| Przechodniość ocen                                          | Jeżeli $i$ -ty czynnik ma przypisany jeden z powyższych stopni podczas porównania do $j$ -tego czynnika, wtedy $j$ -ty czynnik ma odwrotną wartość, gdy porównuje się do $i$ -tego (jeżeli porównując $X$ z $Y$ przyporządkowujemy wartość $\alpha$ , to wtedy automatycznie musimy przyjąć, że wynikiem porównania $Y$ z $X$ musi być $1/\alpha$ ). | odwrotności powyższych wartości |

Źródło: Opracowanie własne na podstawie Saaty (1980).

Następnie sprawdza się, czy porównania zostały przeprowadzone poprawnie. W tym celu oblicza się wskaźnik niezgodności  $CR$ , który mierzy koherencję porównań parami, czyli określa, w jakim stopniu wzajemne porównania ważności kryteriów są niezgodne:

$$CR = \frac{CI}{RI} \cdot 100\%, \quad (2)$$

we wzorze (2)  $CI = \frac{\lambda_{\max} - n}{n - 1}$  jest indeksem niezgodności, przy czym  $\lambda_{\max}$  jest maksymalną lub główną wartością własną macierzy porównań<sup>2</sup>  $A$ , a  $n$  jest liczbą wierszy (kolumn) w macierzy  $A$ , natomiast  $RI$  jest średnim losowym indeksem niezgodności obliczonym z losowo generowanej macierzy o wymiarach  $n \times n$  (tab. 3).

Tabela 3.

Średni losowy indeks niezgodności  $RI$ 

| Rząd macierzy | $n$  | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
|---------------|------|------|------|------|------|------|------|------|------|------|------|
| Indeks losowy | $RI$ | 0,00 | 0,00 | 0,58 | 0,90 | 1,12 | 1,24 | 1,32 | 1,41 | 1,45 | 1,49 |

Źródło: Hanratty, Joseph (1992).

Wskaźnik niezgodności określa, w jakim stopniu wzajemne porównania ważności są niezgodne (niekonsekwentne). W analizie procesu hierarchicznego oczekuje się, aby wskaźnik  $CR$  przyjmował wartości mniejsze lub równe 10%. Wtedy porównania są konsekwentne. W przeciwnym przypadku porównania (wszystkie lub niektóre) należy powtórzyć, w celu usunięcia niezgodności porównań parami<sup>3</sup>. Jeżeli porównania ważności kryteriów zostały przeprowadzone poprawnie oblicza się wagi ważności kryteriów podrzędnych  $w_j$  ( $j = 1, \dots, n$ ), którymi są znormalizowane wektory własne macierzy  $A$ , czyli priorytety lokalne<sup>4</sup>.

W analogiczny sposób można obliczyć priorytety lokalne dla cech prostych. Priorytety lokalne stanowią podstawę do obliczenia priorytetów globalnych. Ostateczne wagi (priorytety globalne)  $w_k$  dla cech uzyskuje się mnożąc priorytety lokalne dla cech przez priorytety dla kryteriów podrzędnych.

Wartości cech  $x_{ik}^*$  ( $i$  – numeruje obiekty,  $i = 1, 2, \dots, N$ ;  $k$  – numeruje cechy,  $k = 1, 2, \dots, K$ ) zostają przemnożone przez współczynniki wagowe ważności cech uzyskane z AHP (jako priorytety globalne) i zestawione w macierz obserwacji:

$$X = [x_{ik}]_{N \times K}, \quad \text{przy czym} \quad x_{ik} = x_{ik}^* w_k. \quad (3)$$

Wybrane cechy metryczne zazwyczaj mają różne miana, dlatego należy stosować odpowiednie procedury normalizacyjne (etap IV) (Wysocki 2010). W pracy zastosowane zostały przekształcenia ilorazowe dla cech metrycznych:

– dla stymulant

$$z_{ik} = \frac{x_{ik}}{\max_i \{x_{ik}\}}, \quad \max_i \{x_{ik}\} > 0, \quad (4)$$

<sup>2</sup> Wartości własne macierzy  $A$  są pierwiastkami wielomianu charakterystycznego:  $w(\lambda) \equiv \det(A - \lambda I)$ , gdzie  $I$  oznacza macierz jednostkową.

<sup>3</sup> W przypadku pełnej zgodności porównań opinii zachodzi  $\lambda_{\max} = n$ ,  $CI = 0$  i  $CR = 0$ .

<sup>4</sup> Wagi lokalne i globalne dla kryteriów podrzędnych są identyczne.

– dla destymulant

$$z_{ik} = \frac{\min_i \{x_{ik}\}}{x_{ik}}, \quad x_{ik} > 0, \quad (5)$$

– dla nominant

$$z_{ik} = \frac{x_{ik}}{\max_i \{x_{ik}\}}, \quad x_{ik} \leq \max_i \{x_{ik}\}, \quad \max_i \{x_{ik}\} > 0, \quad (6)$$

$$z_{ik} = \frac{\max_i \{x_{ik}\}}{x_{ik}}, \quad x_{ik} > \max_i \{x_{ik}\}, \quad x_{ik} > 0. \quad (7)$$

Natomiast cechy porządkowe nie wymagają normalizacji jednak pod warunkiem, że są mierzone na tej samej skali punktowej.

Następnie ustalone zostają współrzędne obiektów modelowych – wzorca:

$$A^+ = \left( \max_i (x_{i1}), \max_i (x_{i2}), \dots, \max_i (x_{iK}) \right) = (x_1^+, x_2^+, \dots, x_K^+) \quad (8)$$

i antywzorca rozwoju:

$$A^- = \left( \min_i (x_{i1}), \min_i (x_{i2}), \dots, \min_i (x_{iK}) \right) = (x_1^-, x_2^-, \dots, x_K^-). \quad (9)$$

W etapie V oblicza się odległości każdego ocenianego obiektu ( $i = 1, \dots, N$ ) od obiektu – wzorca  $N+1$  (+) i obiektu – antywzorca rozwoju  $N+2$  (–). Jeżeli w zbiorze cech oprócz cech metrycznych występują cechy porządkowe, to do obliczenia tych odległości można zastosować uogólnioną miarę odległości dla cech z różnych skal pomiaru (zob. Walesiak 2011):

$$d_i^{(\bullet)} = d_{ij} = \frac{w_1 d_{ij}^P + w_2 d_{ij}^I + w_3 d_{ij}^R}{w_1 + w_2 + w_3}, \quad i = 1, \dots, N, j = N+1, N+2, \quad (10)$$

gdzie:  $d_{ij}$  – miara odległości GDM  $i$ -tego obiektu od obiektu – wzorca ( $N+1$ ) oraz od obiektu – antywzorca ( $N+2$ ),  $(\bullet)$  oznacza alternatywnie obiekt – wzorzec (+) lub obiekt antywzorzec (–),  $P, I, R$  – podzbiór cech porządkowych, przedziałowych, ilorazowych,  $d_{ij}^P, d_{ij}^I, d_{ij}^R$  – miara odległości dla cech porządkowych, przedziałowych, ilorazowych,  $w_1, w_2, w_3$  – wagi przyporządkowane odległościom wyznaczonym na podstawie cech porządkowych, przedziałowych, ilorazowych, przy czym  $w_1 + w_2 + w_3 = K$  (liczba cech); wagi te mogą również wyrażać merytoryczną ważność poszczególnych cech i ich podzbiorów.

Konstrukcja uogólnionej miary odległości określona jest wzorem (Walesiak 2011):

$$d_{ij}^{(\bullet)} = \frac{1}{2} - \frac{\sum_{k=1}^K a_{ijk} b_{jik} + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq i,j}}^{N+2} a_{ilk} b_{jlk}}{2 \left[ \left( \sum_{k=1}^K a_{ijk}^2 + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq i,j}}^{N+2} a_{ilk}^2 \right) \left( \sum_{k=1}^K b_{jik}^2 + \sum_{k=1}^K \sum_{\substack{l=1 \\ l \neq i,j}}^{N+2} b_{jlk}^2 \right) \right]^{\frac{1}{2}}}, \quad (11)$$

gdzie:  $(\bullet)$  oznacza alternatywnie  $P, I, R$ ;  $i, l = 1, \dots, N, N+1, N+2, j = N+1, N+2$  – numeruje obiekty.

Odległości  $d_{ij}^P$  dla cech mierzonych na skali porządkowej wyznacza się stosując podstawienie:

$$a_{iuk}(b_{jtk}) = \begin{cases} 1 & x_{ik} > x_{uk} (x_{jk} > x_{tk}), \\ 0 & x_{ik} = x_{uk} (x_{jk} = x_{tk}), \\ -1 & x_{ik} < x_{uk} (x_{jk} < x_{tk}), \end{cases} \quad \text{dla } u = j, l; t = i, l, \quad (12)$$

gdzie:  $x_{ik} (x_{jk}, x_{lk}, x_{uk}, x_{tk})$  –  $i$ -ta ( $j$ -ta,  $l$ -ta,  $u$ -ta,  $t$ -ta) obserwacja na  $k$ -tej cesze.

Odległości  $d_{ij}^I$  oraz  $d_{ij}^R$  wyznacza się dla cech mierzonych na skali ilorazowej i/lub przedziałowej podstawiając w powyższej formule:

$$a_{iuk} = x_{ik} - x_{uk} \quad \text{dla } u = j, l, \quad b_{jtk} = x_{jk} - x_{tk} \quad \text{dla } t = i, l, \quad (13)$$

gdzie:  $x_{ik} (x_{jk}, x_{lk})$  –  $i$ -ta ( $j$ -ta,  $l$ -ta) obserwacja na  $k$ -tej cesze.

W kolejnym etapie VI oblicza się wartości syntetycznego miernika poziomu rozwoju (Hwang, Yoon 1981; Wysocki 2010):

$$S_i = \frac{d_i^-}{d_i^+ + d_i^-}, \quad 0 \leq S_i \leq 1, \quad (i = 1, 2, \dots, N). \quad (14)$$

Im mniejsza jest odległość danego obiektu od obiektu modelowego – wzorca rozwoju, a tym samym większa od drugiego bieguna – antywzorca rozwoju, tym wartość miernika syntetycznego jest bliższa 1.

Ostatnim VII etapem jest uporządkowanie liniowe obiektów i wyodrębnienie ich klas typologicznych. Wyodrębnienie klas dla całego obszaru zmienności cechy syntetycznej może zostać przeprowadzone metodami statystycznymi z zastosowaniem średniej i odchylenia standardowego z wartości cechy syntetycznej lub w sposób arbitralny,

przyjmując np. przedziały liczbowe wartości miernika  $S_i$ :  $\langle 0,00; 0,20 \rangle$  – poziom bardzo niski,  $\langle 0,20; 0,40 \rangle$  – poziom niski,  $\langle 0,40; 0,50 \rangle$  – poziom średni-niższy,  $\langle 0,50; 0,60 \rangle$  – poziom średni-wyższy,  $\langle 0,60; 0,80 \rangle$  – poziom wysoki,  $\langle 0,80; 1,00 \rangle$  – poziom bardzo wysoki.

### 3. WYNIKI BADAŃ

W pierwszym etapie utworzono strukturę hierarchiczną oceny poziomu rozwoju społeczno-gospodarczego powiatów, przyjmując następujące kryteria podrzędne: warunki przyrodnicze, warunki społeczne, wyposażenie infrastrukturalne i rozwój gospodarczy. W ramach tych kryteriów dokonano wyboru 18 cech charakteryzujących poziom rozwoju społeczno-gospodarczego w przekroju powiatów ziemskich województwa wielkopolskiego obejmujących *cechy metryczne*, opisujące:

- *warunki społeczne*:
    1. Udział pracujących w rolnictwie, leśnictwie, łowiectwie i rybactwie (%),
    2. Udział pracujących w przemyśle i budownictwie (%),
    3. Stopa bezrobocia (%),
  - *wyposażenie infrastrukturalne*:
    4. Odsetek ludności korzystający z instalacji kanalizacyjnej w % ogółu ludności,
    5. Odsetek ludności korzystający z instalacji gazowej w % ogółu ludności,
    6. Miejsca noclegowe na 1000 ludności,
    7. Uczniowie przypadający na 1 komputer z dostępem do Internetu w gimnazjach dla dzieci i młodzieży (bez szkół specjalnych),
  - *rozwój gospodarczy*
    8. Podmioty gospodarcze 10-49 zatrudnionych na 10 tys. ludności,
    9. Podmioty gospodarcze 50 i więcej zatrudnionych na 10 tys. ludności,
    10. Produkcja sprzedana przemysłu ogółem na 1 mieszkańca w zł,
    11. Nakłady inwestycyjne w przedsiębiorstwach na 1 mieszkańca w zł (z 2008 roku),
    12. Dochody własne gmin w dochodach ogółem w % (średnia z 5 lat),
- i *cechy porządkowe* (wyrażające oceny ekspertów w punktach<sup>5</sup>) opisujące:
- *warunki przyrodnicze*:
    13. Walory środowiska przyrodniczego (las, jeziora, rzeki, parki),
  - *wyposażenie infrastrukturalne*:
    14. Jakość dróg gminnych i powiatowych,
    15. Poziom oczyszczalni ścieków,
    16. Jakość edukacji,

<sup>5</sup> W starostach powiatowych województwa wielkopolskiego została przeprowadzona ankieta nt. *Stanu i możliwości rozwoju powiatów województwa wielkopolskiego*, w której eksperci określali poziomy cech niemetrycznych na skali pięciostopniowej, przyjmując oceny: 5 – bardzo wysoki, 4 – wysoki, 3 – dostateczny, 2, niski, 1- bardzo niski.

– *rozwój gospodarczy:*

17. Poziom kultury rolnej,

18. Poziom rozwoju bazy przetwórczej przemysłu rolno-spożywczego.

W drugim etapie przyjęto, że trzy cechy mają charakter destymulant tj.: udział pracujących w rolnictwie, leśnictwie, łowiectwie i rybactwie<sup>6</sup> (%), stopa bezrobocia (%), uczniowie przypadający na 1 komputer z dostępem do Internetu w gimnazjach dla dzieci i młodzieży (bez szkół specjalnych). Pozostałe cechy są stymulantami.

Stosując metodę Saaty'ego analitycznego procesu hierarchicznego ustalono współczynniki wagowe w odniesieniu do kryteriów (tab. 4) (Wysocki 2010). Najwyższy współczynnik wagowy ma kryterium związane z gospodarką (0,565), a najniższy – ze środowiskiem przyrodniczym (0,055) (tab. 4). Dalej dla uproszczenia przyjęto, że współczynniki wagowe cech prostych w ramach danego kryterium podrzędnego są jednakowe (dzieląc wagę kryterium podrzędnego przez liczbę cech otrzymuje się wagi dla cech prostych) (zob. tab. 4).

Tabela 4.

Ważność kryteriów podrzędnych i cech opisujących poziom rozwoju społeczno-gospodarczego

| Kryterium                      | Przyrodnicze<br>( $w_1$ ) | Społeczne<br>( $w_2$ ) | Infrastrukturalne<br>( $w_3$ ) | Gospodarcze<br>( $w_4$ ) |
|--------------------------------|---------------------------|------------------------|--------------------------------|--------------------------|
| Współczynniki wagowe           | 0,055                     | 0,262                  | 0,118                          | 0,565                    |
| Liczba cech w ramach kryterium | 1                         | 3                      | 7                              | 7                        |
| Wagi cech                      | 0,055                     | 0,087                  | 0,017                          | 0,081                    |

Źródło: Opracowanie własne na podstawie Wysocki (2010).

W kolejnym etapie wartości cech metrycznych poddano normalizacji poprzez przekształcenia ilorazowe (etap IV). Następnie ustalono wzorzec przyjmując wartości maksymalne cech w zbiorze wszystkich powiatów Polski oraz antywzorzec, który stanowiły wartości minimalne cech w tym zbiorze.

W etapie V obliczono odległości od wzorca i antywzorca rozwoju stosując uogólnioną miarę odległości GDM<sup>7</sup>. Stanowiły one podstawę do obliczenia wartości cechy syntetycznej – syntetycznego miernika poziomu rozwoju społeczno-gospodarczego dla powiatów  $S_i$  ( $i = 1, 2, \dots, 31$ ).

Tabela 5 pokazuje wartości syntetycznego miernika rozwoju społeczno-gospodarczego dla powiatów według proponowanej metody (podejście I) oraz uzyskane w tym

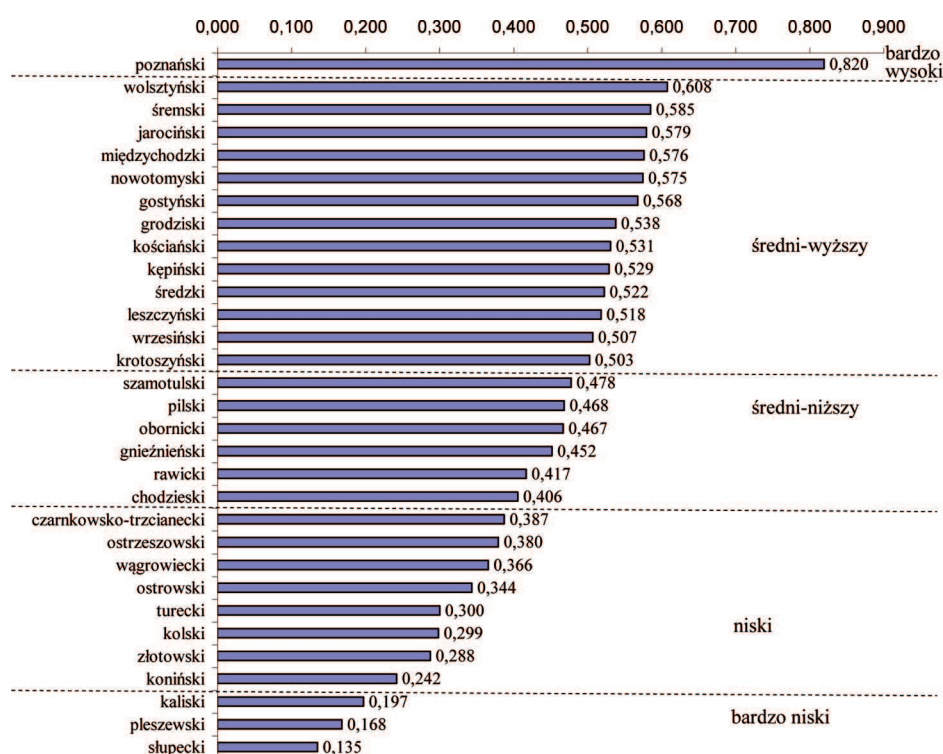
<sup>6</sup> Przyjęto, że ta cecha ma charakter destymulanty, gdyż w rolnictwie polskim, a także w poszczególnych regionach Wielkopolski występuje zjawisko nieefektywnego zatrudnienia w rodzinnych gospodarstwach rolnych (tzw. bezrobocie ukryte) (zob. Kołodziejczak, Wysocki 2012). Z tego względu można przewidywać, że w perspektywie najbliższych lat optymalny (pożądany) poziom zatrudnienia nie zostanie osiągnięty w żadnym regionie Wielkopolski, zaś konieczny będzie odpływ siły roboczej z rolnictwa do innych działów gospodarki.

<sup>7</sup> W obliczeniach wykorzystano pakiet clusterSim programu R.

procesie rangi powiatów. Dla porównania w tabeli 5 zamieszczono również wartości syntetycznego miernika w przypadku podejścia uwzględniającego wagi jednakowe dla kryteriów i cech (podejście II) oraz podejścia III z zastosowaniem metody Hellwiga z uwzględnieniem wag zróżnicowanych. Jednym z powiatów, dla którego wystąpiły znaczące różnice w klasyfikacji rankingowej jest powiat chodzieski. W podejściu I powiat ten uzyskał rangę 20, a w podejściu II uplasował się na 14 pozycji, czyli o sześć miejsc wyżej. Natomiast w procesie porządkowania liniowego metodą Hellwiga powiat ten uplasowałby się na 22 miejscu, czyli o dwa miejsca niżej, niż w podejściu I i aż o osiem miejsc niżej niż w podejściu II.

W metodzie Hellwiga pojawia się problem bliskich zer (powiaty pleszewski – 0,024, kaliski – 0,055) lub nawet ujemnej wartości syntetycznego miernika rozwoju, która wystąpiła w przypadku powiatu słupeckiego (-0,024). Oceny te prowadzą do błędnej identyfikacji poziomu rozwoju tych powiatów, gdyż w rzeczywistości sytuacja społeczno-gospodarcza nie może być aż do tego stopnia niekorzystna.

Rysunek 1 pokazuje uporządkowanie liniowe powiatów według nierosnących wartości rzeczywistych cechy syntetycznej, uzyskanych za pomocą proponowanego podejścia.



Rysunek 1. Uporządkowanie liniowe powiatów województwa wielkopolskiego według poziomu rozwoju społeczno-gospodarczego oraz ich klasyfikacja typologiczna (metoda TOPSIS)

Źródło: Opracowanie własne na podstawie tabeli 5.



Tabela 5.

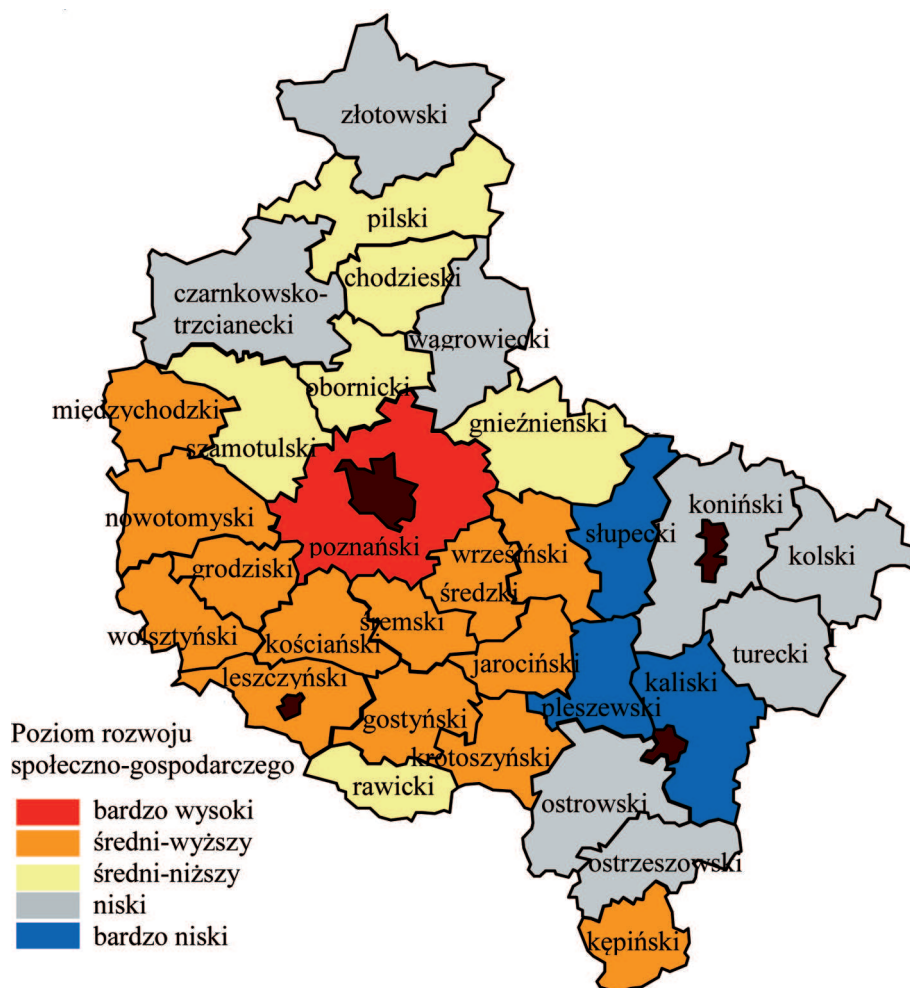
Wartości syntetycznych mierników poziomu rozwoju społeczno-gospodarczego oraz rangi dla powiatów ziemskich województwa wielkopolskiego

| Lp. | Powiaty                 | Podejście |                  |                                        | Rangi  |                  |                                        |
|-----|-------------------------|-----------|------------------|----------------------------------------|--------|------------------|----------------------------------------|
|     |                         | TOPSIS    |                  | metoda Hellwiga <sup>b)</sup><br>(III) | TOPSIS |                  | metoda Hellwiga <sup>b)</sup><br>(III) |
|     |                         | I         | II <sup>a)</sup> |                                        | I      | II <sup>a)</sup> |                                        |
| 1   | chodzieski              | 0,406     | 0,483            | 0,269                                  | 20     | 14               | 22                                     |
| 2   | czarnkowsko-trzcianecki | 0,387     | 0,349            | 0,283                                  | 21     | 22               | 20                                     |
| 3   | gnieźnieński            | 0,452     | 0,494            | 0,367                                  | 18     | 13               | 18                                     |
| 4   | gostyński               | 0,568     | 0,588            | 0,486                                  | 7      | 4                | 7                                      |
| 5   | grodziski               | 0,538     | 0,507            | 0,450                                  | 8      | 12               | 8                                      |
| 6   | jarociński              | 0,579     | 0,577            | 0,496                                  | 4      | 5                | 4                                      |
| 7   | kaliski                 | 0,197     | 0,168            | 0,055                                  | 29     | 29               | 29                                     |
| 8   | kępiński                | 0,529     | 0,521            | 0,405                                  | 10     | 10               | 14                                     |
| 9   | kolski                  | 0,299     | 0,244            | 0,184                                  | 26     | 27               | 25                                     |
| 10  | koniński                | 0,242     | 0,192            | 0,095                                  | 28     | 28               | 28                                     |
| 11  | kościański              | 0,531     | 0,541            | 0,446                                  | 9      | 8                | 9                                      |
| 12  | krotoszyński            | 0,503     | 0,475            | 0,419                                  | 14     | 15               | 12                                     |
| 13  | leszczyński             | 0,518     | 0,468            | 0,424                                  | 12     | 17               | 11                                     |
| 14  | międzychodzki           | 0,576     | 0,550            | 0,489                                  | 5      | 7                | 6                                      |
| 15  | nowotomyski             | 0,575     | 0,561            | 0,492                                  | 6      | 6                | 5                                      |
| 16  | obornicki               | 0,467     | 0,468            | 0,371                                  | 17     | 16               | 17                                     |
| 17  | ostrowski               | 0,344     | 0,371            | 0,240                                  | 24     | 21               | 24                                     |
| 18  | ostrzeszowski           | 0,380     | 0,340            | 0,275                                  | 22     | 24               | 21                                     |
| 19  | pilski                  | 0,468     | 0,512            | 0,400                                  | 16     | 11               | 15                                     |
| 20  | pleszewski              | 0,168     | 0,152            | 0,024                                  | 30     | 30               | 30                                     |
| 21  | poznański               | 0,820     | 0,810            | 0,777                                  | 1      | 1                | 1                                      |
| 22  | rawicki                 | 0,417     | 0,447            | 0,310                                  | 19     | 20               | 19                                     |
| 23  | słupecki                | 0,135     | 0,136            | -0,024                                 | 31     | 31               | 31                                     |
| 24  | szamotulski             | 0,478     | 0,458            | 0,388                                  | 15     | 18               | 16                                     |
| 25  | średzki                 | 0,522     | 0,530            | 0,432                                  | 11     | 9                | 10                                     |
| 26  | śremski                 | 0,585     | 0,599            | 0,508                                  | 3      | 2                | 3                                      |
| 27  | turecki                 | 0,300     | 0,262            | 0,166                                  | 25     | 26               | 26                                     |
| 28  | wągrowiecki             | 0,366     | 0,341            | 0,249                                  | 23     | 23               | 23                                     |
| 29  | wolsztyński             | 0,608     | 0,589            | 0,529                                  | 2      | 3                | 2                                      |
| 30  | wrzesiński              | 0,507     | 0,448            | 0,418                                  | 13     | 19               | 13                                     |
| 31  | złotowski               | 0,288     | 0,289            | 0,157                                  | 27     | 25               | 27                                     |

<sup>a)</sup> Wagi jednakowe. <sup>b)</sup> Proponowane podejście zostało zaproponowane przez Hellwiga (1968).

Źródło: Obliczenia własne na podstawie ankiety przeprowadzonej w starostwach powiatowych w województwie wielkopolskim nt. *Stanu i możliwości rozwojowych powiatów województwa wielkopolskiego* (2000) oraz danych statystycznych Głównego Urzędu Statystycznego z *Banku Danych Lokalnych* (2010).

Delimitację przestrzenną typów poziomu rozwoju społeczno-gospodarczego powiatów przedstawiono na mapie województwa wielkopolskiego (rys. 2).



Rysunek 2. Delimitacja przestrzenna powiatów województwa wielkopolskiego według wartości syntetycznego miernika poziomu rozwoju społeczno-gospodarczego (metoda TOPSIS)

Źródło: Opracowanie własne na podstawie tabeli 5.

Pierwszy typ utworzył powiat poznański, najlepiej rozwinięty pod względem społeczno-gospodarczym. Wartość syntetycznego miernika rozwoju (0,820) była o ponad 0,2 większa niż dla pozostałych powiatów. Oznacza to, że powiat ten w znacznym stopniu wyróżnia się pod względem poziomu rozwoju społeczno-gospodarczego na tle pozostałych powiatów. Wpływ na jego rozwój ma oddziaływanie aglomeracji miejskiej Poznania. Drugi typ utworzyło trzynaście powiatów leżących na południowy-zachód od miasta Poznania. Są to tereny o średnim-wyższym poziomie rozwoju. Kolejny trzeci typ

obejmuje sześć powiatów położonych na północ od Poznania. Powiaty te cechują się średnim-niższym poziomem rozwoju społeczno-gospodarczego. Czwarty typ obejmuje obszar ośmiu powiatów głównie z północnej i wschodniej części województwa. Są to tereny o niskim poziomie rozwoju. Ostatni piąty typ to tereny o bardzo niskim poziomie rozwoju społeczno-gospodarczego. Ten typ utworzyły trzy powiaty (słupecki, kaliski, pleszewski) zdominowane przez gospodarkę rolną, położone we wschodniej części województwa.

#### 4. PODSUMOWANIE

Na podstawie przeprowadzonych obliczeń i analiz można sformułować następujące stwierdzenia i wnioski:

1. Zaproponowane podejście do porządkowania liniowego obiektów wykorzystujące uogólnioną miarę odległości GDM i metodę TOPSIS może być zastosowane do wyznaczenia syntetycznego miernika poziomu rozwoju dla różnych typów danych – metrycznych oraz porządkowych. Wprowadzenie zróżnicowanych wag kryteriów podrzędnych i cech zmieniło uporządkowanie liniowe niektórych powiatów ze względu na poziom rozwoju społeczno-gospodarczego.
2. Dla powiatów pleszewskiego i kaliskiego wartość syntetycznego miernika obliczonego według metody Hellwiga jest bliska zero a w przypadku powiatu słupeckiego osiągnęła nawet wartość poniżej zera, co rodzi problemy związane z identyfikacją poziomu rozwoju społeczno-gospodarczego tych powiatów.
3. Najwyższy poziom rozwoju społeczno-gospodarczego obejmuje obszar powiatu poznańskiego w zasięgu oddziaływania aglomeracji miejskiej Poznania. Natomiast najniższy jest charakterystyczny dla obszarów obejmujących powiaty zdominowane przez gospodarkę rolną, tj. powiatów: słupeckiego, pleszewskiego i kaliskiego.

*Uniwersytet Przyrodniczy w Poznaniu*

#### LITERATURA

- [1] *Bank Danych Lokalnych*, (2010), <http://www.stat.gov.pl/bdl>.
- [2] Hanratty P.J., Joseph B., (1992), *Decision making in chemical engineering and expert systems: application of the analytic hierarchy process to reactor selection*. Comp. Chem. Eng. 16, s. 849-860.
- [3] Heffner K., (2007), *Rozwój społeczno-gospodarczy obszarów wiejskich. Definicje – uwarunkowania – zależności – czynniki – skutki. Badania zróżnicowania rozwoju obszarów wiejskich. W: Zróżnicowanie poziomu rozwoju społeczno-gospodarczego obszarów wiejskich a zróżnicowanie dynamiki przemian*. Wyd. IRWiR, PAN, Warszawa.
- [4] Hellwig Z., (1968), Zastosowania metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny*, z. 4, str. 307-327.
- [5] Hwang C.L., Yoon K., (1981), *Multiple attribute decision-making: Methods and applications*. Springer, Berlin.

- [6] Kołodziejczak W., Wysocki F., (2012), *Identyfikacja charakteru bezrobocia w Polsce według klas miejscowości zamieszkania w latach 2006-2009*. Stowarzyszenie Ekonomistów Rolnictwa i Agrobiznesu. Roczniki Naukowe, tom XIV, z. 4.
- [7] Łuczak A., Wysocki F., (2005), *Wykorzystanie metod taksonometrycznych i analitycznego procesu hierarchicznego do programowania rozwoju obszarów wiejskich*. Wydawnictwo Akademii Rolniczej im. Augusta Cieszkowskiego w Poznaniu, Poznań.
- [8] Saaty T.L., (1980), *The Analytic Hierarchy Process*, McGraw-Hill, New York.
- [9] *Stan i możliwości rozwojowe powiatów województwa wielkopolskiego* (2000): Ankieta przeprowadzona w starostwach powiatowych w województwie wielkopolskim, Uniwersytet Przyrodniczy w Poznaniu.
- [10] Walesiak M., (1993), *Statystyczna analiza wielowymiarowa w badaniach marketingowych*. Prace Naukowe AE we Wrocławiu nr 654. Seria: Monografie i opracowania nr 101, Wrocław.
- [11] Walesiak M., (2011), *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R*. Wyd. Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
- [12] Wysocki F., (2010), *Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich*. Wyd. Uniwersytetu Przyrodniczego w Poznaniu, Poznań.

#### ZASTOSOWANIE UOGÓLNIONEJ MIARY ODLEGŁOŚCI GDM ORAZ METODY TOPSIS DO OCENY POZIOMU ROZWOJU SPOŁECZNO-GOSPODARCZEGO POWIATÓW WOJEWÓDZTWA WIELKOPOLSKIEGO

##### Streszczenie

Celem pracy jest rozpoznanie poziomu rozwoju społeczno-gospodarczego powiatów ziemskich województwa wielkopolskiego. Do jego określenia wykorzystano cechę syntetyczną, która jest funkcją cech prostych – wyznaczników cząstkowych poziomu rozwoju społeczno-gospodarczego. W procedurze tworzenia cechy syntetycznej zastosowano do mierzenia odległości od wzorca i antywzorca rozwoju uogólnioną miarę odległości GDM (Walesiak 1993, 2011), a do agregacji cech metodę TOPSIS (*Technique for Order Preference by Similarity to an Ideal Solution*) (Hwang, Yoon 1981, Wysocki 2010). Miara Walesiaka (Walesiak 1993) pozwala na obliczanie odległości pomiędzy cechami opisywanymi na różnych typach skal tj. porządkowych, przedziałowych i ilorazowych. Natomiast metoda TOPSIS jest metodą wzorcową służącą do agregacji cech. Zaproponowane podejście pozwoliło dokonać oceny poziomu rozwoju społeczno-gospodarczego powiatów opisywanych przez cechy proste, zarówno metryczne, jak i porządkowe.

W procesie oceny poziomu rozwoju społeczno-gospodarczego powiatów województwa wielkopolskiego wykorzystano dane statystyczne z ankiety przeprowadzonej w starostwach powiatowych w województwie wielkopolskim nt. *Stanu i możliwości rozwojowych powiatów województwa wielkopolskiego* (2000) oraz Banku Danych Lokalnych Głównego Urzędu Statystycznego za rok 2010.

**Słowa kluczowe:** poziom rozwoju społeczno-gospodarczego, porządkowanie liniowe, cechy metryczne, cechy porządkowe, uogólniona miara odległości GDM, metoda TOPSIS

---

APPLICATION OF GENERALISED DISTANCE MEASURE AND TOPSIS METHOD TO  
EVALUATE THE LEVEL OF SOCIOECONOMIC DEVELOPMENT OF POWIATS IN  
WIELKOPOLSKA PROVINCE

A b s t r a c t

The paper aims to evaluate the level of socioeconomic development of powiats (i.e. second level administrative units) in Wielkopolska province. With this purpose in mind a synthetic feature was utilized that aggregates simple features: partial indicators of socioeconomic development. The procedure of designing this feature used the generalized distance measure GDM to compute the distance from the ideal and anti-ideal points (Walesiak 1993, 2011), and TOPSIS method to aggregate the features (Hwang, Yoon 1981, Wysocki 2010).

The measure of Walesiak (1993) facilitates calculating distances between features described on different scale types: ordinal, interval, or ratio, while TOPSIS is the standard method for aggregation of features. Proposed approach allows evaluating the level of socioeconomic development of powiats when expressed through simple features, both metric and ordinal.

The evaluation process made use of the data gathered from the survey conducted in the powiat offices of the Wielkopolska province: *“The present state and development potential of the powiats of Wielkopolska”* (2000) and from the Local Data Bank of the Central Statistical Office report of 2010.

**Key words:** level of socioeconomic development, linear ordering, metric features, ordinal features, generalized distance measure (GDM), Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS)

IWONA BĄK

## ZASTOSOWANIE WYBRANYCH METOD STATYSTYCZNO-EKONOMETRYCZNYCH W BADANIU AKTYWNOŚCI TURYSTYCZNEJ SENIORÓW W POLSCE

### 1. WSTĘP

Jednym z najważniejszych problemów społecznym współczesnego świata jest starzenie się ludności. Odnosi się to zwłaszcza do Europy, w tym także i do Polski. Jak wynika z prognoz demograficznych, w ciągu następnych lat proces starzenia się ludności nadal będzie się nasilał. W Polsce w ostatnich 50 latach długość życia wzrosła średnio o 14 lat u kobiet i o 11 lat u mężczyzn. Dwukrotnie zwiększył się też procentowy udział osób w wieku 60 lat i więcej (w 2010 roku wyniósł on 19,7%), a w jeszcze szybszym tempie wzrastała liczba osób powyżej 65. roku życia (*Ludność. Stan... 2010*).

Wzrastająca liczba osób starszych w strukturze populacji to nie tylko wyzwanie dla polityki gospodarczej i społecznej, ale także nowe potężne wyzwanie dla rynku dóbr i usług konsumpcyjnych. Rosnąca w swej masie siła nabywcza uczestników tej grupy wieku oddziaływać będzie coraz silniej na rynek towarów i usług związanych zarówno z życiem codziennym oraz towarów i usług ponadstandardowych, szczególnie w branży motoryzacyjnej, gastronomicznej i turystycznej.

Rosnąca liczba osób starszych może stać się także ważnym czynnikiem wpływającym na wzrost koniunktury szeroko pojętego przemysłu turystycznego. Pozytywny wpływ aktywności turystycznej na zdrowie i samopoczucie osób starszych jest niezaprzeczalny i potwierdzony wieloma wynikami badań (por. np. Napierała, 2002; Chabior, 2005; Fabiś, 2008). Turystyka jest jednym z ważnych czynników zaspakajających podstawowe potrzeby biologiczne, psychiczne i społeczne ludzi starszych. Może ona nie tylko wypełnić czas wolny, lecz także wpływać na wydłużanie się życia ludzkiego. Stanowi ona także ważny element rehabilitacji geriatrycznej, mającej na celu utrzymanie człowieka w podeszłym wieku w możliwie dobrej sprawności fizycznej i psychicznej oraz jak najdłuższe pozostawienie go w dotychczasowym środowisku życiowym i społecznym. Zwiększanie się udziału grupy osób w starszym wieku może również prowadzić do poprawy funkcjonowania rynku usług turystycznych, zarówno krajowych, jak i zagranicznych, a lepsze wyposażenie obiektów (niezbędne dla ludzi starszych i o ograniczonej sprawności) może przyczynić się do poprawy jakości usług. Ze względu na większą swobodę dysponowania czasem wolnym korzystanie przez emerytów z usług biur turystycznych może odbywać się poza sezonem. Stosowane

zachęty cenowe w tym okresie pozwolą na pełniejsze wykorzystanie istniejącej bazy noclegowej.

Polscy seniorzy, na tle seniorów zachodnioeuropejskich czy amerykańskich, a także innych grup wiekowych własnego kraju wykazują stosunkowo niski poziom konsumpcji turystycznej (Śniadek, 2007). Główną przyczyną tej sytuacji jest nie tylko znacznie niższa siła nabywcza współczesnego, polskiego seniora, ale również wyznaczany przez niego system wartości, w którym turystyka oraz inne formy aktywności spędzania czasu wolnego nie zajmują poczesnego miejsca. Biorąc pod uwagę rosnące zasoby czasu wolnego oraz zwiększającą się liczbę osób w tym wieku należy sądzić, że będą one stymulująco oddziaływać na kształtowanie koniunktury na rynku usług turystycznych. W tym celu niezbędne jest podejmowanie badań w sferze konsumpcji turystycznej, które dostarczą istotnych informacji o motywach i zachowaniach konsumpcyjnych turystów-seniorów oraz szacunkowych wielkościach środków, które mogą oni wydatkować na wypoczynek.

Celem artykułu jest badanie prawidłowości w kształtowaniu się zachowań turystycznych seniorów oraz analiza zależności między zmienną określającą wystąpienie wyjazdu turystycznego a zmiennymi ekonomicznymi i społeczno-demograficznymi charakteryzującymi gospodarstwa domowe seniorów w Polsce. Podstawę informacyjną badań zależności będą stanowiły indywidualne dane nieidentyfikowalne dotyczące wyjazdów turystycznych gospodarstw domowych, w których głowa gospodarstwa jest w wieku 60 lat i więcej. Zostały one udostępnione odpłatnie przez Główny Urząd Statystyczny i pochodzą z przeprowadzonego w 2009 roku badania ankietowego nt. „Turystyka i wypoczynek w gospodarstwach domowych. Dane te zostały odpowiednio przetworzone i skojarzone z wynikami badań Budżetów Gospodarstw Domowych przeprowadzonych przez GUS w tym samym roku. W badaniu istotności wpływu czynników determinujących wystąpienie wyjazdu turystycznego wykorzystano test niezależności  $\chi^2$  oraz model logitowy. Omówienie wyników badania zależności zostanie poprzedzone analizą aktywności turystycznej osób w wieku 65 lat i więcej.

## 2. AKTYWNOŚĆ TURYSTYCZNA SENIORÓW

Aktywność turystyczna gospodarstw domowych, w znacznym stopniu determinowana jest przez strukturę według wieku członków tego gospodarstwa. Najwyższą aktywność turystyczną w roku 2005 i 2009 wykazywały dzieci w wieku 7-14 lat oraz osoby w wieku 25-44 lata (tab. 1). Najniższym odsetkiem uczestnictwa w podróżach charakteryzowały się osoby starsze (w wieku 65 lat i więcej) i dotyczyło to wszystkich rodzajów podróży.

Analiza ulubionych sposobów spędzania czasu wolnego przeznaczonego na wypoczynek wykazała, że zdecydowanie największą popularnością wśród członków gospodarstw domowych, cieszyły się następujące cztery sposoby spędzania czasu wolnego (tab. 2): słuchanie radia, muzyki, oglądanie telewizji, wypoczynek bierny, udział w spotkaniach towarzyskich, czytanie. Odsetek słuchających radia, muzyki, oglądających

Tabela 1.

Uczestnictwo w podróżach według grup wieku w latach: 2005 i 2009 (w %)

| Wiek                   | Wyjeżdżające w kraju |                | Wyjeżdżające za granicę<br>(2 dni i dłużej) |
|------------------------|----------------------|----------------|---------------------------------------------|
|                        | 2 – 4 dni            | 5 dni i dłużej |                                             |
| <b>do 6 lat</b>        |                      |                |                                             |
| 2005                   | 22,9                 | 25,3           | 4,1                                         |
| 2009                   | 23,1                 | 28,7           | 5,4                                         |
| <b>7-14 lat</b>        |                      |                |                                             |
| 2005                   | 21,8                 | 36,8           | 5,5                                         |
| 2009                   | 27,4                 | 37,7           | 7,8                                         |
| <b>15-19</b>           |                      |                |                                             |
| 2005                   | 21,1                 | 25,3           | 6,9                                         |
| 2009                   | 21,9                 | 27,4           | 7,0                                         |
| <b>20-24</b>           |                      |                |                                             |
| 2005                   | 21,9                 | 23,9           | 5,5                                         |
| 2009                   | 26,0                 | 20,3           | 9,5                                         |
| <b>25-44</b>           |                      |                |                                             |
| 2005                   | 24,2                 | 24,9           | 7,4                                         |
| 2009                   | 27,2                 | 29,1           | 9,6                                         |
| <b>45-64</b>           |                      |                |                                             |
| 2005                   | 19,2                 | 18,7           | 5,4                                         |
| 2009                   | 23,0                 | 20,8           | 8,2                                         |
| <b>65 lat i więcej</b> |                      |                |                                             |
| 2005                   | 16,0                 | 14,4           | 3,0                                         |
| 2009                   | 15,1                 | 16,5           | 3,1                                         |

Badanie dotyczyło podróży w okresach: 1.10.2004-30.09.2005 oraz 1.10.2008-30.09.2009.

Dla uproszczenia jako okres badania przyjęto rok 2005 i 2009. Dla roku 2001 brak było danych statystycznych.

Źródło: Opracowanie własne na podstawie *Turystyka i wypoczynek ...* (2006, 2010).

telewizję dominował we wszystkich grupach osób ankietowanych, ale najczęściej dotyczył osób w wieku 65 lat i więcej (od 44,6% w roku 2001 do prawie 30% w roku 2009). Drugą popularną formą spędzania czasu wolnego przez osoby starsze było czytanie (od 14-15,2% w badanym okresie), a następnie wypoczynek bierny (spadek z 18% w roku 2001 do 13,4% w roku 2009) i praktyki religijne (wzrost udziału z 8,8% w roku 2001 do 12% w roku 2009). Aktywność ruchowa jest ściśle związana z wiekiem respondentów. Najniższy odsetek uprawiających sport i ćwiczenia fizyczne dotyczył osób w wieku 65 lat i więcej. Najchętniej seniorzy brali udział w spotkaniach towarzyskich (wzrost z 2,2% w roku 2001 do 8,4% w roku 2009) albo pracowali na działce. Stosunkowo rzadko uprawiali sporty i ćwiczenia fizyczne oraz odwiedzali muzea, wystawy, zabytki i brali udział w wycieczkach i imprezach objazdowych.

Z danych statystycznych o kształtowaniu się struktury podróży turystycznych seniorów wynika, że krajowe podróże długoterminowe cechowały się dużym natężeniem sezonowości, przy czym większość z nich przypadała na okres letni (tab. 3). Według badań GUS, ponad 55% uczestników takich podróży realizowało je właśnie w lecie. Podróże w pozostałych porach roku cieszyły się znacznie mniejszą, choć podobną względem siebie, bo kilkuprocentową popularnością. Wyjątek stanowił jedynie miesiąc wrzesień, w którym odbyto w roku 2009 około 15% podróży (w okresach



Tabela 2.

Ulubiony sposób spędzania czasu wolnego przeznaczonego na wypoczynek według wieku członków gospodarstw domowych w latach: 2001, 2005 i 2009 (w %)

| Wyszczególnienie                                                           | Ogółem | 7-14 | 15-19 | 20-24 | 25-44 | 45-64 | 65 i więcej |
|----------------------------------------------------------------------------|--------|------|-------|-------|-------|-------|-------------|
| <b>Wypoczynek bierny – opalanie się, relaks</b>                            | 19,0   | 14,1 | 14,3  | 16,4  | 23,7  | 19,9  | 18,0        |
| 2001                                                                       | 13,6   | 11,4 | 11,2  | 13,0  | 15,4  | 13,9  | 13,1        |
| 2005                                                                       | 14,6   | 12,7 | 11,4  | 12,2  | 16,9  | 15,4  | 13,4        |
| 2009                                                                       |        |      |       |       |       |       |             |
| <b>Czytanie</b>                                                            | 12,9   | 11,7 | 11,2  | 14,4  | 12,9  | 13,3  | 14,0        |
| 2001                                                                       | 12,3   | 10,7 | 9,9   | 9,2   | 11,8  | 13,9  | 15,5        |
| 2005                                                                       | 11,3   | 9,7  | 8,3   | 8,0   | 10,0  | 13,3  | 15,2        |
| 2009                                                                       |        |      |       |       |       |       |             |
| <b>Słuchanie radia, muzyki, oglądanie telewizji, video</b>                 | 37,7   | 41,1 | 39,7  | 31,5  | 35,0  | 37,2  | 44,6        |
| 2001                                                                       | 27,0   | 29,8 | 26,7  | 25,3  | 25,0  | 26,9  | 30,2        |
| 2005                                                                       | 25,8   | 29,0 | 24,5  | 23,8  | 23,9  | 25,8  | 29,6        |
| 2009                                                                       |        |      |       |       |       |       |             |
| <b>Udział w spotkaniach towarzyskich</b>                                   | 6,4    | 4,2  | 10,4  | 14,5  | 6,7   | 4,8   | 2,2         |
| 2001                                                                       | 12,6   | 9,9  | 16,6  | 19,2  | 14,1  | 10,7  | 8,2         |
| 2005                                                                       | 12,7   | 10,0 | 17,2  | 19,6  | 14,4  | 10,8  | 8,4         |
| 2009                                                                       |        |      |       |       |       |       |             |
| <b>Chodzenie do kawiarni, restauracji</b>                                  | 0,5    | -    | 1,0   | 2,0   | 0,5   | 0,2   | 0,1         |
| 2001                                                                       | 1,6    | 0,6  | 3,2   | 4,5   | 1,9   | 0,9   | 0,3         |
| 2005                                                                       | 1,7    | 0,6  | 2,8   | 4,0   | 2,4   | 1,1   | 0,5         |
| 2009                                                                       |        |      |       |       |       |       |             |
| <b>Uprawianie sportu i ćwiczeń fizycznych</b>                              | 8,2    | 18,9 | 13,1  | 10,0  | 6,1   | 5,0   | 2,2         |
| 2001                                                                       | 7,4    | 16,2 | 11,7  | 8,1   | 6,2   | 5,1   | 3,6         |
| 2005                                                                       | 7,4    | 16,6 | 11,8  | 8,7   | 6,9   | 4,7   | 2,8         |
| 2009                                                                       |        |      |       |       |       |       |             |
| <b>Kibicowanie na imprezach sportowych</b>                                 | 2,2    | 2,6  | 3,1   | 3,3   | 2,3   | 1,9   | 0,4         |
| 2001                                                                       | 3,9    | 4,7  | 6,2   | 4,7   | 4,3   | 3,3   | 1,2         |
| 2005                                                                       | 4,1    | 4,9  | 5,8   | 4,9   | 4,9   | 3,5   | 1,0         |
| 2009                                                                       |        |      |       |       |       |       |             |
| <b>Odwiedzanie muzeów, wystaw, zabytków, imprezy objazdowe, zwiedzanie</b> | 1,3    | 0,8  | 0,9   | 1,5   | 1,7   | 1,4   | 0,7         |
| 2001                                                                       | 2,2    | 2,0  | 1,9   | 2,1   | 2,6   | 2,5   | 1,7         |
| 2005                                                                       | 2,1    | 1,4  | 1,2   | 1,7   | 2,6   | 2,3   | 1,8         |
| 2009                                                                       |        |      |       |       |       |       |             |
| <b>Chodzenie do kina, teatru, udział w imprezach kulturalnych</b>          | 0,9    | 0,8  | 2,1   | 1,8   | 0,9   | 0,4   | 0,2         |
| 2001                                                                       | 3,1    | 4,1  | 4,6   | 5,4   | 3,7   | 1,9   | 1,2         |
| 2005                                                                       | 3,4    | 5,0  | 6,1   | 6,1   | 3,8   | 1,8   | 1,4         |
| 2009                                                                       |        |      |       |       |       |       |             |
| <b>Praktyki religijne</b>                                                  | 2,1    | 0,6  | 0,2   | 0,4   | 1,0   | 2,6   | 8,8         |
| 2001                                                                       | 3,4    | 1,6  | 0,7   | 1,0   | 1,6   | 3,8   | 11,9        |
| 2005                                                                       | 3,6    | 1,5  | 1,1   | 1,2   | 1,6   | 4,2   | 12,0        |
| 2009                                                                       |        |      |       |       |       |       |             |

cd. Tabela 2.

|                                             |     |     |     |     |     |      |     |
|---------------------------------------------|-----|-----|-----|-----|-----|------|-----|
| <b>Uprawianie hobby, pogłębianie wiedzy</b> | 2,7 | 3,6 | 2,7 | 3,1 | 3,0 | 2,4  | 1,4 |
| 2001                                        | 3,4 | 4,4 | 3,6 | 3,5 | 3,4 | 3,4  | 2,0 |
| 2005                                        | 3,7 | 4,6 | 4,5 | 4,0 | 3,7 | 3,5  | 2,5 |
| 2009                                        |     |     |     |     |     |      |     |
| <b>Praca na działce</b>                     | 5,0 | 0,3 | 0,4 | 0,5 | 4,7 | 10,1 | 6,6 |
| 2001                                        | 5,7 | 0,5 | 0,6 | 1,1 | 5,4 | 9,9  | 8,7 |
| 2005                                        | 5,5 | 0,4 | 0,6 | 0,9 | 4,3 | 10,2 | 8,2 |
| 2009                                        |     |     |     |     |     |      |     |
| <b>Zakupy dla przyjemności</b>              | 0,6 | 0,3 | 0,5 | 0,5 | 0,9 | 0,7  | 0,1 |
| 2001                                        | 2,2 | 2,6 | 1,1 | 0,7 | 1,4 | 1,5  | 1,4 |
| 2005                                        | 2,6 | 1,6 | 2,8 | 3,4 | 3,4 | 2,4  | 1,4 |
| 2009                                        |     |     |     |     |     |      |     |
| <b>Inne</b>                                 | 0,5 | 1,0 | 0,6 | 0,3 | 0,6 | 0,2  | 0,6 |
| 2001                                        | 1,5 | 2,6 | 1,1 | 0,7 | 1,4 | 1,5  | 1,4 |
| 2005                                        | 1,4 | 1,9 | 2,0 | 1,4 | 1,0 | 1,2  | 1,9 |
| 2009                                        |     |     |     |     |     |      |     |

Badanie dotyczyło wyjazdów w okresie 1.10.2000-30.09.2001. Dla uproszczenia jako okres badania przyjęto rok 2001.

Źródło: Opracowanie własne na podstawie *Turystyka i wypoczynek* ... (2002, 2006, 2010).

wcześniejszych ok. 11%). Wskaźnik uczestnictwa seniorów w podróżach trwających 2-4 dni w celach typowo turystycznych w okresie letnim jest niższy niż wskaźnik uczestnictwa w podróżach krajowych długoterminowych w tym samym okresie. Latem w podróżach turystycznych brało udział około 44,0% uczestników. Natomiast więcej seniorów wyjeżdżało w miesiącach wiosennych, szczególnie w maju oraz pod koniec roku. Sezonowy rozkład podróży zagranicznych przedstawia się nieco inaczej niż podróży krajowych. Nasilenie podróży zaczyna się już w maju i trwa aż do września. W miesiącach typowo zimowych (od grudnia do lutego) osoby starsze sporadycznie wyjeżdżają w celach turystycznych za granicę, a w roku 2009 w ogóle nie uczestniczyli w takich podróżach.

Tabela 3.

Struktura podróży turystycznych seniorów według miesięcy w latach: 2001, 2005 i 2009 (w %)

| Miesiące | Podróże krajowe długoterminowe |      |      | Podróże krajowe krótkoterminowe |      |      | Podróże zagraniczne |      |      |
|----------|--------------------------------|------|------|---------------------------------|------|------|---------------------|------|------|
|          | 2001                           | 2005 | 2009 | 2001                            | 2005 | 2009 | 2001                | 2005 | 2009 |
| styczeń  | 3,2                            | 3,3  | 1,8  | 1,7                             | 2,0  | 2,0  | 2,2                 | 3,4  | 0,0  |
| luty     | 3,5                            | 2,2  | 1,2  | 4,0                             | 2,4  | 2,4  | 2,1                 | 0,0  | 0,0  |
| marzec   | 2,1                            | 4,1  | 1,3  | 4,0                             | 4,1  | 4,9  | 0,6                 | 5,2  | 5,0  |
| kwiecień | 6,9                            | 6,1  | 6,6  | 10,8                            | 10,6 | 8,4  | 3,0                 | 8,4  | 3,6  |
| maj      | 6,6                            | 9,7  | 8,2  | 17,3                            | 10,0 | 13,7 | 22,9                | 9,1  | 11,3 |
| czerwiec | 9,6                            | 11,1 | 12,1 | 10,4                            | 12,6 | 11,2 | 15,1                | 22,5 | 18,2 |
| lipiec   | 21,5                           | 19,8 | 23,1 | 16,6                            | 16,2 | 16,5 | 15,5                | 20,4 | 13,2 |
| sierpień | 26,3                           | 24,3 | 20,6 | 14,0                            | 15,5 | 17,1 | 21,5                | 15,3 | 19,7 |

cd. Tabela 3.

|             |      |     |      |     |      |     |      |      |      |
|-------------|------|-----|------|-----|------|-----|------|------|------|
| wrzesień    | 10,4 | 11  | 14,9 | 9,0 | 12,9 | 6,8 | 10,9 | 10,1 | 21,6 |
| październik | 2,4  | 3,7 | 4,1  | 3,9 | 2,9  | 6,4 | 1,0  | 1,7  | 5,0  |
| listopad    | 3,3  | 2,8 | 1,1  | 3,6 | 3,6  | 3,5 | 1,0  | 1,0  | 2,5  |
| grudzień    | 4,2  | 1,9 | 4,9  | 4,5 | 7,3  | 7,2 | 6,7  | 3,8  | 0,0  |

Źródło: Opracowanie własne na podstawie *Turystyka i wypoczynek ...* (2002, 2006, 2010)

Tabela 4.

Podróże w gospodarstwach domowych seniorów według głównego celu w latach:  
2001, 2005 i 2009 (w %)

| Cel podróży                                     | Podróże krajowe<br>długoterminowe | Podróże krajowe<br>krótkoterminowe | Podróże<br>zagraniczne |
|-------------------------------------------------|-----------------------------------|------------------------------------|------------------------|
| wypoczynek, rekreacja, wakacje                  |                                   |                                    |                        |
| 2001                                            | 27,2                              | 14,6                               | 21,5                   |
| 2005                                            | 29,2                              | 7,4                                | 8,8                    |
| 2009                                            | 40,4                              | 22,0                               | 9,0                    |
| zwiedzanie (architektura, kultura,<br>przyroda) |                                   |                                    |                        |
| 2001                                            | 1,0                               | 2,6                                | 17,6                   |
| 2005                                            | 1,1                               | 3,0                                | 21,3                   |
| 2009                                            | 2,6                               | 3,6                                | 30,8                   |
| odwiedziny u krewnych lub znajomych             |                                   |                                    |                        |
| 2001                                            | 47,2                              | 66,3                               | 40,5                   |
| 2005                                            | 50,7                              | 66,6                               | 57,9                   |
| 2009                                            | 37,0                              | 58,6                               | 49,3                   |
| uroczystości rodzinne                           |                                   |                                    |                        |
| 2001                                            | 1,0                               | 6,6                                | 9,7                    |
| 2005                                            | 4,5                               | 6,9                                | 3,4                    |
| 2009                                            | 2,4                               | 5,9                                | 6,5                    |
| zdrowotny                                       |                                   |                                    |                        |
| 2001                                            | 17,1                              | –                                  | –                      |
| 2005                                            | 14,1                              | 0,6                                | –                      |
| 2009                                            | 15,7                              | 1,1                                | –                      |
| religijny                                       |                                   |                                    |                        |
| 2001                                            | 1,0                               | 5,6                                | 4,6                    |
| 2005                                            | 0,3                               | 7,2                                | 3,3                    |
| 2009                                            | 0,6                               | 3,7                                | 4,3                    |
| inny                                            |                                   |                                    |                        |
| 2001                                            | 5,5                               | 4,3                                | 6,1                    |
| 2005                                            | 0,1                               | 8,3                                | 5,3                    |
| 2009                                            | 1,3                               | 5,1                                | 0,1                    |

Źródło: Opracowanie własne na podstawie *Turystyka i wypoczynek ...* (2002, 2006, 2010)

Głównym celem podróży członków gospodarstw domowych poza miejsce zamieszkania były odwiedziny krewnych lub znajomych (tab. 4). Na drugim miejscu wskazywano wypoczynek, rekreację i wakacje. Jednak w 2009 roku podróże, których celem jest wypoczynek, rekreacja i wakacje były najczęściej wskazywane jako główny cel wyjazdów wakacyjnych. Nie bez znaczenia były również podróże wakacyjne w celu

poratowania zdrowia. Cele podróży zagranicznych są nieco odmienne niż wyjazdów krajowych. Co prawda największy udział miały podróże związane z odwiedzinami krewnych i znajomych, ale uwagę zwraca wysoki udział podróży, w czasie których turyści zwiedzali architekturę, kulturę i przyrodę. Istotny poziom wskazań dotyczył też wypoczynku, rekreacji, wakacji oraz uroczystości rodzinnych. Formy podróży były dosyć zróżnicowane, ale najwięcej turystów wybrało wycieczkę i imprezę objazdową, a następnie wczasy.

Z odpowiedzi respondentów wynika, że przy organizowaniu podróży rzadko korzystali oni z usług pośrednicko-organizatorskich, szczególnie w przypadku podróży krajowych (tab. 5). W 2009 roku z takich usług skorzystano w odniesieniu do niespełna 23% podróży wakacyjnych (w roku 2001 31,2%). W 2009 roku dominował zakup pełnych pakietów usług (12,4%), a w roku 2001 i 2005 – zakup noclegów. Krótkie podróże organizowane są na ogół całkowicie samodzielnie<sup>1</sup>. Tylko 7,5% gospodarstw domowych seniorów w 2009 roku wykupiło jakiekolwiek usługi u pośredników. Cztery lata wcześniej takiego zakupu dokonało nieco ponad 11% podróżujących. Najczęściej kupowano: pełny pakiet usług, noclegi lub transport. Większość wyjeżdżających za granicę przebywała na wypoczynku zorganizowanym indywidualnie (za wyjątkiem roku 2001), choć odsetek osób, które zakupiły usługi był zdecydowanie wyższy niż przy wyjazdach krajowych. Jest to zgodne z prawidłowościami obserwowanymi praktycznie we wszystkich krajach, gdzie podróże zagraniczne charakteryzują się znacznie wyższym poziomem zorganizowania niż podróże krajowe. Według GUS pełne pakiety turystyczne zakupiła co trzecia osoba wyjeżdżająca za granicę (w 2005 roku – co czwarta). Natomiast wśród pojedynczych usług najczęściej kupowano noclegi i transport.

Tabela 5.

Podróże zrealizowane w gospodarstwach domowych seniorów według rodzajów zakupionych usług w latach: 2001, 2005 i 2009 (w %)

| Rodzaje zakupionych usług | Podróże krajowe długoterminowe | Podróże krajowe krótkoterminowe | Podróże zagraniczne |
|---------------------------|--------------------------------|---------------------------------|---------------------|
| ogółem zakupione usługi   | 31,2                           | 6,8                             | 51,2                |
| 2001                      | 19,6                           | 11,3                            | 47,6                |
| 2005                      | 22,9                           | 7,5                             | 42,7                |
| 2009                      |                                |                                 |                     |
| odsetek zakupionych usług |                                |                                 |                     |
| pełny pakiet              | 11,3                           | 1,3                             | 35,0                |
| 2001                      | 7,3                            | 0,8                             | 24,7                |
| 2005                      | 12,4                           | 3,7                             | 31,0                |
| 2009                      |                                |                                 |                     |
| nocleg                    | 18,4                           | 3,1                             | -                   |
| 2001                      | 11,4                           | 6,7                             | 10,5                |
| 2005                      | 8,2                            | 3,5                             | 7,6                 |
| 2009                      |                                |                                 |                     |

<sup>1</sup> Podobne wnioski sformułował w pracy Łaciak (2007).

cd. Tabela 5.

|                             |      |     |      |
|-----------------------------|------|-----|------|
| transport                   | 1,0  | 4,2 | 9,9  |
| 2001                        | 0,7  | 3,1 | 12,4 |
| 2005                        | 0,7  | 1,3 | 6,3  |
| 2009                        |      |     |      |
| pełne wyżywienie            | 10,2 | 0,8 | –    |
| 2001                        | 7,7  | 0,5 | –    |
| 2005                        | 6,7  | 0,4 | 5,4  |
| 2009                        |      |     |      |
| częściowe wyżywienie        | 4,0  | 0,5 | 6,3  |
| 2001                        | 0,9  | 1,0 | –    |
| 2005                        | 1,5  | 0,7 | 2,2  |
| 2009                        |      |     |      |
| lokalne wycieczki i imprezy | 2,0  | –   | –    |
| 2001                        | 1,7  | 0,4 | –    |
| 2005                        | 1,2  | 0,0 | 2,2  |
| 2009                        |      |     |      |
| inne                        | 3,0  | 0,3 | –    |
| 2001                        | 3,5  | 2,1 | –    |
| 2005                        | 1,8  | 0,0 | 0,0  |
| 2009                        |      |     |      |

Źródło: Opracowanie własne na podstawie *Turystyka i wypoczynek* ... (2002, 2006, 2010)

## 3. ANALIZA ZALEŻNOŚCI

W celu zidentyfikowania czynników wpływających na poziom uczestnictwa w wyjazdach turystycznych gospodarstw domowych seniorów, w których głowa gospodarstwa była w wieku 60 lat i więcej oraz związków zachodzących pomiędzy nimi, wykorzystano test niezależności  $\chi^2$ . Pozwala on określić statystyczną istotność związku między dwoma zmiennymi zarówno mierzalnymi, jak i niemierzalnymi (Luszniewicz, Słaby, 2008).

W tab. 6-16 w wierszach, poza ostatnim, podano liczebności składowe i liczebności ogółem kategorii zmiennych, których współzależność była badana. W celach porównawczych podane zostały także odsetki gospodarstw domowych uczestniczących lub nieuczestniczących w wyjazdach turystycznych. Natomiast ostatnie wiersze tych tabel zawierają informacje o wartościach statystyk  $\chi^2$ , liczbie stopni swobody (df) oraz krytycznym poziomie istotności (p).

Pierwszym z analizowanych czynników była płeć głowy gospodarstwa domowego. Okazało się, że płeć zasadniczo nie wpływa na uczestnictwo w wyjazdach turystycznych (tab. 6). Wniosek ten jest zgodny z wynikami badania, przeprowadzonego w skali regionalnej i ponadregionalnej, przedstawionymi w literaturze z zakresu turystyki (Alejziak, 2009, Brudnicki, 2006). Należy jednocześnie zauważyć, że odsetki gospodarstw, w których odnotowano wyjazd według płci są bardzo zbliżone, z nieznaczną przewagą gospodarstw, w których głową gospodarstwa są kobiety.

Tabela 6.

Płeć głowy gospodarstwa domowego a uczestnictwo w wyjazdach turystycznych

| Płeć          | Wyjazd turystyczny                    |       |     |       |        |        |
|---------------|---------------------------------------|-------|-----|-------|--------|--------|
|               | Tak                                   |       | Nie |       | Ogółem |        |
|               | N                                     | %     | N   | %     | N      | %      |
| mężczyźni     | 251                                   | 38,91 | 394 | 61,09 | 645    | 48,21  |
| kobiety       | 275                                   | 39,68 | 418 | 60,32 | 693    | 51,79  |
| Ogółem        | 526                                   | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$ | $\chi^2 = 0,083; df = 1; p = 0,774^2$ |       |     |       |        |        |

Źródło: obliczenia własne

Tabela 7.

Wiek głowy gospodarstwa domowego a uczestnictwo w wyjazdach turystycznych

| Wiek          | Wyjazd turystyczny                   |       |     |       |        |        |
|---------------|--------------------------------------|-------|-----|-------|--------|--------|
|               | Tak                                  |       | Nie |       | Ogółem |        |
|               | N                                    | %     | N   | %     | N      | %      |
| 60-64         | 198                                  | 45,31 | 239 | 54,69 | 437    | 32,66  |
| 65-69         | 119                                  | 42,65 | 160 | 57,35 | 279    | 20,85  |
| 70-74         | 99                                   | 41,42 | 140 | 58,58 | 239    | 17,86  |
| 75-79         | 64                                   | 31,53 | 139 | 68,47 | 203    | 15,17  |
| 80 i więcej   | 46                                   | 25,56 | 134 | 74,44 | 180    | 13,45  |
| Ogółem        | 526                                  | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$ | $\chi^2 = 27,773; df = 4; p = 0,000$ |       |     |       |        |        |

Źródło: obliczenia własne

Z informacji zawartych w tab. 7 wynika, że w przeciwieństwie do płci, wiek jest tym czynnikiem, który w dużym stopniu wpływa na aktywność turystyczną gospodarstw domowych seniorów. Badania wykazały istotny statystycznie związek pomiędzy wiekiem głowy gospodarstwa domowego a wyjazdami turystycznymi. Wraz ze wzrostem wieku głowy gospodarstwa domowego spada odsetek gospodarstw, w których odnotowano tego rodzaju wyjazdy. Największy odsetek uczestnictwa dotyczył najmłodszej grupy badanych (w wieku 60-64 lat) i wynosił ponad 45%. Grupa ta obejmuje gospodarstwa, w których głowa gospodarstwa przeszła już na emeryturę i które posiadają zabezpieczenie emerytalne i dużo wolnego czasu, który mogą przeznaczyć

<sup>2</sup> Jako poziom istotności przyjęto  $\alpha = 0,05$ .

na turystykę. Najmniejszym odsetkiem uczestniczących w wyjazdach charakteryzowały się gospodarstwa, w których głowa gospodarstwa miała 80 lat i więcej.

Zmienną, która wpływa na poziom aktywności turystycznej jest także wykształcenie (tab. 8). Stwierdzono istotny statystycznie związek poziomu wykształcenia głowy gospodarstwa domowego z odpowiedzią na pytanie dotyczące uczestnictwa w wyjazdach turystycznych. Rozkład odpowiedzi wskazuje, że im wyższe jest wykształcenie, tym większy jest udział gospodarstw uczestniczących w wyjazdach turystycznych. Najwyższym odsetkiem uczestnictwa w wyjazdach charakteryzowały się gospodarstwa domowe, w których głowa gospodarstwa posiadała wykształcenie wyższe, a następnie średnie. Zaś najniższy odsetek uczestnictwa dotyczył gospodarstw, w których głowa gospodarstwa legitymowała się wykształceniem co najwyżej gimnazjalnym.

Tabela 8.

Wykształcenie głowy gospodarstwa domowego a uczestnictwo w wyjazdach turystycznych

| Wykształcenie           | Wyjazd turystyczny                    |       |     |       |        |        |
|-------------------------|---------------------------------------|-------|-----|-------|--------|--------|
|                         | Tak                                   |       | Nie |       | Ogółem |        |
|                         | N                                     | %     | N   | %     | N      | %      |
| co najwyżej gimnazjalne | 117                                   | 24,79 | 355 | 75,21 | 472    | 35,28  |
| zasadnicze zawodowe     | 100                                   | 36,50 | 174 | 63,50 | 274    | 20,48  |
| średnie                 | 214                                   | 49,54 | 218 | 50,46 | 432    | 32,29  |
| wyższe                  | 95                                    | 59,38 | 65  | 40,63 | 160    | 11,96  |
| Ogółem                  | 526                                   | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$           | $\chi^2 = 88,570$ ; df = 3; p = 0,000 |       |     |       |        |        |

Źródło: obliczenia własne

Również miejsce zamieszkania gospodarstwa domowego jest istotnie powiązane z wyjazdami turystycznymi (tab. 9). Najniższym udziałem w wyjazdach turystycznych charakteryzowali się mieszkańcy wsi, ponieważ tylko nieco ponad 30% gospodarstw mieszkających na wsi wyjeżdżało w celach turystycznych. Natomiast najwyższy udział w wyjazdach turystycznych dotyczył mieszkańców dużych miast. Należy zauważyć, że im wyższy poziom urbanizacji, tym większy jest odsetek gospodarstw uczestniczących w wyjazdach.

Stwierdzono także występowanie statystycznie istotnego związku pomiędzy wyjazdem a stanem cywilnym głowy gospodarstwa domowego (tab. 10). Najbardziej aktywne turystycznie okazały się gospodarstwa, w których głowy gospodarstw były rozwiedzione lub żyły w separacji. Natomiast najmniejsza aktywność dotyczyła osób nie będących nigdy w związku małżeńskim.

Badania wykazały również istotny statystycznie związek pomiędzy typem biologicznym gospodarstwa domowego a uczestnictwem w wyjazdach turystycznych (tab. 11).

Tabela 9.

Miejsce zamieszkania gospodarstwa domowego a uczestnictwo w wyjazdach turystycznych

| Miejsce zamieszkania | Wyjazd turystyczny                    |       |     |       |        |        |
|----------------------|---------------------------------------|-------|-----|-------|--------|--------|
|                      | Tak                                   |       | Nie |       | Ogółem |        |
|                      | N                                     | %     | N   | %     | N      | %      |
| wieś                 | 157                                   | 30,84 | 352 | 69,16 | 509    | 38,04  |
| miasto <20 tys.      | 49                                    | 38,58 | 78  | 61,42 | 127    | 9,49   |
| 20 – 99 tys.         | 109                                   | 40,67 | 159 | 59,33 | 268    | 20,03  |
| 100 – 199 tys.       | 42                                    | 43,30 | 55  | 56,70 | 97     | 7,25   |
| 200 – 499 tys.       | 63                                    | 44,68 | 78  | 55,32 | 141    | 10,54  |
| 500 tys. i więcej    | 106                                   | 54,08 | 90  | 45,92 | 196    | 14,65  |
| Ogółem               | 526                                   | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$        | $\chi^2 = 35,803$ ; df = 5; p = 0,000 |       |     |       |        |        |

Źródło: obliczenia własne

Tabela 10.

Stan cywilny głowy gospodarstwa domowego a uczestnictwo w wyjazdach turystycznych

| Stan cywilny               | Wyjazd turystyczny                    |       |     |       |        |        |
|----------------------------|---------------------------------------|-------|-----|-------|--------|--------|
|                            | Tak                                   |       | Nie |       | Ogółem |        |
|                            | N                                     | %     | N   | %     | N      | %      |
| kawaler/panna              | 20                                    | 28,57 | 50  | 71,43 | 70     | 5,23   |
| żonaty/mężatka             | 286                                   | 40,86 | 414 | 59,14 | 700    | 52,32  |
| rozwódziona(a)/w separacji | 44                                    | 58,67 | 31  | 41,33 | 75     | 5,61   |
| wdowiec/wdowa              | 176                                   | 35,70 | 317 | 64,30 | 493    | 36,85  |
| Ogółem                     | 526                                   | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$              | $\chi^2 = 18,558$ ; df = 3; p = 0,000 |       |     |       |        |        |

Źródło: obliczenia własne

Tabela 11.

Typ biologiczny gospodarstwa domowego a uczestnictwo w wyjazdach turystycznych

| Typ                       | Wyjazd turystyczny                    |       |     |       |        |        |
|---------------------------|---------------------------------------|-------|-----|-------|--------|--------|
|                           | Tak                                   |       | Nie |       | Ogółem |        |
|                           | N                                     | %     | N   | %     | N      | %      |
| małżeństwo bez dzieci     | 237                                   | 43,17 | 312 | 56,83 | 549    | 41,03  |
| małżeństwo z dziećmi      | 9                                     | 52,94 | 8   | 47,06 | 17     | 1,27   |
| gospodarstwo jednoosobowe | 196                                   | 38,28 | 316 | 61,72 | 512    | 38,27  |
| pozostałe                 | 84                                    | 32,31 | 176 | 67,69 | 260    | 19,43  |
| Ogółem                    | 526                                   | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$             | $\chi^2 = 10,322$ ; df = 3; p = 0,016 |       |     |       |        |        |

Źródło: obliczenia własne



Najwyższym udziałem w wyjazdach turystycznych charakteryzowały się małżeństwa z dziećmi, najniższym zaś osoby samotne.

Na poziom aktywności turystycznej nie wpływa w istotny sposób wielkość gospodarstwa domowego (tab. 12). Niemniej jednak najwyższy odsetek uczestniczących w wyjazdach dotyczył członków gospodarstw czteroosobowych, a następnie dwuosobowych.

Tabela 12.

Wielkość gospodarstwa domowego zamieszkania a uczestnictwo w wyjazdach turystycznych

| Liczba osób<br>w gosp. domowym | Wyjazd turystyczny                   |       |     |       |        |        |
|--------------------------------|--------------------------------------|-------|-----|-------|--------|--------|
|                                | Tak                                  |       | Nie |       | Ogółem |        |
|                                | N                                    | %     | N   | %     | N      | %      |
| 1                              | 196                                  | 38,28 | 316 | 61,72 | 512    | 38,27  |
| 2                              | 257                                  | 41,19 | 367 | 58,81 | 624    | 46,64  |
| 3                              | 37                                   | 31,36 | 81  | 68,64 | 118    | 8,82   |
| 4                              | 19                                   | 46,34 | 22  | 53,66 | 41     | 3,06   |
| 5 i więcej                     | 17                                   | 39,53 | 26  | 60,47 | 43     | 3,21   |
| Ogółem                         | 526                                  | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$                  | $\chi^2 = 5,127$ ; df = 4; p = 0,274 |       |     |       |        |        |

Źródło: obliczenia własne

Stwierdzono występowanie statystycznie istotnego związku między wielkością dochodów przypadającą na jednego członka gospodarstwa domowego a wyjazdami turystycznymi (tab. 13). Wraz ze wzrostem wysokości dochodów w każdym kolejnym przedziale wzrasta odsetek gospodarstw wyjeżdżających. Najwyższe odsetki wyjeżdżających dotyczą gospodarstw, w których dochód miesięczny na jedną osobę przekraczał 1600 zł.

W badaniach wzięto także pod uwagę subiektywne odczucia respondentów na temat własnej sytuacji materialnej. Uzyskane wyniki przedstawiono w tab. 14. Pozwalają one stwierdzić, że występuje istotny statystycznie związek pomiędzy analizowaną zmienną a wyjazdami turystycznymi. Im lepsza ocena sytuacji materialnej, tym wyższy odsetek gospodarstw uczestniczących w wyjazdach. Można zauważyć, że w przypadku gospodarstw, które oceniły swoje warunki materialne jako bardzo dobre lub raczej dobre, w wyjazdach brało udział ok. 53-54% gospodarstw. Natomiast wśród gospodarstw oceniających swoją sytuację materialną gorzej, odsetki były już co najmniej o połowę niższe.

Badania nie wykazały istotnych zależności pomiędzy uczestnictwem w wyjazdach turystycznych a posiadaniem oszczędności (tab. 15) i zadłużeniem (tab. 16) gospodarstwa domowego. Należy przy tym zauważyć, że gospodarstwa seniorów w zasadzie

Tabela 13.

Dochody w przeliczeniu na jedną osobę w gospodarstwie domowym a uczestnictwo w wyjazdach turystycznych

| Dochody na jedną osobę | Wyjazd turystyczny                    |       |     |       |        |        |
|------------------------|---------------------------------------|-------|-----|-------|--------|--------|
|                        | Tak                                   |       | Nie |       | Ogółem |        |
|                        | N                                     | %     | N   | %     | N      | %      |
| do 799 zł              | 51                                    | 23,72 | 164 | 76,28 | 215    | 16,07  |
| 800-999 zł             | 65                                    | 28,63 | 162 | 71,37 | 227    | 16,97  |
| 1000-1199 zł           | 68                                    | 30,49 | 155 | 69,51 | 223    | 16,67  |
| 1200-1399 zł           | 82                                    | 41,21 | 117 | 58,79 | 199    | 14,87  |
| 1400-1599 zł           | 78                                    | 48,15 | 84  | 51,85 | 162    | 12,11  |
| 1600 złi więcej        | 182                                   | 58,33 | 130 | 41,67 | 312    | 23,32  |
| Ogółem                 | 526                                   | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$          | $\chi^2 = 92,940$ ; df = 5; p = 0,000 |       |     |       |        |        |

Źródło: obliczenia własne

Tabela 14.

Warunki materialne gospodarstwa domowego a uczestnictwo w wyjazdach turystycznych

| Warunki materialne | Wyjazd turystyczny                    |       |     |       |        |        |
|--------------------|---------------------------------------|-------|-----|-------|--------|--------|
|                    | Tak                                   |       | Nie |       | Ogółem |        |
|                    | N                                     | %     | N   | %     | N      | %      |
| bardzo dobre       | 7                                     | 53,85 | 6   | 46,15 | 13     | 0,97   |
| raczej dobre       | 126                                   | 52,94 | 112 | 47,06 | 238    | 17,79  |
| przeciętne         | 309                                   | 39,97 | 464 | 60,03 | 773    | 57,77  |
| raczej złe         | 63                                    | 27,04 | 170 | 72,96 | 233    | 17,41  |
| złe                | 21                                    | 25,93 | 60  | 74,07 | 81     | 6,05   |
| Ogółem             | 526                                   | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$      | $\chi^2 = 40,619$ ; df = 4; p = 0,000 |       |     |       |        |        |

Źródło: obliczenia własne

rzadko posiadają oszczędności – dotyczy to nieco ponad 7% gospodarstw, jeszcze rzadziej zaciągają pożyczki (tylko 5.46% gospodarstw).

Tabela 15.

Posiadanie oszczędności w gospodarstwie domowym a uczestnictwo w wyjazdach turystycznych

| Oszczędności  | Wyjazd turystyczny                   |       |     |       |        |        |
|---------------|--------------------------------------|-------|-----|-------|--------|--------|
|               | Tak                                  |       | Nie |       | Ogółem |        |
|               | N                                    | %     | N   | %     | N      | %      |
| Tak           | 44                                   | 45,36 | 53  | 54,64 | 97     | 7,25   |
| Nie           | 482                                  | 38,84 | 759 | 61,16 | 1241   | 92,75  |
| Ogółem        | 526                                  | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$ | $\chi^2 = 1,604$ ; df = 1; p = 0,205 |       |     |       |        |        |

Źródło: obliczenia własne

Tabela 16.

Zadłużenie gospodarstwa domowego a uczestnictwo w wyjazdach turystycznych

| Zadłużenie    | Wyjazd turystyczny                   |       |     |       |        |        |
|---------------|--------------------------------------|-------|-----|-------|--------|--------|
|               | Tak                                  |       | Nie |       | Ogółem |        |
|               | N                                    | %     | N   | %     | N      | %      |
| Tak           | 34                                   | 46,58 | 39  | 53,42 | 73     | 5,46   |
| Nie           | 492                                  | 38,89 | 773 | 61,11 | 1265   | 94,54  |
| Ogółem        | 526                                  | 39,31 | 812 | 60,69 | 1338   | 100,00 |
| Test $\chi^2$ | $\chi^2 = 1,707$ ; df = 1; p = 0,191 |       |     |       |        |        |

Źródło: obliczenia własne

## 4. ANALIZA LOGITOWA

W tej części artykułu podjęta zostanie próba równoczesnej analizy zależności pomiędzy zmienną określającą wystąpienie wyjazdu turystycznego a analizowanymi wyżej zmiennymi o charakterze ekonomicznym oraz społeczno-demograficznym. W celu określenia, które z tych zmiennych w istotny sposób wpływają na prawdopodobieństwo zaistnienia wyjazdu turystycznego wykorzystany zostanie model logitowy.

Model analizy logitowej jest szczególnym rodzajem modelu regresji, w którym zmienna zależna  $y_i^*$  jest nieobserwowalna (Maddala, 2006, 2008, Johnston, 1991):

$$y_i^* = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + u_i. \quad (1)$$

Nazywa się ją zmienną ukrytą i przyjmuje ona następujące wartości:

$$y_i = \begin{cases} 1 & \text{dla } y_i^* > 0, \\ 0 & \text{w pozostałych przypadkach.} \end{cases} \quad (2)$$

W przeprowadzonym badaniu dotyczącym wystąpienia bądź braku wyjazdu turystycznego powyższe równanie przyjmie postać:

$$y_i = \begin{cases} 1 & \text{gdy wystąpił wyjazd turystyczny,} \\ 0 & \text{w pozostałych przypadkach.} \end{cases} \quad (3)$$

Wówczas model logistyczny regresji dla zmiennej dychotomicznej określony jest równaniem (Kurkiewicz, 2007, Stanisław, 2007):

$$P_i(Y = 1 / x_1, x_2, \dots, x_k) = \frac{e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}}}{1 + e^{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}}}, \quad (4)$$

gdzie:  $\beta_j$  – współczynniki regresji,

$x_1, x_2, \dots, x_k$  – zmienne niezależne, które mogą być ilościowe lub jakościowe,

$i$  – numer obserwacji,

$j$  – numer zmiennej objaśniającej.

Prawdopodobieństwo  $P_i(Y = 1 / x_1, x_2, \dots, x_k)$  jest warunkowym prawdopodobieństwem, że zmienna  $y$  przyjmie wartość 1 dla wartości zmiennych objaśniających  $x_1, x_2, \dots, x_k$ .

Po oszacowaniu parametrów modeli można wyznaczyć teoretyczne wartości zmiennej  $Y$  według standardowej zasady prognozy (Gruszczyński, 2010):

$$\hat{y}_i = \begin{cases} 1 & \text{gdy } 0,5 < \hat{p}_i \leq 1, \\ 0 & \text{gdy } 0 \leq \hat{p}_i \leq 0,5, \end{cases} \quad (5)$$

gdzie  $\hat{p}_i$  – prawdopodobieństwa teoretyczne uzyskane z oszacowanego na podstawie próby losowej modelu regresji logistycznej.

W sytuacji, gdy próba jest niezbilansowana<sup>3</sup>, czyli taka, w której liczba jedynek znacznie różni się od liczby zer, do prognozowania wartości teoretycznych można zastosować modyfikację standardowej zasady i liczyć prognozy według zasady optymalnej wartości granicznej  $\alpha$  (Gruszczyński, 2002):

$$\hat{y}_i = \begin{cases} 1 & \text{gdy } \alpha < \hat{p}_i \leq 1, \\ 0 & \text{gdy } 0 < \hat{p}_i \leq \alpha. \end{cases} \quad (6)$$

Wartość graniczną  $\alpha$  ustala się jako udział jedynek w próbie. Poprawność oszacowanego modelu można sprawdzić wykorzystując test ilorazu wiarygodności, służący do weryfikacji hipotezy zerowej mówiącej o tym, że wszystkie parametry modelu, poza wyrazem wolnym, są równe zero. Statystyka testu wyrażona jest wzorem (Gruszczyński, 2010):

$$\chi^2 = 2 (\ln L_{UR} - \ln L_R), \quad (7)$$

<sup>3</sup> Użyta próba do modelowania wyjazdów turystycznych była niezbilansowana.

gdzie:

$L_{UR}$  – wartość funkcji wiarygodności dla modelu pełnego,

$L_R$  – wartość funkcji wiarygodności dla modelu zredukowanego do wyrazu wolnego.

Do pomiaru efektywności prognoz można wykorzystać miary wyznaczone na podstawie tablicy trafności klasyfikacji przypadków (tab. 17).

Tabela 17.

Budowa tablicy trafności

| Faktyczne | Przewidywane    |                 | Razem    |
|-----------|-----------------|-----------------|----------|
|           | $\hat{y}_i = 1$ | $\hat{y}_i = 0$ |          |
| $y_i = 1$ | $n_{11}$        | $n_{10}$        | $n_{1.}$ |
| $y_i = 0$ | $n_{01}$        | $n_{00}$        | $n_{0.}$ |
| Razem     | $n_{.1}$        | $n_{.0}$        | $n$      |

Źródło: (Gruszczyński, 2010, s. 73)

Korzystając z tablicy trafności wyznaczono dwie miary:

$$\text{zliczeniowy współczynnik } R^2 \quad R^2 = \frac{n_{00} + n_{11}}{n} \cdot 100\%, \quad (8)$$

$$\text{iloraz szans} \quad IS = \frac{n_{11} \cdot n_{00}}{n_{01} \cdot n_{10}}. \quad (9)$$

Ocena zliczeniowego współczynnika  $R^2$  oznacza procent trafnie zaklasyfikowanych przypadków. Im bliższa jednocy jest wartość tej miary tym lepsze dopasowanie modelu do danych empirycznych badanego zjawiska. Model dobrze sprawdza się w prognozowaniu, gdy zliczeniowy  $R^2 > 50\%$ . Oznacza to, że klasyfikacja na podstawie modelu jest lepsza od przypadkowej. W przypadku ilorazu szans, im wartość tej miary jest większa od 1, tym lepsza jest klasyfikacja na podstawie modelu.

Do określenia zależności między wystąpieniem wyjazdu turystycznego a zmiennymi związanymi z sytuacją ekonomiczną i społeczno-demograficzną gospodarstwa domowego uwzględniono następujące zmienne objaśniające (tab. 18): wiek głowy gospodarstwa domowego (*WIEK*), wykształcenie głowy gospodarstwa domowego (*WYKSZ*), miejsce zamieszkania (*MZAM*), stan cywilny głowy gospodarstwa domowego (*STAN CYW*), typ biologiczny gospodarstwa domowego (*TYP BIOL*), dochód gospodarstwa domowego w przeliczeniu na jedną osobę w zł (*DOCH*), warunki materialne gospodarstwa domowego (*SYT MAT*). Ponieważ większość zmiennych (oprócz wieku i dochodu) przedstawionych w tabeli została podzielona na pewne warianty, należało więc dokonać wyboru kategorii bazowych, które są płaszczyzną odniesienia dla dokonywanych w analizie ekonometrycznej interpretacji parametrów. Wprowadzenie kategorii bazowych było konieczne ze względów estymacyjnych, ponieważ suma każdej zmiennej była równa 1, a więc zmiennej odpowiadającej wyrazowi wolnemu. Modelu

bez pominiętych zmiennych bazowych nie da się oszacować. Wybrano następujące kategorie bazowe: *WYKSZ 1*, *MZAM 1*, *STAN CYW 1*, *TYP BIOL 3*, *SYT MAT 5*.

Tabela 18.

Zestawienie zmiennych objaśniających

| Symbol zmiennej   | Opis zmiennej                                                | Symbol zmiennej   | Opis zmiennej                                                 |
|-------------------|--------------------------------------------------------------|-------------------|---------------------------------------------------------------|
| <i>WIEK</i>       | wiek w latach                                                | <i>STAN CYW 3</i> | rozwód(a)/w separacji – 1,<br>0 w pozostałych przypadkach     |
| <i>WYKSZ 1</i>    | co najwyżej gimnazjalne – 1,<br>0 w pozostałych przypadkach  | <i>STAN CYW 4</i> | wdowiec/wdowa – 1,<br>0 w pozostałych przypadkach             |
| <i>WYKSZ 2</i>    | zasadnicze zawodowe – 1,<br>0 w pozostałych przypadkach      | <i>TYP BIOL 1</i> | małżeństwo bez dzieci – 1,<br>0 w pozostałych przypadkach     |
| <i>WYKSZ 3</i>    | średnie – 1,<br>0 w pozostałych przypadkach                  | <i>TYP BIOL 2</i> | małżeństwo z dziećmi – 1,<br>0 w pozostałych przypadkach      |
| <i>WYKSZ 4</i>    | wyższe – 1,<br>0 w pozostałych przypadkach                   | <i>TYP BIOL 3</i> | gospodarstwo jednoosobowe – 1,<br>0 w pozostałych przypadkach |
| <i>MZAM 1</i>     | wieś – 1,<br>0 w pozostałych przypadkach                     | <i>TYP BIOL 4</i> | pozostałe – 1,<br>0 w pozostałych przypadkach                 |
| <i>MZAM 2</i>     | miasto <20 tys. – 1,<br>0 w pozostałych przypadkach          | <i>DOCH</i>       | dochód w zł na 1 osobę                                        |
| <i>MZAM 3</i>     | miasto 20 – 99 tys. – 1,<br>0 w pozostałych przypadkach      | <i>SYT MAT 1</i>  | bardzo dobre – 1,<br>0 w pozostałych przypadkach              |
| <i>MZAM 4</i>     | miasto 100 – 199 tys. – 1,<br>0 w pozostałych przypadkach    | <i>SYT MAT 2</i>  | raczej dobre – 1,<br>0 w pozostałych przypadkach              |
| <i>MZAM 5</i>     | miasto 200 – 499 tys. – 1,<br>0 w pozostałych przypadkach    | <i>SYT MAT 3</i>  | przeciętne – 1,<br>0 w pozostałych przypadkach                |
| <i>MZAM 6</i>     | miasto 500 tys. i więcej – 1,<br>0 w pozostałych przypadkach | <i>SYT MAT 4</i>  | raczej złe – 1,<br>0 w pozostałych przypadkach                |
| <i>STAN CYW 1</i> | kawaler/panna – 1,<br>0 w pozostałych przypadkach            | <i>SYT MAT 5</i>  | złe – 1,<br>0 w pozostałych przypadkach                       |
| <i>STAN CYW 2</i> | żonaty/mężatka – 1,<br>0 w pozostałych przypadkach           |                   |                                                               |

Źródło: opracowanie własne

Oceny parametrów modelu logitowego przedstawiono w tab. 19. Zmienne nieistotne statystycznie eliminowano z modelu metodą krokową (*a posteriori*). Procedura ta polega na porównaniu wartości prawdopodobieństwa  $p$  z przyjętym poziomem istotności  $\alpha$ . Jeżeli  $p$  było większe od  $\alpha$ , zmienną eliminowano z modelu i dokonywano ponownej estymacji. Działanie to powtarzano aż do uzyskania modelu z wartościami  $p$  mniejszymi lub równymi  $\alpha$ . Wartość  $\alpha$  przyjęto na poziomie 5%. W ostatnim wierszu tabeli zamieszczono statystykę testu ilorazu wiarygodności oraz wartości dwóch miar: ilorazu szans i zliczeniowego  $R^2$  (ogólna trafność klasyfikacji). Dla badanego modelu statystyka empiryczna  $\chi^2$  świadczy o tym, że na poziomie istotności 0,05 należy odrzucić hipotezę zerową mówiącą o łącznej nieistotności parametrów modeli. Również iloraz szans jest dość wysoki.

Tabela 19.

Wyniki oszacowania modelu logitowego wyjazdów turystycznych gospodarstw domowych seniorów

| Zmienna                                                                               | Ocena parametru | Błąd standardowy | Statystyka t-Studenta | Wartość p |
|---------------------------------------------------------------------------------------|-----------------|------------------|-----------------------|-----------|
| Wyraz Wolny                                                                           | 0,3768          | 0,4847           | 0,7776                | 0,4370    |
| WIEK                                                                                  | -0,0398         | 0,0073           | -5,4167               | 0,0000    |
| WYKSZ 3                                                                               | 0,6301          | 0,1358           | 4,6393                | 0,0000    |
| WYKSZ 4                                                                               | 0,7085          | 0,2061           | 3,4371                | 0,0006    |
| STAN CYW 3                                                                            | 1,0907          | 0,2950           | 3,6972                | 0,0002    |
| STAN CYW 4                                                                            | 0,6008          | 0,2035           | 2,9526                | 0,0032    |
| TYP BIOL 1                                                                            | 0,5778          | 0,1877           | 3,0777                | 0,0021    |
| DOCH                                                                                  | 0,5915          | 0,1236           | 4,7856                | 0,0000    |
| SYT MAT 2                                                                             | 0,5199          | 0,2090           | 2,4879                | 0,0130    |
| SYT MAT 3                                                                             | 0,4003          | 0,1570           | 2,5499                | 0,0109    |
| $\chi^2 = 166,30$ df = 9 p = 0,000, IS = 3,97<br>Ogólna trafność klasyfikacji: 67,74% |                 |                  |                       |           |

Źródło: obliczenia własne na podstawie danych GUS

Jak wynika z oszacowanego modelu spośród zmiennych obejmujących czynniki społeczno-demograficzne istotnie statystycznie okazało się miejsce zamieszkania<sup>4</sup>. Oznacza to, że takie same szanse wyjazdu turystycznego mają zarówno osoby zamieszkałe na wsi jak i w mieście. Prawdopodobieństwo wyjazdu turystycznego maleje wraz z wiekiem głowy gospodarstwa domowego, a rośnie wraz ze wzrostem dochodu gospodarstwa w przeliczeniu na jedną osobę.

Istotnym czynnikiem okazały się również subiektywne odczucia respondentów na temat własnej sytuacji materialnej. Otóż gospodarstwa, które oceniają swoją sytuację jako dobrą lub przeciętną mają większą szansę wyjazdu niż gospodarstwa w złej sytuacji materialnej. Istotne statystycznie znaczenie dla wyjazdów turystycznych ma wykształcenie głowy gospodarstwa domowego. Prawdopodobieństwo wyjazdu okazało się większe dla osób z wykształceniem wyższym i średnim w porównaniu z osobami z wykształceniem co najwyżej gimnazjalnym.

Czynnikiem różnicującym szanse wyjazdu jest także stan cywilny. Osoby rozwiedzione, żyjące w separacji oraz te, które utraciły współmałżonka w wyniku jego zgonu mają większe prawdopodobieństwo wyjazdu niż osoby, które nigdy nie żyły w związku małżeńskim. Czynnikiem sprzyjającym wyjazdom jest również typ biolo-

<sup>4</sup> Może to być związane ze współwystępowaniem tej zmiennej z innymi zmiennymi, które weszły do modelu.

giczny gospodarstwa. Małżeństwa bez dzieci posiadają większe prawdopodobieństwo wyjazdu w porównaniu z gospodarstwami jednoosobowymi.

## 5. PODSUMOWANIE

Z przeprowadzonej analizy wynika, że czynne wykorzystanie czasu wolnego przez seniorów wciąż jeszcze jest mało popularne. Chętniej oni słuchają radia, muzyki, oglądają telewizję niż uprawiają sport, ćwiczenia fizyczne lub biorą udział w wycieczkach i imprezach objazdowych. Znacznie rzadziej uczestniczą w imprezach turystycznych. Najczęściej ich plany urlopowe związane były z wyjazdami do rodziny lub znajomych a później z pobytem wypoczynkowym. Podróże, szczególnie krajowe, organizowane były na ogół całkowicie samodzielnie.

Badaniami objęto zależności zachodzące pomiędzy 11 zmiennymi a uczestnictwem w wyjazdach turystycznych. Okazało się, że aż siedem zmiennych wykazywało istotną statystycznie zależność. Brak zależności dotyczył: płci, wielkości gospodarstwa domowego, posiadanych oszczędności i zadłużenia.

W artykule wykazano także, że na prawdopodobieństwo wystąpienia wyjazdu turystycznego wpływają: wiek głowy gospodarstwa domowego, jego wykształcenie, stan cywilny, typ biologiczny gospodarstwa domowego oraz sytuacja ekonomiczna gospodarstwa.

Projektowane zmiany w systemie emerytalnym, polegające na wydłużeniu okresu uprawniającego do otrzymania emerytury, mogą wpłynąć na zmniejszenie aktywności turystycznej gospodarstw domowych seniorów. Przy założeniu dotychczasowego czteroletniego cyklu badań aktywności turystycznej gospodarstw domowych pierwsze następstwa tych zmian będzie można zaobserwować w badaniu w roku 2017.

*Zachodniopomorski Uniwersytet Technologiczny w Szczecinie*

## LITERATURA

- [1] Alejziak W., (2009), *Determinanty i zróżnicowanie społeczne aktywności turystycznej*, Studia i Monografie nr 56, Akademia Wychowania Fizycznego im. Bronisława Czecha w Krakowie, Kraków.
- [2] Brudnicki R., (2006), *Społeczno-ekonomiczne uwarunkowania zachowań turystycznych mieszkańców Torunia*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- [3] Chabior A., (2005), *Aktywność życiowa i jakość życia seniorów*, [w:] Seniorzy w rodzinie, instytucji i społeczeństwie, red. Fabiś A. Wybrane zagadnienia współczesnej gerontologii, Wydawnictwo Wyższej Szkoły Zarządzania i Marketingu, Sosnowiec.
- [4] Fabiś A. (red.), (2008), *Aktywność społeczna, kulturalna i oświatowa seniorów*, Wyd. Wyższa Szkoła Administracji, Uniwersytet Łódzki, Bielsko-Biała-Łódź.
- [5] Gruszczyński M. (2002), *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Oficyna wydawnicza AGH, Warszawa.
- [6] Gruszczyński M. (red.), (2010), *Mikroekonometria. Modele i metody analizy danych indywidualnych*, Wolters Kluwer Polska Sp. z o.o., Warszawa.
- [7] Johnston J., (1991), *Econometric methods*. McGraw-Hill Book Company, Singapore.



- [8] Kurkiewicz J.(red.), (2007), *Ludzie starsi w rodzinie i społeczeństwie*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków.
- [9] *Ludność. Stan i struktura w przekroju terytorialnym*, (2010), GUS, Warszawa.
- [10] Luszniwicz A., Słaby T., (2008), *Statystyka z pakietem komputerowym Statistica Pl. Teoria i zastosowania*, Wydawnictwo C. H. Beck, Warszawa.
- [11] Łaciak J., (2007), *Uczestnictwo Polaków w wyjazdach turystycznych w 2006 roku*, Instytut Turystyki, Warszawa.
- [12] Maddala G.S., (2006, 2008), *Ekonometria*. PWN, Warszawa.
- [13] Napierała M., (2002), *Turystyka w „jesieni życia”*, [w:] Partnerstwo nauki i praktyki w turystyce – fakty, intencje, potrzeby rozwoju, red. R. Przybyszewska-Gudelis, M. Grabiszewski, Wyższa Pomorska Szkoła Turystyki i Hotelarstwa w Bydgoszczy, Bydgoszcz.
- [14] *Turystyka i wypoczynek w 2001 roku*, (2002), Informacje i Opracowania Statystyczne, GUS, Warszawa.
- [15] *Turystyka i wypoczynek w gospodarstwach domowych w 2005 roku*, (2006), Informacje i Opracowania Statystyczne, GUS, Warszawa.
- [16] *Turystyka i wypoczynek w gospodarstwach domowych w 2009 roku*, (2010), Informacje i Opracowania Statystyczne GUS, Warszawa.
- [17] Stanisław A., (2007), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, Statsoft Polska Sp z o.o. Kraków.
- [18] Śniadek J., (2007), *Konsumpcja turystyczna polskich seniorów na tle globalnych tendencji w turystyce*, Gerontologia Polska, tom 15, nr 1-2, 21-30.

## STATYSTYCZNA ANALIZA AKTYWNOŚCI TURYSTYCZNEJ SENIORÓW W POLSCE

### Streszczenie

Celem artykułu jest analiza zależności między zmienną określającą wystąpienie wyjazdu turystycznego a zmiennymi ekonomicznymi i społeczno-demograficznymi charakteryzującymi gospodarstwa domowe seniorów w Polsce. Podstawą informacyjną badań są dane statystyczne dotyczące wyjazdów turystycznych gospodarstw domowych, w których głowa gospodarstwa jest w wieku 60 lat i więcej. Zostały one zaczerpnięte z przeprowadzonego w 2009 roku przez Główny Urząd Statystyczny badania ankietowego nt. „Turystyka i wypoczynek w gospodarstwach domowych. W badaniu istotności wpływu czynników determinujących wystąpienie wyjazdu turystycznego wykorzystano test niezależności  $\chi^2$  oraz model logitowy.

**Słowa kluczowe:** aktywność turystyczna seniorów, analiza zależności, model logitowy

## APPLICATION OF SELECTED STATISTICAL AND ECONOMETRIC METHODS IN THE ANALYSIS OF SENIOR TOURISM ACTIVITIES IN POLAND

### Abstract

Purpose of this paper is to present an analysis of the dependences between a variable determining occurrence of a tourist trip and economic and socio-demographic variables that characterize households

of seniors in Poland. The study is based on statistical data regarding tourist trips taken by members of households, where half of the household is of the age of 60 or more years. It has been culled from a survey performed in 2009 by the National Statistical Office, called „Tourism and leisure at households”. In researches about significance of factors determining occurrence of a tourist trip were used  $\chi^2$  independence test and logit model.

**Key words:** seniors tourism activity, analysis of relationship, logit model

MARZENA PIOTROWSKA-TRYBULL, STANISŁAW SIRKO

### WPLYW JEDNOSTKI WOJSKOWEJ NA ROZWÓJ GMINY<sup>1</sup>

Rozwój lokalny jest kategorią złożoną. Określa się go jako proces zróżnicowania i wzbogacania działalności ekonomicznej i społecznej na danym terytorium, polegający na mobilizacji i koordynacji zasobów i energii (Jewtuchowicz, 2005). Można go rozpatrywać w płaszczyznach: gospodarczej, społecznej, infrastrukturalnej oraz środowiskowej, które z kolei są identyfikowane poprzez zmienne ilustrujące lokalne przeobrażenia.

W pewnym stopniu o sytuacji społeczno-ekonomicznej w gminie może przesądzać jednostka wojskowa, która znajduje się na jej terenie. Więzi występujące pomiędzy jednostką wojskową a otoczeniem mają charakter dwustronny. Jednostka czerpiąc z otoczenia zasoby, realizuje przypisane jej zadania, natomiast do otoczenia dostarcza dobro publiczne, jakim jest bezpieczeństwo obywateli oraz generuje efekty wynikające z użytkowania zasobów. Relacje, które występują pomiędzy jednostką wojskową a środowiskiem lokalnym, z jednej strony są złożone, a drugiej strony często niedostrzegane. Dopiero likwidacja jednostki wojskowej sprawia, że ujawniają się zagrożenia na gruncie społecznym i ekonomicznym. Świadomość powyższego przyczyniła się do podjęcia badań zmierzających do określenia roli jednostki wojskowej w funkcjonowaniu lokalnej społeczności<sup>2</sup>. W artykule przedstawiono wyniki badań sondażowych przeprowadzonych wśród mieszkańców gmin oraz żołnierzy, w trakcie których zmierzano do poznania ich opinii w zakresie wpływu jednostki wojskowej na rozwój gminy.

<sup>1</sup> Przedstawiono wybrane wyniki badań sondażowych, które uzyskano podczas realizacji pracy naukowej pt. *Uwarunkowania lokalne i regionalne funkcjonowania jednostek wojskowych na terytorium kraju* finansowanej ze środków na naukę, w latach 2010-2013, jako projekt badawczy Ministerstwa Nauki i Szkolnictwa Wyższego, nr O N115 311238.

<sup>2</sup> Konsekwencje pozostawiania jednostki wojskowej na terenie gminy, w czasie kryzysu i wojny nie były rozpatrywane. Nie oznacza to jednak, że w czasie pokoju nie mogą pojawić się takie sytuacje, związane z obecnością jednostki wojskowej na terenie gminy, które nie wpłyną negatywnie na sytuację w jej otoczeniu np. wybuch składu amunicji.

## 1. ORGANIZACJA BADAŃ, PODSTAWOWE INFORMACJE

Badania rozpoczęto od identyfikacji gmin, na terenie których stacjonują jednostki wojskowe. Spośród nich wylosowano<sup>3</sup> siedemnaście gmin rozmieszczonych na terenie całego kraju, w których przeprowadzono badania sondażowe. W każdej z gmin, a były wśród nich gminy miejskie, miejsko-wiejskie i wiejskie (9 gmin miejskich, 1 gmina miejsko-wiejska, 7 gmin wiejskich), badaniami, które zrealizowano metodą CATI (Computer Assisted Telephone Interview), objęto minimum pięćdziesiąt osób. Zalety i wady metody zostały omówione w pracy Szredera (2010). Dobór osób do badania był losowy (w sumie uczestniczyło w nich 856 mieszkańców gmin). Próba była próbą kwotowo-losową. Charakterystyka metod doboru próby znajduje się w pracach Aczela (2000) oraz Szredera (2010). Podział badanych ze względu na wiek (18-34, 35-44, 45-54, 55-75 lat) i płeć (kobiety, mężczyźni) uwzględniał proporcje osób mieszkających w każdej z gmin. Losowanie numerów telefonów respondentów, bez zwracania, zrealizowano z jednakowym prawdopodobieństwem dla każdego numeru. Badaniami objęto także 624 żołnierzy pełniących służbę w 19 jednostkach wojskowych głównych rodzajów sił zbrojnych. Dobór jednostek wojskowych i żołnierzy był również losowy, a ich udział w badaniach dobrowolny (losowanie przeprowadziło Wojskowe Biuro Badań Społecznych). Położenie wylosowanych gmin i jednostek wojskowych pokazano na rysunku 1.

Wywiady z mieszkańcami gmin przeprowadziły na zlecenie autorów artykułu, firmy: MillwardBrown SMG/KRC w Warszawie oraz 4P. Autorzy uczestniczyli w przygotowaniu pytań, które zostały zadane respondentom. Badania w jednostkach wojskowych wykonało Wojskowe Biuro Badań Społecznych (WBBS), wykorzystując przygotowany przez autorów kwestionariusz ankiety. Pytania zamieszczone w kwestionariuszu ankiety były zasadniczo pytaniami skategoryzowanymi. Respondenci mogli też przekazać swoje opinie w ramach pytań otwartych. W artykule zamieszczono wyniki własnych analiz, które przeprowadzono na podstawie danych uzyskanych od SMG/KRC, 4P i WBBS. Wykonano je za pomocą programów komputerowych Excel i Statistica v.9.1. Informacje dotyczące gmin uzyskano ze stron WWW Głównego Urzędu Statystycznego.

## 2. ZASADNICZE INFORMACJE DOTYCZĄCE RESPONDENTÓW I GMIN

Wśród badanych mieszkańców gmin były osoby zróżnicowane wiekowo (24,3% w wieku 18-34 lat, 16,6% 35-44 lat, 26,6% 45-54 lat, 32,5% 55-75 lat), zarówno aktywne zawodowo (67,7% ankietowanych), jak i pozostające bez pracy (32,3%)<sup>4</sup>. Na podstawie uzyskanych wyników można wskazać gminy, w których odsetek zatrudnio-

<sup>3</sup> Wylosowano je spośród zbioru gmin, który zaczerpnięto z Rozporządzenia Ministra Obrony Narodowej z 24 listopada 2009 r. w sprawie utworzenia garnizonów oraz określenia zadań, siedzib i terytorialnego zasięgu właściwości i ich dowódców (Dz. U. Nr 206, poz. 1592).

<sup>4</sup> W tej grupie poza bezrobotnymi znalazły się osoby uczące się oraz emeryci.



Rysunek 1. Położenie geograficzne gmin, w których realizowano badania

Źródło: opracowanie własne

nych i bezrobotnych respondentów był zbliżony (Hrubieszów, Chełmno, Sandomierz, Braniewo, Brzeg, Nowa Dęba) oraz takie, gdzie przeważały osoby zatrudnione (pozostałe gminy)<sup>5</sup>. Ankietowani posiadali różne wykształcenie, najczęściej średnie (46,2%), pozostali: wyższe (15,3%), pomaturalne (21,1%), zawodowe (16,0%) i podstawowe (1,4%). Badani, którzy zajmowali stanowiska kierownicze lub byli pracownikami umysłowymi, uzyskiwali wyższe dochody. Wysokie zarobki nie były też obce pracownikom fizycznym, bowiem co piąty z nich zarabiał ponad 4000,00 zł. Ankietowani pochodzili ze zróżnicowanych liczebnie rodzin. W połowie gmin były to jedna lub dwie osoby w rodzinie, w pozostałych cztery i więcej.

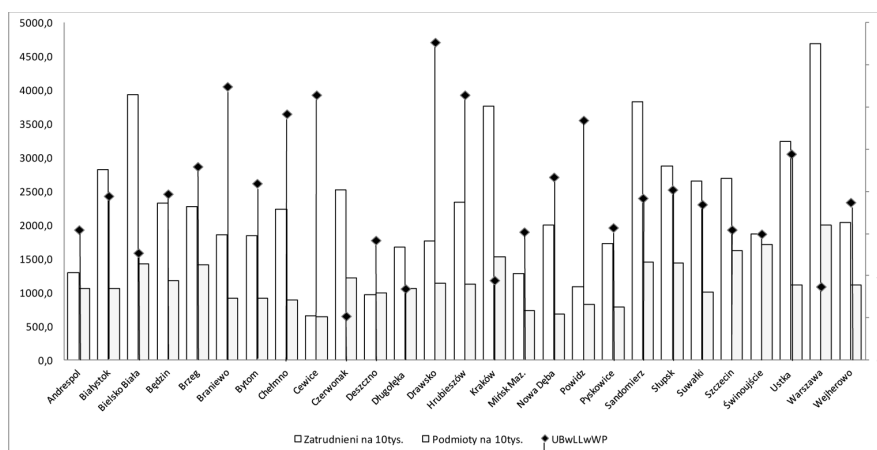
W jednostkach wojskowych badaniami objęto oficerów (23,4% – w tym oficerowie starsi 13,3%, oficerowie młodszy 10,1%), podoficerów (38,3%) i szeregowych zawodowych (35,9%)<sup>6</sup>. Stosunkowo duży odsetek oficerów starszych wynikał z faktu, że w badaniach, między innymi, uczestniczyli przedstawiciele Ministerstwa Obrony Narodowej, Sztabu Generalnego Wojska Polskiego, pracownicy sztabów poszczególnych rodzajów sił zbrojnych, gdzie na wielu stanowiskach zatrudnieni są oficerowie

<sup>5</sup> Przedstawione wyniki były istotne statystycznie na poziomie 0,05. Do weryfikacji hipotez statystycznych stosowano test  $\chi^2$ .

<sup>6</sup> Informacje podane przez 2,4% badanych były niekompletne, dlatego też w dalszych analizach ich nie uwzględniano.

starsi. Wśród oficerów młodszych (88,9% badanych z tej grupy), podoficerów (86,2%) i szeregowych zawodowych (97,9%), dominowały osoby znajdujące zatrudnienie w jednostkach wojskowych niżej usytuowanych w strukturze sił zbrojnych. Badani żołnierze to osoby w różnym wieku (średnia – 33,1 lat, rozstęp – 37 lat), z różnym stażem służby wojskowej (rozstęp – 36,5 lat). Na stałe mieszkają oni w różnych środowiskach. Blisko jedna czwarta spośród nich na wsi (zasadniczo szeregowi zawodowi), pozostali w różnej wielkości miastach.

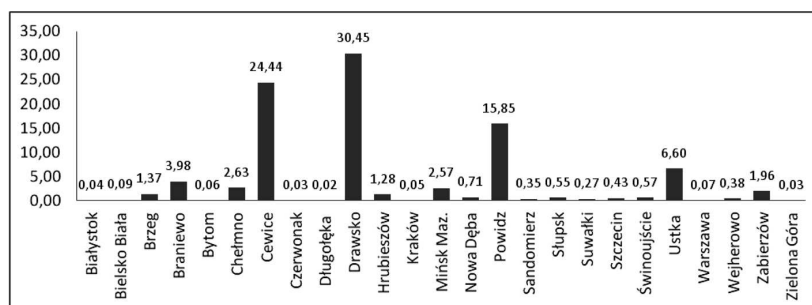
Podstawowe informacje dotyczące sytuacji na rynku pracy w gminach, gdzie przeprowadzono badania oraz wpływów z podatków do budżetu gmin, z racji stacjonowania na ich terenie jednostki wojskowej, przedstawiono na rysunkach 2 i 3.



UBwLLwWP - udział bezrobotnych zarejestrowanych w liczbie ludności w wieku produkcyjnym (prawa oś OY)

Rysunek 2. Sytuacja na rynku pracy w gminach, w grudniu 2010 roku

Źródło: opracowanie własne na podstawie <http://www.stat.gov.pl/bdl/app/portret.dims> (dostęp 27.04.2012r.)



Ze względu na brak danych z Rejonowych Zarządów Infrastruktury na rysunku brakuje gmin: Andrespol, Będzin, Deszczno, Pyskowice, Wysokie Mazowieckie.

Rysunek 3. Udział procentowy wpływów z podatków od: nieruchomości, rolnego i leśnego, w dochodach własnych poszczególnych gmin, w 2010 roku, z racji stacjonowania na ich terenie jednostki wojskowej

Źródło: opracowanie własne

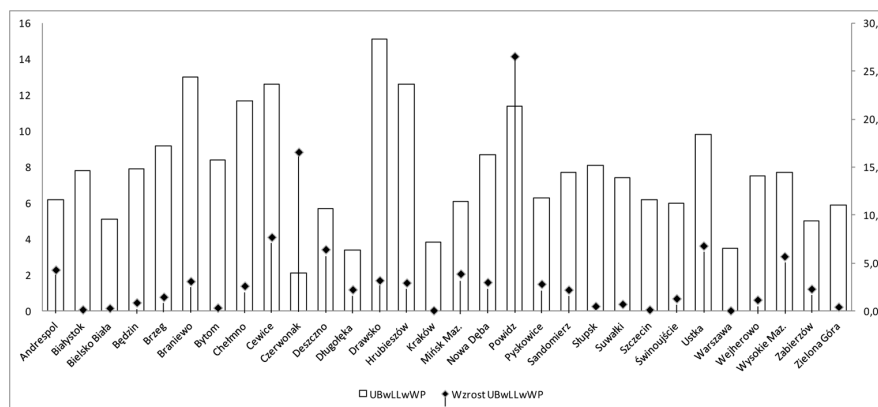
### 3. OPINIE RESPONDENTÓW NA TEMAT WPŁYWU JEDNOSTKI WOJSKOWEJ NA ROZWÓJ GMINY

Opinie mieszkańców na temat obecności jednostki wojskowej na terenie ich gminy były zróżnicowane. Z jednej strony większość z nich (51,9%) pozytywnie postrzegała ten fakt, co korespondowało z wynikami prezentowanymi w raportach Felisiaka (2010, 2012). Głównie byli to mieszkańcy gmin znajdujących się blisko granicy kraju (Deszczno, Świnoujście, Braniewo, Hrubieszów), z drugiej duży odsetek (43,5%) wyborów neutralnych odnotowano wśród mieszkańców gmin położonych w pobliżu centralnej części kraju (Andrespol, Czerwonak). Można przypuszczać, że taka ocena w pewnym stopniu może wynikać z poczucia bezpieczeństwa, jakie daje mieszkańcom położonym w różnych regionach kraju, obecność jednostki wojskowej. Obecność ta powoduje, że blisko połowa (48,0%) respondentów czuje się bezpieczniej. Można to odnieść do prezentowanego w literaturze ujęcia, gdzie bezpieczeństwo postrzegane jest jako stan, kiedy określone podmioty nie dostrzegają zagrożeń dla swojego rozwoju (Cieślarczyk, 2009). Jednocześnie należy mieć na uwadze to, że w czasie pokoju pojawia się szereg zagrożeń zarówno dla jednostek, jak i poszczególnych grup, którym należy przeciwdziałać tworząc warunki dla trwania, przetrwania i rozwoju. Obok innych sfer, można odnieść to do sfery bezpieczeństwa ekonomicznego, w tym bezpieczeństwa zatrudnienia. W tym względzie obecność jednostek wojskowych była częściej dostrzegana przez mieszkańców w gminach miejskich, niż w wiejskich. Można jedynie przypuszczać, że ma to związek z sytuacją na poszczególnych rynkach pracy, jak również położeniem jednostek wojskowych. W miastach życie toczy się niejako wokół nich i łatwiej jest zaobserwować powiązania jednostek wojskowych z rynkiem pracy, w przeciwieństwie do terenów wiejskich, w których lokalizowane są tzw. zielone garnizony.

Rynek pracy w badanych gminach charakteryzuje się dużym stopniem złożoności<sup>7</sup>. W latach 2002-2010 liczba podmiotów gospodarczych zarejestrowanych w REGON (średnio ujmując), w sferze publicznej zmniejszyła się z 337 w 2002 roku, do 283 w 2010 roku. W sferze prywatnej wzrosła ona z 10161 w 2002 roku do 11067 w 2010 roku. Jednostki wojskowe są więc jednym z wielu pracodawców i w pewnym stopniu wpływają na lokalne rynki pracy. Z jednej strony pełnią one rolę pracodawcy dla mieszkańców gminy i żołnierzy, z drugiej – w wyniku ich likwidacji, pojawia się więcej bezrobotnych (rysunek 4)<sup>8</sup>. Likwidacja jednostki może się także przyczynić do spadku popytu na towary i usługi firm, które dostarczały je wojsku, jako instytucji oraz na które zgłaszali zapotrzebowanie pracownicy wojska.

<sup>7</sup> Porównywanie zróżnicowanych rynków pracy tych gmin jest trudne i niekiedy może prowadzić do błędnych wniosków.

<sup>8</sup> Dla wszystkich gmin przyjęto jednakową liczbę zwolnionych w wyniku likwidacji jednostki wojskowej. W obliczeniach uwzględniono żołnierzy i pracowników wojska. Przyjęto, że w gminie pozostanie tylko pewien odsetek żołnierzy, pozostali znajdą zatrudnienie w innych jednostkach na terenie kraju.

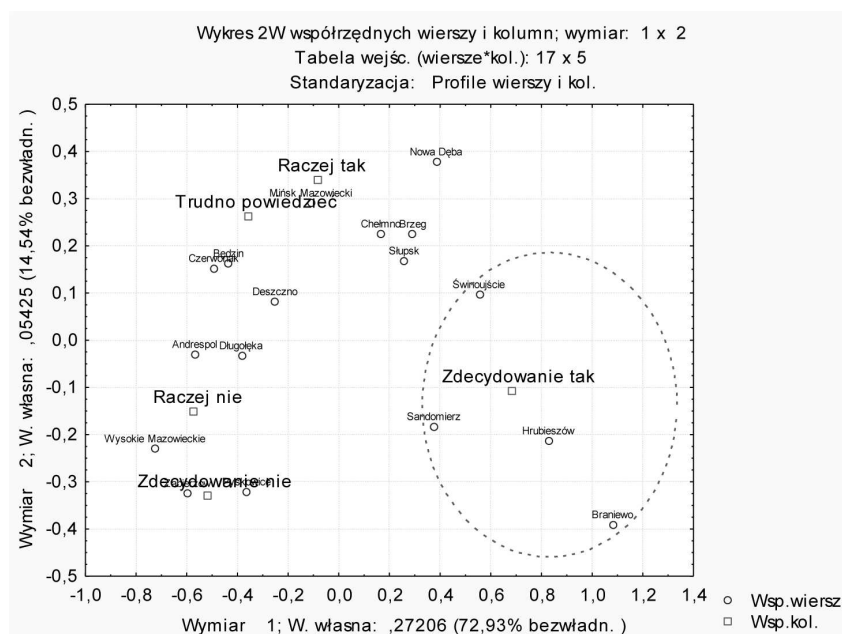


Wzrost UBwLLwWP – wzrost udziału bezrobotnych zarejestrowanych w liczbie ludności w wieku produkcyjnym (prawa oś OY)

Rysunek 4. Pogorszenie się sytuacji na rynku pracy poszczególnych gmin w wyniku likwidacji jednostki wojskowej (wariant)

Źródło: opracowanie własne na podstawie <http://www.stat.gov.pl/bdl/app/portret.dims> (dostęp 27.04.2012r.)

Radykalne pogorszenie się sytuacji na lokalnym rynku pracy, w wyniku likwidacji jednostki wojskowej dostrzegali głównie mieszkańcy gmin: Braniewo (88,0% badanych z tej gminy), Hrubieszów (74,0%), Świnoujście (56,0%), Sandomierz (52,0%) – rysunek 5. W pozostałych gminach odsetek takich odpowiedzi był mniejszy.



Rysunek 5. Ocena wpływu likwidacji jednostki wojskowej na lokalny rynek pracy

Źródło: opracowanie własne



Wyniki analizy korespondencji zaprezentowane na wykresach, interpretuje się na podstawie położenia punktów odzwierciedlających kategorie określonych zmiennych. Punkty znajdujące się daleko od centrum rzutowania, wnoszą największy wkład do odrzucenia hipotezy o niezależności zmiennych. Bliskie położenie punktów należących do różnych zmiennych, wskazuje na występowanie powiązań między kategoriami. Niedalekie położenie dwóch punktów należących do tej samej zmiennej, wskazuje na duże podobieństwo ich profili. Dla lepszego zobrazowania, położenie niektórych punktów wyróżniono za pomocą linii przerywanej. Zagadnienia te omówione zostały w pracach Stanimir (2005) i Stanisław (2007).

Na zasadnicze pogorszenie sytuacji na lokalnym rynku pracy zwracało uwagę od 38,0% do 44,0% respondentów z gmin: Chełmno, Słupsk, Brzeg, Nowa Dęba. Najniższy odsetek takich odpowiedzi odnotowano w gminach Andrespol i Czerwonak – po 8,0%.

Rolę jednostki wojskowej jako pracodawcy dla mieszkańców gminy dostrzegli również żołnierze. Stosunkowo duża grupa spośród nich (28,2%) oceniła, że obecność jednostki wojskowej na terenie gminy sprzyja powstawaniu nowych miejsc pracy lub utrzymaniu istniejących. Byli to głównie respondenci, których jednostki wojskowe stacjonują w gminach, w których możliwości zatrudnienia są ograniczone (gminy: Powidz, Cewice, Braniewo). W ocenie jednej trzeciej ankietowanych żołnierzy (33,8%) obecność jednostki wojskowej wzbogaca lokalny rynek pracy. Pozostali (39,3%) nie dostrzegli takiego oddziaływania lub ocenili ten wpływ na średnim poziomie (26,9%).

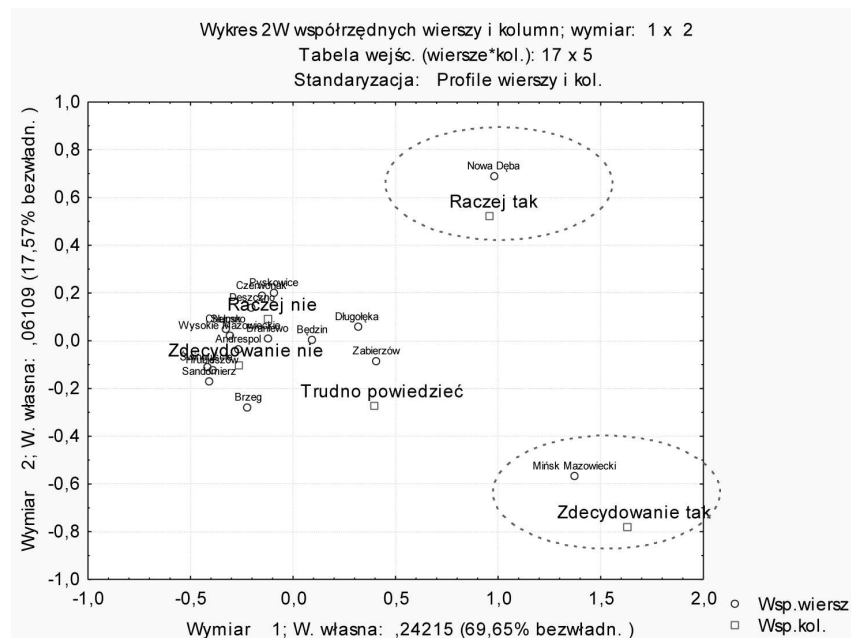
Obecność jednostki wojskowej – zdaniem mieszkańców – w różnym stopniu sprzyja rozwojowi lokalnych firm, głównie poprzez indywidualne zakupy dóbr i usług. W ich ocenie na obecności wojska w gminie korzystają głównie sklepy spożywcze (19,8% ankietowanych), piekarnie (18,3%), bary i restauracje (4,7%) i inne. Tak, więc z jednej strony utrata podmiotu (jednostki wojskowej), oznaczałaby kurczenie się zasobów wewnętrznych (procesy migracyjne, spadek siły nabywczej mieszkańców), a w konsekwencji zmniejszenie potencjału społeczno-gospodarczego gmin. Z drugiej strony, likwidacja jednostki wojskowej może dla gminy oznaczać korzyści. Tereny po jednostce mogą być przeznaczone na: powstanie stref gospodarczych oraz inwestycje mieszkaniowe i inne cele publiczne. Paradoksalnie oznaczałoby to, że likwidacja jednostki sprzyja wzrostowi potencjału gospodarczego i społecznego gminy.

Badani żołnierze oceniając oddziaływanie jednostki wojskowej na rozwój gminy, wskazywali na wpływy z podatków lokalnych do budżetu samorządu w postaci podatku od nieruchomości (częściej zwracali na to uwagę żołnierze pracujący w dużych jednostkach wojskowych). W ocenie respondentów dla niektórych firm obecność jednostki wojskowej jest okazją do sprzedaży swoich produktów (w ocenie co czwartego z nich – 25,8% oddziaływanie to jest istotne) lub świadczenia określonych usług (na wysoki stopień oddziaływania wskazało 20,8% ankietowanych). Żołnierze ocenili, że spośród lokalnych firm to głównie: stacje benzynowe (według 38,1% badanych), bary/puby (35,9%), samochodowe warsztaty naprawcze (32,1%), dyskoteki (29,2%)

i sklepy ogólnie – spożywcze (27,1%) są tymi, które w istotnym stopniu korzystają z obecności wojska.

Pewien obraz na rolę jednostki wojskowej jako partnera dla lokalnych firm można uzyskać analizując odpowiedzi mieszkańców gmin na temat tego czy podejmując decyzje biznesowe, biorą one pod uwagę fakt istnienia na swoim terenie jednostki wojskowej. Zarówno w gminach miejskich, jak i wiejskich odpowiedź „tak” wybrało 6,8% badanych.

Jednostki wojskowe są niekiedy zlokalizowane w centralnej części gminy, innym razem na jej obrzeżach. W ocenie dużej grupy badanych mieszkańców (65,2%) jednostka wojskowa w niewielkim stopniu lub w ogóle nie ogranicza wykorzystania terenów gminy na cele turystyczne. Najwięcej takich wskazań odnotowano wśród respondentów z gmin: Wysokie Mazowieckie (84,0%), Chełmno (78,0%), Hrubieszów (76,0%), Sandomierz (76,0%), Deszczno (72,0%). Odsetek odpowiedzi mieszkańców z gmin miejskich i wiejskich, wskazujących na brak oddziaływania jednostki wojskowej na ten element w gminie był podobny. Uczestniczący w badaniach żołnierze w większości także nie dostrzegali negatywnego wpływu jednostek wojskowych na funkcję turystyczną. Niewielka grupa ankietowanych mieszkańców gmin i niewielu żołnierzy oceniło, że jednostki wojskowe generują uciążliwy dla mieszkańców hałas. Na takie utrudnienia wskazywali mieszkający w pobliżu lotnisk (Mińsk Mazowiecki) lub poligonów (Nowa Dęba).



Rysunek 6. Ocena uciążliwości spowodowana hałasem generowanym przez jednostki

Źródło: opracowanie własne

Jednostki wojskowe stacjonują w nowoczesnych koszarowcach lub w starszych, niekiedy posiadających wartość zabytkową, budynkach. To sprawia, że mogą one wpływać na lokalny krajobraz zarówno upiększając go lub szpecąc. W ocenie większości badanych mieszkańców (68,8%) budynki te nie mają wpływu na lokalny krajobraz. Najwięcej takich wskazań odnotowano wśród mieszkańców gmin: Czerwonak (84,0% badanych) i Andrespol (78,0%). Stosunkowo niewielka grupa respondentów (19,8%) stwierdziła, że „upiększają” go, najczęściej byli to ankietowani z gmin miejskich Hrubieszów (50,0%), Sandomierz (36,0%), Chełmno (34,0%). Według innych (11,5%), budynki wojskowe szpecą krajobraz, w szczególności z gmin Zabierzów (30,2%), Długoleś (20,0%).

W ocenie ponad połowy ankietowanych mieszkańców gmin (52,4%) obecność jednostki wojskowej nie ma wpływu na wykorzystanie terenów w jej sąsiedztwie. Ograniczenia w tym zakresie zostały zauważone przez właścicieli firm, głównie z gmin: Słupsk (12,0%), Długoleś (12,0%) i Mińsk Mazowiecki (10,0%). Opinie żołnierzy dotyczące wpływu lokalizacji jednostek wojskowych na wykorzystanie terenów położonych w ich sąsiedztwie zasadniczo wskazywały na brak takiego oddziaływania.

Według 68,3% ankietowanych mieszkańców obecność jednostki wojskowej nie ma wpływu na rozwój funkcji mieszkaniowych w gminie. Zdecydowane ograniczenia w tym względzie były najczęściej wskazywane przez badanych z gmin: Andrespol (14,0%), Długoleś (14,0%), Słupsk (10,0%). Jeszcze większy odsetek 74,1% respondentów nie dostrzegał takiego wpływu na rozwój funkcji usługowych na terenie gminy. Zdecydowane utrudnienia były zauważane głównie przez badanych z gmin Długoleś (12,0%), Braniewo (8,0%), Andrespol (6,0%). Ankietowani, przeważnie kadra kierownicza, dostrzegali także, że majątek należący do wojska, nie jest należycie wykorzystywany.

Według badanych żołnierzy lokalizacja jednostki wojskowej w minimalnym stopniu ogranicza rozwój funkcji mieszkaniowej w gminie. Takie opinie wyraziło: 90,0% badanych z Zielonej Góry, 75,0% z dzielnicy Warszawa Bemowo, 76,9% z Ustki, 72,2% z Braniewa. Respondenci z innych gmin także dostrzegali niewielki wpływ jednostki na ten element funkcjonowania gminy.

W tabeli 1 przedstawiono zbiorczo oceny respondentów z gmin Braniewo, Chełmno i Mińsk Mazowiecki, w których badano zarówno mieszkańców i żołnierzy. W związku z tym, możliwe było porównanie ich opinii na wybrane kwestie związane z funkcjonowaniem jednostki wojskowej na terenie ich gminy.

Ogólnie ujmując oceny mieszkańców i żołnierzy mieszkających w gminach Braniewo, Chełmno i Mińsk Mazowiecki są podobne. Zarówno mieszkańcy, jak i żołnierze, podkreślili wpływ jednostki wojskowej na poczucie bezpieczeństwa mieszkańców oraz na lokalny rynek pracy. Można dostrzec także to, że żołnierze częściej zauważali wpływ jednostki wojskowej na realizację funkcji mieszkaniowej, usługowej oraz wykorzystanie terenów położonych w sąsiedztwie jednostki.

Tabela 1.

Zestawienie zbiorcze opinii mieszkańców i żołnierzy z gmin Braniewo, Chełmno, Mińsk Mazowiecki

| Pytanie dotyczące wpływu jednostki wojskowej na:        | Ankietowani | Odpowiedzi /%/                |      |             |             |
|---------------------------------------------------------|-------------|-------------------------------|------|-------------|-------------|
|                                                         |             | mały. .... b. duży<br>(-) (+) |      |             |             |
| poczucie bezpieczeństwa mieszkańców                     | mieszkańcy  | 6,7                           | 41,3 | <b>48,7</b> | 3,3         |
|                                                         | żołnierze   | 9,3                           | 11,6 | 23,2        | <b>55,9</b> |
| lokalny rynek pracy                                     | mieszkańcy  | 3,3                           | 15,3 | 24,0        | <b>57,4</b> |
|                                                         | żołnierze   | 13,2                          | 18,6 | 25,6        | <b>42,6</b> |
| wykorzystanie terenów w sąsiedztwie jednostki wojskowej | mieszkańcy  | <b>54,0</b>                   | 30,0 | 10,0        | 6,0         |
|                                                         | żołnierze   | <b>50,4</b>                   | 13,2 | 17,8        | 18,6        |
| wykorzystania terenów na cele turystyczne               | mieszkańcy  | <b>68,7</b>                   | 18,0 | 13,3        | –           |
|                                                         | żołnierze   | <b>57,4</b>                   | 12,4 | 13,9        | 16,3        |
| wykorzystania terenów na rozwój funkcji mieszkaniowych  | mieszkańcy  | <b>61,3</b>                   | 24,7 | 14,0        | –           |
|                                                         | żołnierze   | <b>60,5</b>                   | 8,5  | 13,9        | 17,1        |
| wykorzystania terenów na rozwój funkcji usługowych      | mieszkańcy  | <b>69,3</b>                   | 18,0 | 12,7        | –           |
|                                                         | żołnierze   | <b>55,0</b>                   | 14,7 | 13,9        | 16,4        |
| lokalny krajobraz                                       | mieszkańcy  | <b>68,0</b>                   | 24,0 | 7,3         | 0,7         |
|                                                         | żołnierze   | <b>60,5</b>                   | 17,0 | 10,1        | 12,4        |
| generowanie hałasu                                      | mieszkańcy  | <b>42,0</b>                   | 34,0 | 12,0        | 12,0        |
|                                                         | żołnierze   | <b>37,2</b>                   | 20,9 | 19,4        | 22,5        |

Źródło: opracowanie własne

## 4. PODSUMOWANIE

Podsumowując przedstawione wyniki badań można wysnuć następujące wnioski. Jednostka wojskowa w zróżnicowanym stopniu wpływa na funkcjonowanie lokalnej społeczności. Jej obecność na terenie gmin jest przez mieszkańców akceptowana. Poziom integracji jednostek wojskowych z lokalną społecznością jest stosunkowo niski. W wypowiedziach mieszkańców występował duży odsetek odpowiedzi neutralnych. W ich ocenie jednostka wojskowa nie wpływa w istotnym stopniu na rozwój gminy, ale też nie ogranicza jej funkcjonowania. Opinie badanych żołnierzy często były zbieżne z opiniami mieszkańców.

*Akademia Obrony Narodowej*

## LITERATURA

- [1] Aczel A.D., (2000), *Statystyka w zarządzaniu*, PWN, Warszawa.
- [2] Cieślarczyk M., (2009), *Teoretyczne i metodologiczne podstawy badania problemów bezpieczeństwa i obronności państwa*, AP, Siedlce.
- [3] Feliksiak M., (2010), *Oceny instytucji publicznych*, CBOS, Warszawa.
- [4] Feliksiak M., (2012), *Oceny instytucji publicznych*, CBOS, Warszawa.
- [5] Jewtuchowicz A., (2005), *Terytorium i współczesne dylematy jego rozwoju*, UŁ, Łódź.
- [6] Panek T., (2007), *Statystyka społeczna*, PWE, Warszawa.
- [7] Piotrowska-Trybull M., (2011), *Rozwój miasta Daugavpils w perspektywie militarnej i gospodarczej*, Zeszyty Naukowe Akademii Obrony Narodowej, Nr 3 (84), Warszawa.
- [8] Sirko S., (2011), *Żołnierze i pracownicy wojska na współczesnym rynku pracy* /w:/ Byłok F., Słocińska A., (red.), *Współczesne oblicza kapitału ludzkiego i intelektualnego*, PCz, Częstochowa.
- [9] Sokołowski A., (2010), *Estymacja i testowanie hipotez* /w:/ Statistica w badaniach naukowych i nauczaniu statystyki, StatSoft, Kraków.
- [10] Stanimir A., (2005), *Analiza korespondencji jako narzędzie do badania zjawisk ekonomicznych*, AE, Wrocław.
- [11] Stanisław A., (2007), *Przystępny kurs statystyki t.3. Analizy wielowymiarowe*, StatSoft, Kraków.
- [12] Szreder M., (2010), *Metody i techniki sondażowych badań opinii*, PWE, Warszawa.

## WPŁYW JEDNOSTKI WOJSKOWEJ NA ROZWÓJ GMINY

## Streszczenie

Rozwój poszczególnych gmin jest warunkowany wewnątrznie i zewnątrznie. Stacjonujące na terenie gmin jednostki wojskowe, w zróżnicowanym stopniu, wpływają na poszczególne sfery życia mieszkańców. Ich lokalizacja, na terenie poszczególnych gmin, determinuje szereg powiązań z otoczeniem lokalnym. Powiązania mają charakter dwustronny. Jednostki czerpią z otoczenia zasoby (ludzkie, finansowe, rzeczowe i informacyjne) w oparciu, o które realizują przypisane im zadania. Natomiast do otoczenia dostarczają dobro publiczne, jakim jest bezpieczeństwo obywateli oraz generują do otoczenia efekty wynikające z użytkowania zasobów. W artykule przedstawiono wyniki badań przeprowadzonych wśród mieszkańców gmin oraz wśród żołnierzy jednostek, które stacjonują w różnych regionach na terenie kraju.

**Słowa kluczowe:** jednostka wojskowa, samorząd terytorialny, rozwój lokalny, wpływ, powiązania

## THE INFLUENCE OF THE PRESENCE OF A MILITARY UNIT ON THE DEVELOPMENT OF A GMINA

## Abstract

Development of particular communities is conditioned by internal and external factors. Military units which are stationed on the territory of communities, influence the specific spheres of inhabitants life, in

different ways. Their localization on the territory determines a number of ties with local surrounding. These ties have bilateral character. Military units acquire from local surrounding resources (human, financial, material, information), which are the basis to fulfil their tasks and they transfer public good like safety of citizens and, what is more, they generate other outcomes as a result of uses of resources. The article presents outcomes of research carried out among inhabitants of communities and soldiers of military units, which are located in different regions on the territory of Poland.

**Key words:** military units, local authority, local development, influence, connections

## SPIS TREŚCI

### Numer Specjalny 1

|                                                                                                                                                                         |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Od Redakcji .....                                                                                                                                                       | 4   |
| Kazimierz Kruszk a – Polskie Towarzystwo Statystyczne (1912-2012) .....                                                                                                 | 5   |
| Elżbieta Gołata – Kongres Statystyki Polskiej z okazji – 100-lecia Polskiego Towarzystwa<br>Statystycznego, Poznań, 18-20 kwietnia 2012 r. ....                         | 17  |
| Czesław Domański – Laureaci Medalu im. Jerzego Sławy-Neymana .....                                                                                                      | 23  |
| Uchwała Kongresu Statystyki Polskiej, 20 kwietnia 2012 r. ....                                                                                                          | 31  |
| Walenty Ostasiewicz – Rozwój myśli statystycznej w Polsce w XIX wieku .....                                                                                             | 33  |
| Tadeusz Caliński – Rozwój i osiągnięcia w biometrii polskiej .....                                                                                                      | 47  |
| Krzysztof Jajuga – Rozwój polskiej myśli statystycznej w naukach ekonomicznych .....                                                                                    | 53  |
| Jan Kordos – Współzależność pomiędzy rozwojem teorii i praktyki badań reprezentacyjnych<br>w Polsce .....                                                               | 61  |
| Daniel Kosirowski – Głębia położenia-rozrzutu w strumieniowej analizie danych ekonomicznych .....                                                                       | 87  |
| Jan W. Owsiński – On the optimal division of an empirical distribution (and some related<br>problems) .....                                                             | 109 |
| Wioletta Grzenda – Badanie determinant pozostawiania bez pracy osób młodych z wykorzystaniem<br>semiparametrycznego modelu Coxa .....                                   | 123 |
| Maria Szumuksta-Zawadzka, Jan Zawadzki – Z badań nad metodami prognozowania na podstawie<br>niekompletnych szeregów czasowych z wahaniami okresowymi (sezonowymi) ..... | 140 |
| Tomasz Klimanek – Using indirect estimation with spatial autocorrelation in social surveys<br>in Poland .....                                                           | 155 |
| Andrzej Czyżewski, Aleksander Grzelak – Możliwości wykorzystania statystyki bilansów<br>przepływów międzygałęziowych dla makroekonomicznych ocen w gospodarce .....     | 173 |

### Numer Specjalny 2

|                                                                                                                                                                                                        |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Jan Paradysz – Statystyka regionalna: stan, problemy i kierunki rozwoju .....                                                                                                                          | 191 |
| Wojciech Roszka – System statystyki publicznej oparty na zintegrowanych źródłach danych .....                                                                                                          | 205 |
| Jacek Kowalewski, Monika Natkowska – Wykorzystanie mapy statystyki krótkookresowej w procesie<br>organizacji badań przedsiębiorstw prowadzonych przez statystykę publiczną ...                         | 222 |
| Magdalena Okupniak – Zastosowanie analizy blokowej na przykładzie badania zróżnicowania<br>dochodów gospodarstw domowych .....                                                                         | 237 |
| Łukasz Wawrowski – Analiza ubóstwa w przekroju powiatów w województwie wielkopolskim z<br>wykorzystaniem metod statystyki małych obszarów .....                                                        | 248 |
| Roma Ryś-Jurek – Zastosowanie hierarchicznej klasyfikacji aglomeracyjnej do grupowania<br>krajów Unii Europejskiej ze względu na strukturę i skalę produkcji gospodarstw rolnych .....                 | 261 |
| Hanna Gruchociak – Delimitacja lokalnych rynków pracy w Polsce .....                                                                                                                                   | 277 |
| Aleksandra Łuczak, Feliks Wysocki – Zastosowanie uogólnionej miary odległości GDM oraz<br>metody TOPSIS do oceny poziomu rozwoju społeczno-gospodarczego powiatów województwa<br>wielkopolskiego ..... | 298 |
| Iwona Bąk – Zastosowanie wybranych metod statystyczno-ekonometrycznych w badaniu aktywności<br>turystycznej seniorów w Polsce .....                                                                    | 312 |
| Marzena Piotrowska-Trybull, Stanisław Sirko – Wpływ jednostki wojaskowej na rozwój gminy .....                                                                                                         | 333 |

## CONTENTS

### Special issue 1

|                                                                                                                                                                   |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| From the Editor .....                                                                                                                                             | 4   |
| Kazimierz K r u s z k a – The Polish Statistical Association (1912–2012) .....                                                                                    | 5   |
| Elżbieta G o ł a t a – The Congress of Polish Statistics to Celebrate the 100th Anniversary of the Polish Statistical Association, Poznan, 18-20 April 2012 ..... | 17  |
| Czesław D o m a ń s k i – Winners of the Jerzy S p ł a w a -Neyman Medal .....                                                                                    | 23  |
| A resolution of the Congress of Polish Statistics, 20 April 2012 .....                                                                                            | 31  |
| Walenty O s t a s i e w i c z – The Development of Statistical Thought in Poland in 19 <sup>th</sup> c. ....                                                      | 33  |
| Tadeusz C a l i ń s k i – The Development and Achievements in Polish Biometry .....                                                                               | 47  |
| Krzysztof J a j u g a – The Development of Polish Statistics in Economic Sciences .....                                                                           | 53  |
| Jan K o r d o s – The Interplay between Sample Survey Theory and Practice in Poland .....                                                                         | 61  |
| Daniel K o s i o r o w s k i – Location-Scale Depth in Economic Data Stream Analysis .....                                                                        | 87  |
| Jan W. O w s i ń s k i – On the optimal division of an empirical distribution (and some related problems) .....                                                   | 109 |
| Wioletta G r z e n d a – An Analysis of Unemployment Duration Determinants among Young People Using Semiparametric Cox Model .....                                | 123 |
| Maria S z m u k s t a - Z a w a d z k a, Jan Z a w a d z k i – Studies of Methods Applied to Forecasting Incomplete Data in Seasonal Time Series .....            | 140 |
| Tomasz K l i m a n e k – Using indirect estimation with spatial autocorrelation in social surveys in Poland .....                                                 | 155 |
| Andrzej C z y ż e w s k i, Aleksander G r z e l a k – Possibilities of Using the Statistic of Input-Output in Macroeconomics Evaluation of the Economy .....      | 173 |

### Special issue 2

|                                                                                                                                                                                                           |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Jan P a r a d y s z – Regional Statistics: Current Status, Problems and Development Directions ..                                                                                                         | 191 |
| Wojciech R o s z k a – The System of Public Statistics Based on Integrated Data Sources .....                                                                                                             | 205 |
| Jacek K o w a l e w s k i, Monika N a t k o w s k a – Using a Map of Short Term Statistics in the Process of Organizing Business Surveys Conducted by Public Statistics .....                             | 222 |
| Magdalena O k u p n i a k – Application of the Block Analysis to the Investigation of Household Income Diversity .....                                                                                    | 237 |
| Łukasz W a w r o w s k i – Poverty Analysis in Districts of Wielkopolska with Small Area Estimation Methods .....                                                                                         | 248 |
| Roma R y ś - J u r e k – The Use of the Agglomeration Hierarchical Classification to Group the EU Countries according to the Structure and Scale of Farms' Production .....                               | 261 |
| Hanna G r u c h o c i a k – Delimitation of Local Labor Markets in Poland .....                                                                                                                           | 277 |
| Aleksandra Ł u c z a k, Feliks W y s o c k i – Application of Generalised Distance Measure and TOPSIS Method to Evaluate the Level of Socioeconomic Development of Powiats in Wielkopolska Province ..... | 298 |
| Iwona B ą k – Application of Selected Statistical and Econometric Methods in the Analysis of Senior Tourism Activities in Poland .....                                                                    | 312 |
| Marzena P i o t r o w s k a - T r y b u l l, Stanisław S i r k o – The Influence of the presence of a Military Unit on the Development of a Gmina .....                                                   | 333 |