РЕЗЮМЕ

Темой статьи является генезис, деятельность и финансовые результаты фирм кредитного посредничества. Этот сегмент финансового рынка определился в Польше в половине девяностых лет XX века и с этого момента очень быстро развивается. На результаты обследуемых фирм в 2007 и 2008 гг положительно повлияли хорошая конъюнктура и высокий спрос на банковские продукты в этот период. Из проведенного анализа кредитной политики банков вытекает, что текущий год будет труднее для кредитных посредников. Главным образом это касается фирм, которые только начинают деятельность, или планируют ее расширение. Некоторые банки из опасения перед понижением качества портфеля активов, потеряли заинтересованность сотрудничеством с кредитными брокерами. Можно поставить гипотез, что шансы развития будут иметь не только, наиболее богатые узко специализированные брокерские компании, но прежде всего наиболее универсальные фирмы, сотрудничающие со многими партнерами, предлагающими большое разнообразие финансовых услуг.

Andrzej MŁODAK

Zróżnicowanie kapitału ludzkiego na rynku pracy

Jednym z podstawowych zadań, jakie postawiła przed sobą kierowana przez prof. dra hab. Jana Paradysza podgrupa do spraw metod statystyczno-matematycznych, funkcjonująca w ramach przygotowań do Powszechnego Spisu Rolnego 2010 oraz Narodowego Spisu Powszechnego Ludności i Mieszkań 2011, jest ocena jakości spisu powszechnego przeprowadzonego w 2002 r. Innym zadaniem jest badanie stabilności modeli ekonometrycznych w czasie i w przestrzeni z wykorzystaniem taksonometrycznej analizy podobieństwa regionalnego. Działania te mają na celu wyeliminowanie ułomności metodologicznych oraz doskonalenie konstrukcji efektywnych metod estymacji dla małych obszarów w planowanych spisach. Istotnym elementem tej strategii stało się przeprowadzenie analizy taksonomicznej danych ze spisu w 2002 r. Służy ona ulepszaniu jakości statystyki małych obszarów z wykorzystaniem dostępnych metod weryfikacyjnych oraz konstrukcji mierników kompleksowych i skupień obiektów podobnych na różnych poziomach agregacji danych.

Przedmiotem przeprowadzonego badania[1] była ocena przestrzennego zróżnicowania gmin i powiatów pod względem wartości kapitału ludzkiego w województwach wielkopolskim i mazowieckim. To zjawisko zalicza się do fundamentalnych czynników kształtujących rozwój społeczno-gospodarczy każdego regionu. Biorąc pod uwagę konieczność uzyskania odpowiedniego poziomu przestrzennej porównywalności informacji oraz dostępność danych, przedmiotem analizy obrano 15 cech o charakterze wskaźnikowym. Oto pełne zestawienie badanych cech:

X_1 — udział potencjalnego ludzkiego kapitału rynku pracy w ogólnej liczbie ludności faktycznie zamieszkałej i przybyłej z zagranicy na pobyt czasowy (w %);
X_2 — średnia odwrotność zagęszczenia mieszkań, w których mieszkają badane osoby;
X_3 — udział liczby osób w wieku produkcyjnym mobilnym w liczbie osób w wieku produkcyjnym (w %);
X_4 — liczba osób w wieku produkcyjnym przypadających na 100 osób w wieku nieprodukcyjnym;
X_5 — odsetek osób z wyższym wykształceniem w wieku 13 lat i więcej;
X_6 — odsetek osób z wykształceniem średnim w wieku 13 lat i więcej;
X_7 — odsetek osób kontynuujących naukę w wieku 13 lat i więcej;
X_8 — odsetek osób nieposiadających ograniczeń do wykonywania podstawowych czynności;
X_9 — odsetek osób bez orzeczonej niepełnosprawności;
X_{10} — odsetek osób, dla których przynajmniej jednym z języków domowych jest język polski;
X_{11} — udział pracujących w liczbie osób w wieku 15 lat i więcej (w %);
X_{12} — odsetek pracujących wykonujących pracę dodatkową;
X_{13} — osoby aktywne w poszukiwaniu pracy w % osób niepracujących w wieku 15 lat i więcej;
X_{14} — osoby gotowe podjąć pracę w % osób niepracujących, aktywnie poszukujących pracy w wieku 15 lat i więcej;
X_{15} — odsetek osób dorosłych z obciążeniem dyspozycyjności.

Główną przesłanką, którą brano pod uwagę przy doborze tych cech, był wpływ danego czynnika na przydatność osoby z punktu widzenia potrzeb rynku pracy. Jest rzeczą zrozumiałą, że osoby dobrze wykształcone, sprawne fizycznie, aktywne w rozwoju zawodowym lub poszukiwaniu pracy i samokształceniu oraz z minimalnym obciążeniem dyspozycyjności (rozumianym tutaj jako posiadanie osób pod opieką, głównie dzieci) są bardziej atrakcyjne dla potencjalnych pracodawców. Dlatego też musiało to znaleźć odzwierciedlenie w przeprowadzo-

[1] Serdecznie dziękuję moim Kolegom z Urzędu Statystycznego w Poznaniu: mgr Beacie Kraszewskiej, mgrowi Arkadiuszowi Kowalczykowi i mgrowi Pawłowi Lańduchowi (Oddział w Kaliszu) oraz mgrowi Tomaszowi Józefowskiemu (Ośrodek Statystyki Małych Obszarów) i mgrowi Maciejowi Kaźmierczakowi (Centrum Statystyki Miast) za pomoc w realizacji badania.

nym dociekaniu. Warto też zauważyć, że spis powszechny dostarcza wielu ważnych informacji z tej dziedziny, nieosiągalnych (szczególnie na niskich poziomach przestrzennych) w innych badaniach statystycznych o podobnej tematyce.

Wymienione cechy stanowiły przedmiot analizy taksonomicznej polegającej na konstrukcji syntetycznego miernika rozwoju oraz typologii rozpatrywanych jednostek przestrzennych. Zgodnie ze schematem opisanym w literaturze (Sobczyk, 1995; Śmiłowska, 1997; Młodak, 2006 a), analiza ta składa się z następujących etapów:

1) weryfikacja zmiennościowa cech,
2) weryfikacja korelacyjna cech,
3) stymulacja i normalizacja cech,
4) definicja wzorca rozwojowego,
5) definicja miary odległości obiektów od wzorca,
6) konstrukcja miernika kompleksowego,
7) grupowanie obiektów pod względem podobieństwa.

OCENA JAKOŚCI ZGROMADZONYCH DANYCH

Zanim przystąpimy do zasadniczej części analizy, konieczne jest dokonanie oceny rozkładów poszczególnych cech, co ułatwi uchwycenie prawidłowości w nich występujących oraz wpływu ewentualnych obserwacji odstających na postać i właściwości rozpatrywanego modelu wielowymiarowego.

Tabl. 1 zawiera podstawową charakterystykę statystyczną badanych cech dla gmin i powiatów województw wielkopolskiego i mazowieckiego: średnią arytmetyczną, medianę i współczynnik zmienności wyrażony w %. Ten ostatni wskaźnik będzie miał także istotne znaczenie w późniejszym procesie weryfikacyjnym. Zamieszczamy również trzy wykresy obrazujące wartości wskaźnika asymetrii, stosunku odległości międzykwartylowej (różnicy pomiędzy trzecim a pierwszym kwartylem) do rozstępu (różnicy między maksymalną a minimalną wartością danej cechy) oraz kurtozy (odzwierciedlającej poziom spłaszczenia rozkładu wartości danej cechy w stosunku do rozkładu normalnego).

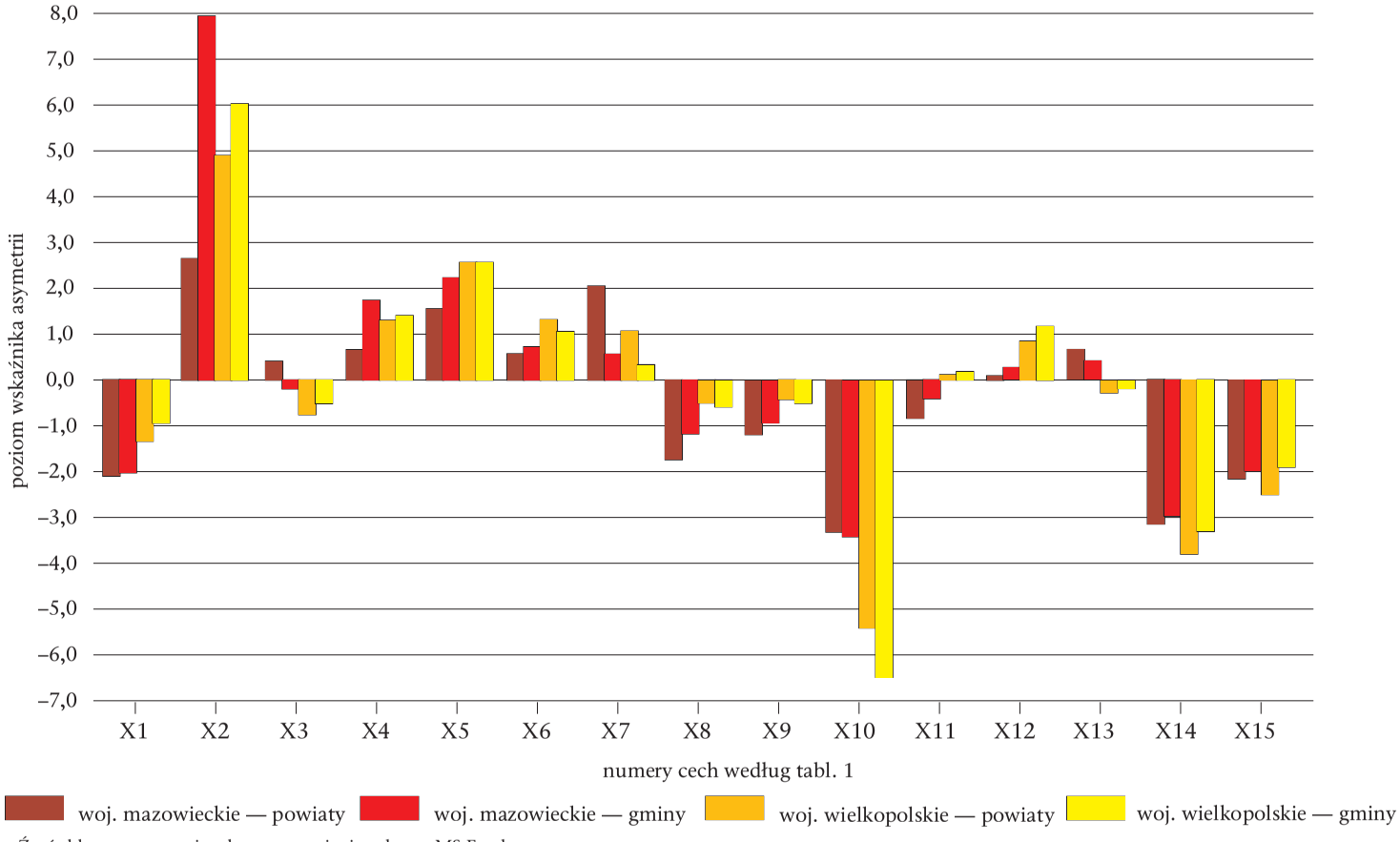
Analizując uzyskane rezultaty dochodzimy do ciekawych konkluzji. Różnice pomiędzy średnimi arytmetycznymi a medianami są raczej niewielkie. Sugeruje to brak występowania istotnych obserwacji odstających. Rozkłady cech w większości wykazują asymetrię, choć jej natężenie i kierunki są różne (czasem nawet dla tego samego województwa — np. cecha X_3 — wykr. 1). Zmienność rozpatrywanych cech okazuje się natomiast bardzo zróżnicowana, zarówno w ramach jednego województwa (statystyka dla gmin często różni się istotnie od statystyki powiatów) jak i pomiędzy województwami. W przypadku powiatów zmienność ta jest generalnie niższa niż dla gmin i tylko w niewielkim stopniu można tłumaczyć to istotną różnicą w liczbie obiektów (331 gmin w woj. mazowieckim i 230 w woj. wielkopolskim, podczas gdy powiatów jest odpowiednio 42 i 35). Wynika stąd, że zróżnicowanie przestrzenne nie musi być hierarchicznie dziedziczne, tzn. zmienność na niższym poziomie agregacji niekoniecznie znajduje swe odzwierciedlenie w zmienności na poziomie wyższym.

TABL. 1. WARTOŚCI STATYSTYKI OPISOWEJ DLA BADANYCH CECH

Zmienne	Woj. mazowieckie						Woj. wielkopolskie					
	gminy			powiaty			gminy			powiaty		
	średnia arytmetyczna	mediana	współczynnik zmienności	średnia arytmetyczna	mediana	współczynnik zmienności	średnia arytmetyczna	mediana	współczynnik zmienności	średnia arytmetyczna	mediana	współczynnik zmienności
X_1	94,465	95,031	2,640	94,234	94,865	2,055	95,833	96,005	1,349	95,574	95,861	1,325
X_2	2,061	1,399	127,291	2,142	1,747	55,083	1,753	1,248	98,852	1,904	1,558	74,334
X_3	65,676	65,803	4,174	64,992	65,090	2,796	66,717	66,899	2,969	65,618	65,830	2,205
X_4	138,726	131,295	19,138	148,282	142,306	14,298	151,635	148,470	10,482	159,210	156,437	7,837
X_5	6,378	3,833	91,328	7,982	6,120	54,380	5,293	4,028	65,334	6,787	5,792	45,953
X_6	21,850	19,374	35,193	25,358	23,578	22,803	21,372	20,077	25,308	24,751	24,006	18,619
X_7	18,707	18,412	12,034	19,762	19,494	8,699	19,096	18,941	7,912	19,481	19,290	4,780
X_8	89,940	90,487	3,051	90,174	90,406	2,063	89,500	89,664	2,215	89,595	89,719	1,232
X_9	89,683	90,728	3,863	89,624	90,101	3,335	86,786	86,973	3,601	86,127	86,486	3,393
X_{10}	99,026	99,649	1,827	99,020	99,426	1,359	99,799	99,949	0,544	99,796	99,903	0,469
X_{11}	49,286	49,357	10,298	47,688	48,185	8,098	47,242	47,160	9,444	46,590	46,309	6,966
X_{12}	12,316	12,303	39,501	10,709	10,682	32,595	9,885	9,038	39,363	8,739	8,282	29,666
X_{13}	19,850	19,438	25,483	21,825	22,096	16,759	21,224	21,484	16,184	21,472	21,970	12,725
X_{14}	91,390	95,238	12,128	91,468	95,120	9,349	92,685	94,004	5,868	92,033	92,750	6,058
X_{15}	59,731	60,142	8,419	59,947	60,537	5,446	65,357	66,020	5,654	64,292	64,948	5,668

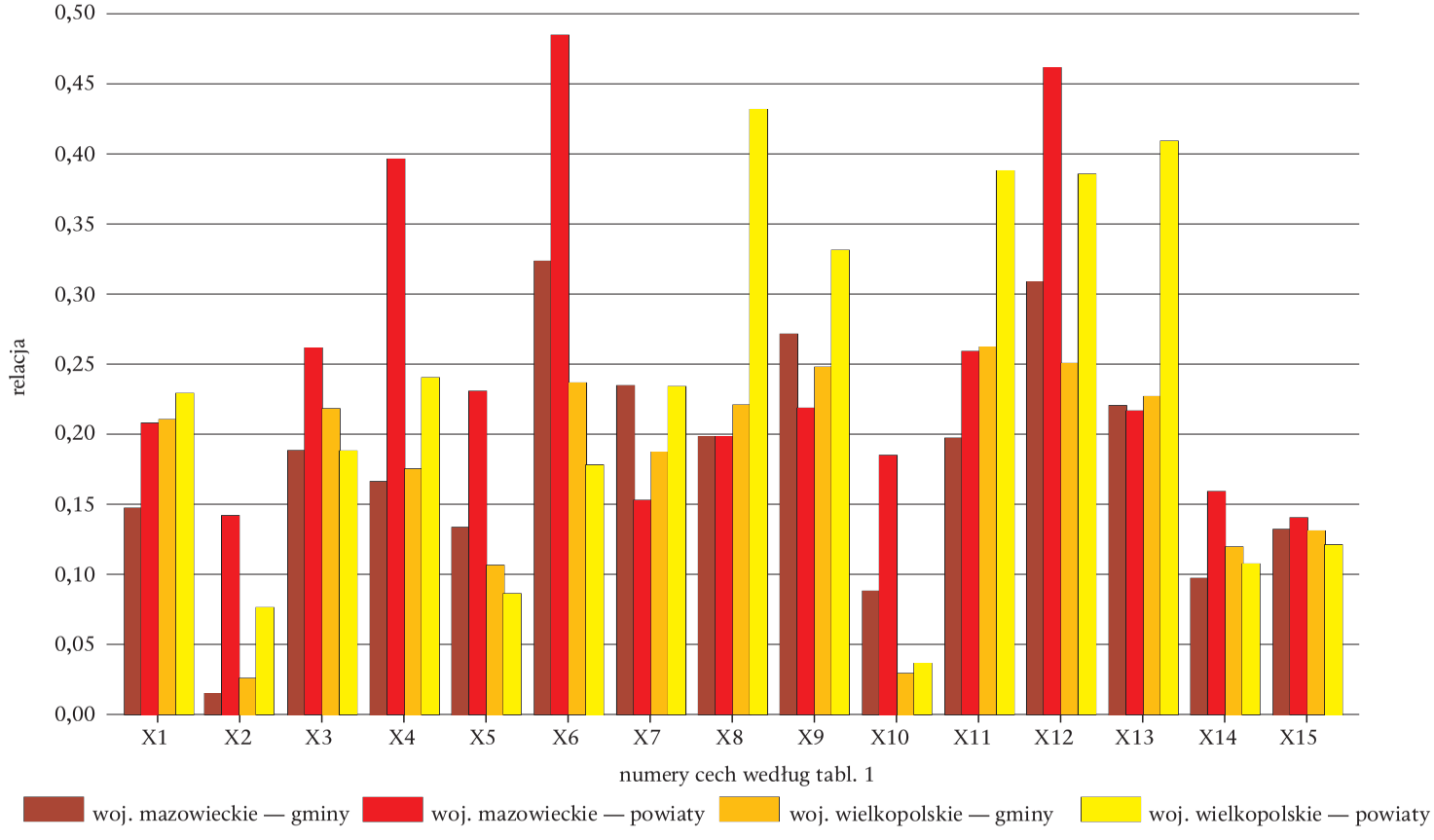
Ź r ó d ł o: opracowanie własne przy użyciu pakietu SAS Enterprise Guide 4.1.

Wykr. 1. WSKAŹNIK ASYMETRII (skośności)



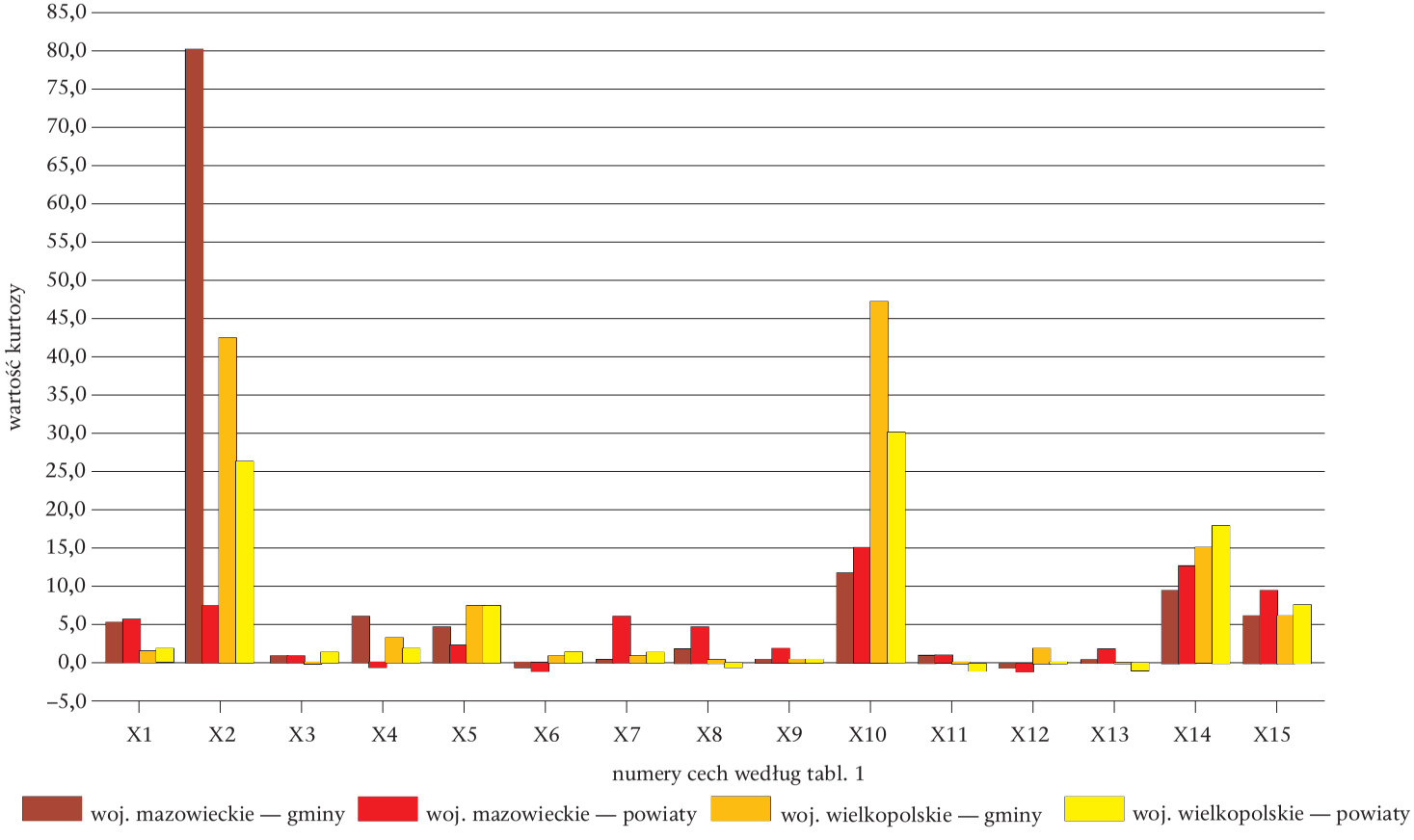
Ź r ó d ł o: opracowanie własne przy użyciu arkusza MS Excel.

Wykr. 2. RELACJA ODLEGŁOŚCI MIĘDZYKWARTYLOWEJ DO ROZSTĘPU



Ź r ó d ł o: opracowanie własne przy użyciu arkusza MS Excel.

Wykr. 3. KURTOZA



Ź r ó d ł o: opracowanie własne przy użyciu arkusza MS Excel.

Na wykr. 2 pokazano kształtowanie się relacji odległości międzykwartylowej do rozstępu. Wskazuje ona, czy źródeł zmienności — w przypadku asymetrii — należy szukać tylko w występowaniu pojedynczych obserwacji odstających (wartości niskie, szczególnie bliskie zera) czy też w ogólnym rozproszeniu wszystkich wartości (wskaźnik bliski 1). W naszym modelu ogólna dyspersja wartości widoczna jest szczególnie w odniesieniu do cech X_6 i X_{12} dla powiatów Mazowsza. W większości przypadków o zmienności decydują jednak obserwacje odstające. Wykr. 3 uwidacznia z kolei kształtowanie się kurtozy, czyli wskaźnika spiczastości (spłaszczenia) gęstości rozkładu. Szczególną wysmukłość w porównaniu z rozkładem normalnym wykazują cechy X_2, X_{10}, X_{14} i X_{15}. Kształt rozkładu pozostałych zmiennych jest zróżnicowany w zależności od poziomu przestrzennego, aczkolwiek na ogół okazuje się być on zbliżony do krzywej Gaussa. Czasem występują niewielkie spłaszczenia (np. cecha X_4 dla powiatów woj. mazowieckiego odróżnia się wyraźnie od pozostałych — średnio leptokurtycznych — ujemną wartością rzeczonego wskaźnika). Nie ma natomiast — co ciekawe — cechy, która w każdym przypadku byłaby platykurtyczna.

WERYFIKACJA CECH

Pierwszy zasadniczy etap badawczy stanowi weryfikacja zmiennościowa zestawu cech. Dokonuje się tutaj eliminacji tych spośród nich, które wykazują zbyt niskie zróżnicowanie — z taksonomicznego punktu widzenia stanowią one bowiem niewielką wartość analityczną. W tym celu ustalamy, dla których spośród badanych cech współczynnik zmienności przyjmuje wartości poniżej arbitralnie ustalonego progu. Zazwyczaj — i tak będzie również w naszym przypadku — przyjmuje się go na poziomie 10%. Opisane obserwacje skłaniają nas do stwierdzenia, że specyfika określonych agregacji przestrzennych powoduje, że mimo identycznej wyjściowej kolekcji zmiennych, dla każdej z nich zbiór cech finalnie poddawanych analizie może być odmienny. Eliminując zatem cechy o wartości współczynnika zmienności poniżej 10% otrzymujemy cztery zestawy:

1) woj. mazowieckie (gminy): X_2, X_4, X_5, X_6, X_7, X_{11}, X_{12}, X_{13}, X_{14};
2) woj. mazowieckie (powiaty): X_2, X_4, X_5, X_6, X_{12}, X_{13};
3) woj. wielkopolskie (gminy): X_2, X_4, X_5, X_6, X_{12}, X_{13};
4) woj. wielkopolskie (powiaty): X_2, X_5, X_6, X_{12}, X_{13}.

Zestaw cech dla powiatów Mazowsza jest zatem taki sam, jak dla gmin Wielkopolski. Istnieje ponadto coś na kształt „żelaznej grupy", czyli cech, które występują w każdym z tych zestawów. Należą tu cechy: X_2, X_5, X_6, X_{12} oraz X_{13}.

Każdy ze wskazanych zestawów został następnie poddany weryfikacji korelacyjnej. Jej cel stanowi ustalenie, które cechy wykazują nadmierne skorelowanie z innymi (i stanowią tym samym nośnik podobnej informacji). Ważną rzeczą jest także spełnienie postulatu, aby owa ocena uwzględniała wszystkie właściwości modelu, czyli reprezentowała podejście kompleksowe. W tym celu wykorzystamy metodę odwróconej macierzy korelacji (Malina, Zeliaś, 1998; Lira i in., 2002; Młodak, 2006 a).

Dla każdego z rozpatrywanych zestawów wyznaczamy zatem macierz współczynników korelacji Pearsona, a następnie macierz do niej odwrotną. Diagonalne elementy tej macierzy należą do przedziału $[1,\infty)$. Jeśli wartość któregoś z nich przekracza arbitralnie ustalony próg (w tym przypadku 10), to oznacza to, że macierz jest wadliwie uwarunkowana numerycznie. Innymi słowy, odpowiadająca jej cecha okazuje się nadmiernie skorelowana z innymi. Można więc ją usunąć. Jeśli takich niepokojących wartości diagonalnych występuje więcej, to należy dokładnie przeanalizować korelację pomiędzy poszczególnymi „podejrzanymi" cechami, a następnie dokonać takiej eliminacji, aby była ona jak najskromniejsza i jednocześnie zapewniła odpowiednie nieskorelowanie pozostałych cech. Weryfikacja korelacyjna wymusza więc niejako zastosowanie pewnej uznaniowości w podejmowaniu decyzji dyskryminacyjnej.

W wyniku podjętych czynności otrzymano ostatecznie zestawy cech, które staną się podstawą analizy taksonomicznej. Noszą one nazwę *cech diagnostycznych*. Oto one:

1) woj. mazowieckie (gminy): X_2, X_4, X_5, X_6, X_7, X_{11}, X_{12}, X_{13}, X_{14};
2) woj. mazowieckie (powiaty): X_2, X_4, X_{12}, X_{13};
3) woj. wielkopolskie (gminy): X_2, X_4, X_5, X_6, X_{12}, X_{13};
4) woj. wielkopolskie (powiaty): X_2, X_6, X_{12}, X_{13}.

Warto nadmienić, że redukcji dokonano jedynie w przypadku modeli powiatowych. Dla woj. mazowieckiego konieczne okazało się usunięcie cech X_5 oraz X_6, gdyż odpowiadające im wartości diagonalne odwróconej macierzy korelacji wyraźnie przekraczały graniczną wartość 10 (wynosząc odpowiednio: 17,47014 i 16,75633). Analiza danych dla powiatów woj. wielkopolskiego wykazała, że próg dopuszczalnej korelacji z innymi cechami przekroczyła cecha X_5 (odpowiednia wartość diagonalna odwróconej macierzy korelacji wyniosła 12,0014). Dlatego też podjęto decyzję o jej eliminacji.

KONSTRUKCJA MIERNIKA KOMPLEKSOWEGO I GRUPOWANIE OBIEKTÓW

Zauważmy, że wszystkie cechy diagnostyczne są stymulantami, tzn. mają tę właściwość, że wyższa wartość świadczy o lepszej sytuacji danego obszaru. Dzięki temu dodatkowa czynność zamiany cech o odmiennych własnościach na stymulanty (czyli stymulacja) nie była tu konieczna[2]. Kolejny istotny krok analizy taksonomicznej stanowi zatem ich normalizacja, mająca na celu ujednolicenie mian. Skorzystamy tutaj z mediany Webera (wektora minimalizującego sumę odległości euklidesowych od punktów obrazujących dane obiekty), ale w wersji uciętej. Oznacza to, że nie bierzemy pod uwagę wszystkich odległości, ale tylko pewną liczbę najmniejszych spośród nich (np. Vandev, 2002).

Ujmując rzecz formalnie, niech n i m będą liczbami naturalnymi, przy czym n oznacza liczbę obiektów, zaś m — liczbę cech diagnostycznych. Oznaczmy przez x_{ij} obserwację cechy X_j dla i-tego obiektu, $i = 1, 2, \ldots, n$,

[2] Postulat ten uwzględniono już w definicji cech.

$j = 1, 2, \ldots, m$. Wówczas każdy obiekt reprezentowany jest przez punkt $\gamma_i = (x_{i1}, x_{i2}, ..., x_{im}) \in \Re^m$, $i = 1, 2, \ldots, n$. Dla dowolnego punktu $y = (y_1, y_2, ..., y_m) \in \Re^m$ oznaczmy jego odległość euklidesową od γ_i przez $\|\gamma_i - y\| = \sqrt{\sum_{j=1}^{m} (x_{ij} - y_j)^2}$, $i = 1, 2, \ldots, n$. Niech $\gamma_{(1)}, \gamma_{(2)}, ..., \gamma_{(n)}$ oznacza takie uporządkowanie punktów $\gamma_1, \gamma_2, ..., \gamma_n$, że $\|\gamma_{(1)} - y\| \le \|\gamma_{(2)} - y\| \le \ldots \le \|\gamma_{(n)} - y\|$. Przyjmijmy, że p jest liczbą naturalną taką, że $1 \le p \le n$. Wówczas *uciętą medianą Webera* na poziomie p nazywamy wektor $\theta = (\theta_1, \theta_2, ..., \theta_m) \in \Re^m$, który spełnia następującą równość:

$$\sum_{i=1}^{p} \|\gamma_{(i)} - \theta\| = \min_{y \in \Re^m} \sum_{i=1}^{p} \|\gamma_{(i)} - y\|$$

Głównym celem wprowadzenia tej modyfikacji (oprócz wykorzystania wzajemnych powiązań pomiędzy cechami) jest zniwelowanie często spotykanego problemu występowania nielicznych obiektów w naturalny sposób różniących się istotnie i kompleksowo od pozostałych, np. miast na prawach powiatu wśród zbiorowości powiatów. Powoduje to zawsze pewne upośledzenie modelu i w związku z tym niejednokrotnie w takich sytuacjach postulowano nawet wyłączenie tego rodzaju obiektów do odrębnej analizy. Ograniczenie ucięcia mediany Webera tylko do liczby obiektów „klasycznych" pozwala nieco zredukować te niedogodności poprzez pewne „wygładzenie" parametrów normalizacyjnych. Formuła normalizacyjna ma postać (Młodak, 2006 b):

$$z_{ij} = \frac{x_{ij} - \theta_j}{1{,}4826 \cdot \underset{i=1, 2, ..., n}{\text{med}} |x_{ij} - \theta_j|} \quad i = 1, 2, ..., n, \;\; j = 1, 2, ..., m$$

Taksonomiczny wzorzec rozwoju konstruujemy jako idealny obiekt, opisany poprzez maksymalne wartości znormalizowanych cech (obiekt opisany wektorem $\Psi = (\Psi_1, \Psi_2, ..., \Psi_m) \in \Re^m$ taki, że $\Psi_j = \max_{i=1, 2, ..., n} z_{ij}$, $j = 1, 2, ..., m$), zaś odległość poszczególnych obiektów od tegoż wzorca wyrażamy miarą medianową: $d_i = \underset{j=1, 2, ..., m}{\text{med}} |z_{ij} - \Psi_j|$, $i = 1, 2, ..., n$.

Miernik kompleksowy określony jest jako metacecha $\mu = (\mu_1, \mu_2, ..., \mu_n)$ taka, że:

$$\mu_i = 1 - \frac{d_i}{\text{med}(d) + 2{,}5 \cdot \text{mad}(d)} \quad i = 1, 2, ..., n$$

przy czym $d = (d_1, d_2, ..., d_n)$, zaś $\text{mad}(d) = \underset{i=1, 2, ..., n}{\text{med}} |d_i - \text{med}(d)|$.

W celu dokonania podziału zbiorowości na klasy typologiczne, wyodrębnijmy dwa podzbiory wartościujące zbioru obiektów. Do pierwszego należą te obiekty, dla których wartości miernika są większe od mediany wartości miernika, med(μ), zaś do drugiego — pozostałe. Medianę elementów pierwszego podzbioru oznaczymy przez $\text{med}_1(\mu)$, natomiast drugiego — $\text{med}_2(\mu)$. Wówczas klasy typologiczne wyznaczone tzw. *metodą trzech median* dla zbioru obiektów $\Gamma = \{\Gamma_1, \Gamma_2, ..., \Gamma_n\}$ są następujące:

klasa 1: $\{\Gamma_i \in \Gamma : \mu_i > \text{med}_1(\mu)\}$,
klasa 2: $\{\Gamma_i \in \Gamma : \text{med}(\mu) < \mu_i \leq \text{med}_1(\mu)\}$,
klasa 3: $\{\Gamma_i \in \Gamma : \text{med}_2(\mu) < \mu_i \leq \text{med}(\mu)\}$,
klasa 4: $\{\Gamma_i \in \Gamma : \mu_i \leq \text{med}_2(\mu)\}$.

Alternatywnym rozwiązaniem jest tutaj klasyfikacja dokonywana tzw. *metodą progową*:

klasa 1: $\{\Gamma_i \in \Gamma : \mu_i \geq \text{med}(\mu) + 2{,}5 \cdot \text{mad}(\mu)\}$,
klasa 2: $\{\Gamma_i \in \Gamma : \text{med}(\mu) \leq \mu_i < \text{med}(\mu) + 2{,}5 \cdot \text{mad}(\mu)\}$,
klasa 3: $\{\Gamma_i \in \Gamma : \text{med}(\mu) - 2{,}5 \cdot \text{mad}(\mu) \leq \mu_i < \text{med}(\mu)\}$,
klasa 4: $\{\Gamma_i \in \Gamma : \mu_i < \text{med}(\mu) - 2{,}5 \cdot \text{mad}(\mu)\}$.

Jakość otrzymanej typologii można ocenić za pomocą odpowiednich wskaźników homogeniczności (wewnętrznej jednorodności), heterogeniczności (wzajemnej odrębności) i poprawności skupień. Konstrukcja pierwszego z nich opiera się odpowiednio na medianach odległości euklidesowych pomiędzy elementami tych samych (i co najmniej dwuelementowych) skupień, a następnie ich uśrednieniu medianą globalną. W drugim przypadku wykorzystuje się mediany odległości elementów danego skupienia od obiektów należących do pozostałych klas. Wskaźnik poprawności stanowi iloraz obu współczynników cząstkowych. Przyjmuje on wartość nieujemną — im jest ona bliższa zera, tym grupowanie okazuje się lepsze.

W celu przeprowadzenia całej procedury taksonomicznej, w środowisku SAS napisano odpowiedni moduł obliczeniowy pod nazwą TAKSONOMIA. Najtrudniejszą kwestią było tutaj opracowanie algorytmu wyznaczania uciętej mediany Webera. Większość dostępnych procedur dotyczy „normalnej" jej postaci, czyli sytuacji, gdy $p = n$. Napisano je w językach: FORTRAN (Gower, 1974; Bedall, Zimmermann, 1979) i TURBO PASCAL (Młodak, 2006 a). Dodatkowo występują w nich poważne ograniczenia co do dopuszczalnej liczby obiektów. Stworzony obecnie algorytm oparto na koncepcji Vandeva (2002), wykorzystującej procedurę optymalizacyjną Newtona-Raphsona. Liczby miast na prawach powiatu w województwach mazowieckim i wielkopolskim wynoszą odpowiednio 5 i 4 (a one są najbardziej nietypowe). Dlatego to ucięcie ustalono na poziomie $p = n - 5$ i $p = n - 4$, gdzie n jest stosowną liczbą gmin/powiatów.

W przypadku gmin mazowieckich, otrzymane wartości miernika wahały się od –0,00169 (Wesoła) do 1,0000 (Warszawa Ochota). Ich zmienność była bardzo

duża (242,1%). Wśród powiatów tego województwa owo zróżnicowanie było już mniejsze, choć i tak dość wyraźne (86,6%). Najniższą wartość miernika zanotowano w powiecie grodziskim (0,072508), zaś najwyższą — w powiecie piaseczyńskim (1,00000). Bardzo zróżnicowane (225,8%) są także gminy woj. wielkopolskiego: od 0,000914 (Tarnowo Podgórne) do 1,00000 (Poznań Nowe Miasto). Wśród powiatów Wielkopolski (zmienność 133,7%) najgorsza sytuacja panuje w grodziskim (wartość miernika 0,038952), a najlepsza — w Poznaniu (1,00000). W tabl. 2 podano granice klas typologicznych wyznaczone dla poszczególnych modeli, w myśl obu omawianych wyżej sposobów — trzech median oraz progowego.

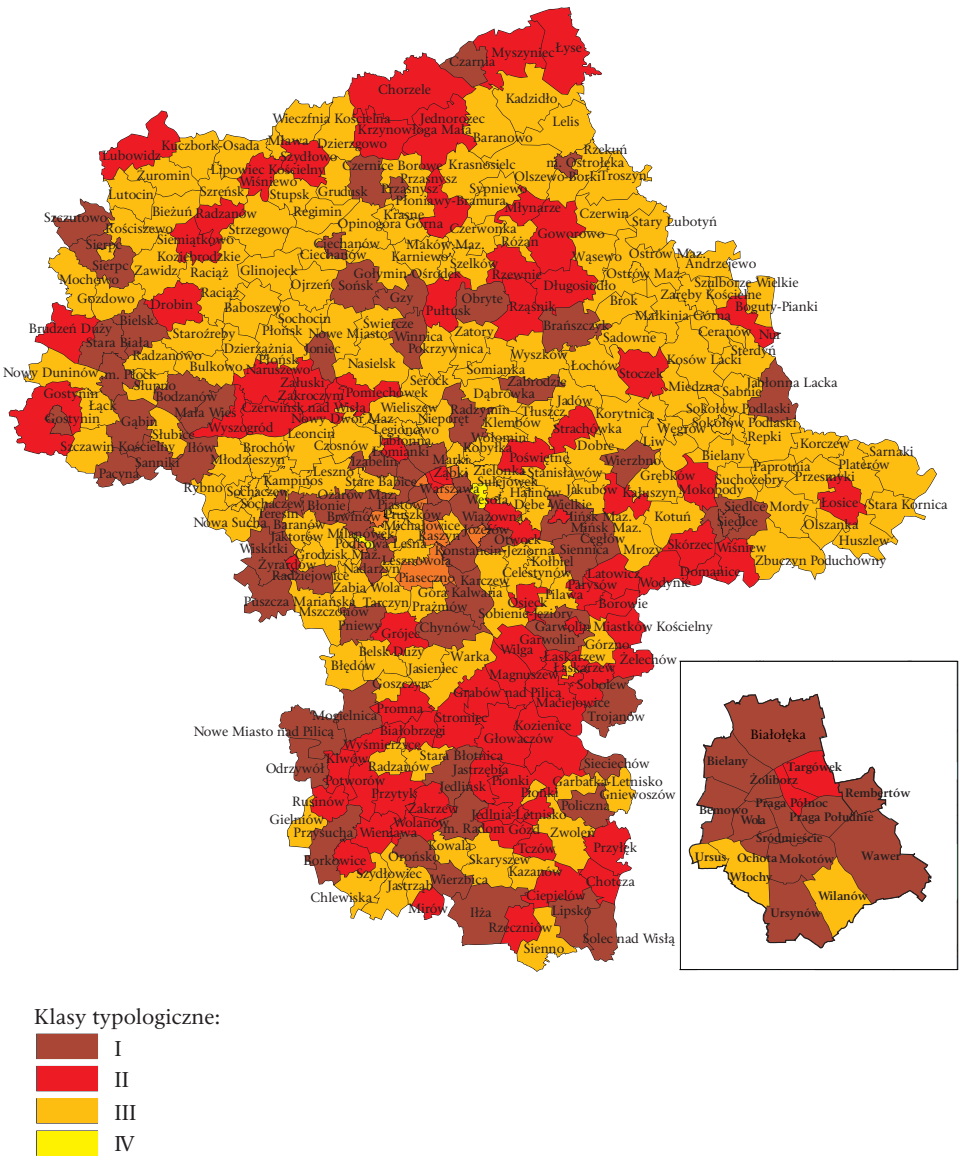
TABL. 2. GRANICE KLAS TYPOLOGICZNYCH

Klasy	Metoda trzech median		Metoda progowa	
	granica dolna	granica górna	granica dolna	granica górna
Woj. mazowieckie				
Gminy				
I	0,0239807	x	0,0255766	x
II	0,0127883	0,0239807	0,0127883	0,0255766
III	0,0088645	0,0127883	0,0000000	0,0127883
IV	x	0,0088645	x	0,0000000
Powiaty				
I	0,2543636	x	0,3336292	x
II	0,1668146	0,2543636	0,1668146	0,3336292
III	0,1204341	0,1668146	0,0000000	0,1668146
IV	x	0,1204341	x	0,0000000
Woj. wielkopolskie				
Gminy				
I	0,0376424	x	0,0326091	x
II	0,0163045	0,0376424	0,0163045	0,0326091
III	0,0114085	0,0163045	0,0000000	0,0163045
IV	x	0,0114085	x	0,0000000
Powiaty				
I	0,1318202	x	0,1607116	x
II	0,0803558	0,1318202	0,0803558	0,1607116
III	0,0535371	0,0803558	0,0000000	0,0803558
IV	x	0,0535371	x	0,0000000

Ź r ó d ł o: opracowanie własne na podstawie obliczeń dokonanych przy użyciu pakietu SAS Enterprise Guide 4.1.

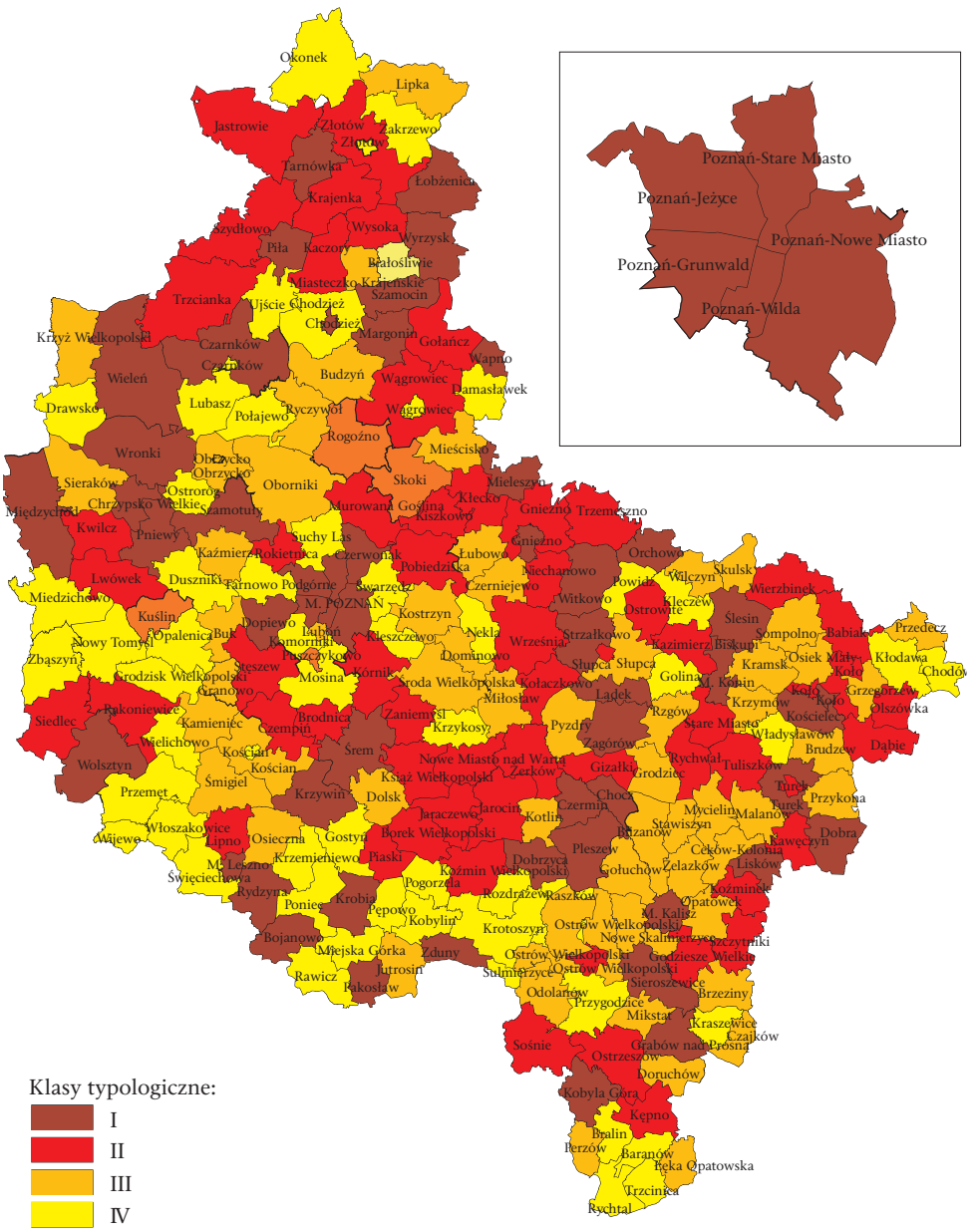
Jakość grupowania różni się w zależności od modelu i rodzaju metody. Opcja trzech median najlepsza jest dla Wielkopolski (wskaźnik poprawności skupień wyniósł tutaj 0,948813, przy 0,9507544 dla podejścia progowego w modelu gminnym i odpowiednio 1,0227516 oraz 1,1339789 w powiatowym), zaś metoda progowa — dla gmin i powiatów Mazowsza (0,8747358 wobec 1,0540528 dla sposobu alternatywnego w pierwszym i odpowiednio 0,8211523 oraz 1,0244203 w drugim przypadku). Najistotniejsze różnice dotyczą najniższej, czwartej klasy. W myśl metody progowej często bywa ona pusta lub co najwyżej jednoelementowa, podczas gdy w przypadku sposobu opartego na trzech medianach może zawierać sporo obiektów. Zamieszczone dalej wykresy obrazują zróżnicowanie przestrzenne gmin według najlepszej typologii dla obu województw.

Wykr. 4. TYPOLOGIA GMIN WOJ. MAZOWIECKIEGO WEDŁUG METODY PROGOWEJ



Ź r ó d ł o: opracowanie własne.

Wykr. 5. TYPOLOGIA GMIN WOJ. WIELKOPOLSKIEGO WEDŁUG METODY TRZECH MEDIAN



Ź r ó d ł o: opracowanie własne.

Najwięcej gmin woj. mazowieckiego należy do, stosunkowo słabej, III grupy. Najlepsza sytuacja panuje w centralnej i południowej części regionu, co może być związane z oddziaływaniem dwóch miast: Warszawy i Radomia. Gminy Wielkopolski są o wiele bardziej zróżnicowane, co częściowo wynika z właściwości samej metody grupowania oraz z faktu, że wiele gmin jest dobrze rozwiniętych rolniczo (np. rejon pilski czy pleszewsko-jarociński), co wywiera niebagatelny wpływ na wartość kapitału ludzkiego.

W przypadku powiatów najlepsze grupowanie na Mazowszu to: klasa I — otwocki, piaseczyński, pułtuski, m. Radom, m. Siedlce, warszawski; klasa II — ciechanowski, garwoliński, gostyniński, lipski, miński, m. Ostrołęka, płocki, m. Płock, pruszkowski, przasnyski, przysuski, sochaczewski, warszawski zachodni, wyszkowski, żyrardowski; klasa III — białobrzeski, grodziski, grójecki, kozienicki, legionowski, łosicki, makowski, mławski, nowodworski, ostrołęcki, ostrowski, płoński, radomski, siedlecki, sierpecki, sokołowski, szydłowiecki, węgrowski, wołomiński, zwoleński, żuromiński. Klasa IV jest pusta.

Dla Wielkopolski (metoda trzech median) sytuacja wygląda następująco: klasa I — czarnkowsko-trzcianecki, gnieźnieński, międzychodzki, ostrzeszowski, pilski, pleszewski, m. Poznań, śremski; klasa II — chodzieski, m. Kalisz, kolski, m. Konin, m. Leszno, słupecki, szamotulski, turecki, wolsztyński; klasa III — gostyński, jarociński, krotoszyński, leszczyński, obornicki, ostrowski, poznański, rawicki, wrzesiński; klasa IV — grodziski, kaliski, kępiński, koniński, kościański, nowotomyski, średzki, wągrowiecki, złotowski. A zatem zdecydowanie lepszą sytuacją w zakresie kapitału ludzkiego charakteryzują się obszary o znacznym potencjale ekonomicznym lub posiadające bogatą tradycję określonej branży gospodarczej (np. powiaty ostrzeszowski czy turecki w Wielkopolsce).

Wnioski

Wyniki przeprowadzonej analizy wskazują, że w zależności od badanego obszaru i poziomu agregacji, modele taksonomiczne mogą istotnie różnić się między sobą. Nie istnieje więc jeden uniwersalny zestaw cech diagnostycznych. W większości przypadków homogeniczność klas typologicznych jest zbliżona do heterogeniczności, a czasem nawet ją przewyższa. To niekorzystne zjawisko może mieć swe źródło w odmienności lokalnych czynników kształtujących zaplecze rynku pracy na poszczególnych obszarach, ale także w arbitralności ustalania liczby granic klas, co bywa krytykowane przez taksonomów. Alternatywę w tej sytuacji stanowi np. endogeniczne wyznaczenie skupień (co jednak częstokroć grozi ich rozdrobnieniem) bądź też zastosowanie klasyfikacji rozmytej, w której każdemu obiektowi przypisywane jest prawdopodobieństwo przynależności do skupienia.

Badanie wykazało, że ludzki potencjał rynku pracy wykazuje znaczne zróżnicowanie terytorialne, a efektywnie skonstruowany miernik kompleksowy może być cennym narzędziem oceny tego aspektu życia społeczno-gospodarczego. Metacechę tę da się skutecznie wykorzystać jako zmienną pomocniczą w testo-

waniu metod estymacji pośredniej dla małych obszarów. Pozwoli to na redukcję wielu kosztów związanych z uzyskiwaniem różnorodnych zmiennych statystycznych w spisach powszechnych.

dr hab. Andrzej Młodak — *Urząd Statystyczny w Poznaniu, Centrum Statystyki Miast*

LITERATURA

Bedall F. K., Zimmermann H. (1979), *The Mediancentre*, Applied Statistics, vol. 28, No. 1

Gower J. C. (1974), *The Mediancentre*, Applied Statistics, vol. 23, No. 1

Lira J., Wagner W., Wysocki F. (2002), *Mediana w zagadnieniach porządkowania obiektów wielocechowych*, [w:] J. Paradysz (red.), *Statystyka regionalna w służbie samorządu lokalnego i biznesu*, Internetowa Oficyna Wydawnicza Centrum Statystyki Regionalnej, Akademia Ekonomiczna w Poznaniu, Poznań

Malina A., Zeliaś A. (1998), *On Building Taxonometric Measures on Living Conditions*, „Statistics in Transition", vol. 3, No. 3

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

Młodak A. (2006 b), *Multilateral normalisations of diagnostic features*, „Statistics in Transition", vol. 7, No. 5

Sobczyk M. (1995), *Wybrane zagadnienia taksonomii numerycznej*, [w:] *Rozwój metodologii badań statystycznych w Polsce*, „Biblioteka Wiadomości Statystycznych", tom 44, GUS i PTS, Warszawa

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

Vandev D. L. (2002), *Computing of Trimmed L_1 — Median*, Laboratory of Computer Stochastics, Institute of Mathematics, Bulgarian Academy of Sciences, www.fmi.uni-sofia.bg/fmi/statist/ /Personal/Vandev/papers/aspap.pdf

SUMMARY

Results of an analysis of spatial diversification of human capital on the labour market, performed within preparatory activities to the National Population and Housing Census 2011, are presented in the article. The study concerned the gminas and powiats of the Mazowieckie and Wielkopolskie Voivodships. Using data collected during the previous census conducted in 2002, 15 indicative variables describing this phenomena were selected and next their verification on the account of variation and correlation was performed. Next, the complex measure was constructed using the trimmed Weber median to the characteristic normalisation. On the basis of the measure value, the collection of areas was divided into typological classes and conclusions for census necessities were formulated.