

Beata KRASZEWSKA

Wykorzystanie analizy skupień w ocenie zróżnicowania zagrożenia ubóstwem w podregionach Polski¹

Streszczenie. *Artykuł poświęcony jest zagadnieniu wykorzystania metod analizy skupień do oceny zróżnicowania zagrożenia ubóstwem w podregionach. Na podstawie danych dotyczących rynku pracy, wynagrodzeń i opieki społecznej, opracowanych na podstawie zasobów Banku Danych Lokalnych GUS (BDL) i ich weryfikacji zmiennościowo-korelacyjnej, określono zestaw cech diagnostycznych, które posłużyły do wyznaczenia skupień podregionów podobnych pod względem zagrożenia ubóstwem.*

Otrzymane rezultaty porównano z wynikami prac studialnych w zakresie przestrzennej dywersyfikacji oszacowań wskaźnika zagrożenia ubóstwem ARPR (at-risk-of-poverty rate) w 2011 r., przeprowadzonych przez Ośrodek Statystyki Małych Obszarów Urzędu Statystycznego w Poznaniu we współpracy z ekspertami Banku Światowego.

Słowa kluczowe: analiza skupień, metoda Warda, odległość euklidesowa, wskaźnik zagrożenia ubóstwem.

Celem artykułu jest opis wykorzystania metod analizy skupień do oceny zróżnicowania zagrożenia ubóstwem w podregionach w 2011 r. Bieda występuje w społeczeństwach od wieków, dlatego badanie ubóstwa wciąż nie traci na ważności. W Polsce wielu naukowców zajmowało i zajmuje się tym zagadnieniem m.in.: Panek, Zwierchowski (2013); Golinowska (2012); Luszniwicz (1982); Panek, Podgórski, Szulc (1999); Kordos, Ochocki (1993) czy Meliczek (1991).

Jak powszechnie wiadomo, ubóstwo jest zjawiskiem wielowymiarowym i mierzy się je przy pomocy wielu różnorodnych wskaźników, których konstrukcja oparta jest częstokroć na istotnie odmiennych zasobach informacji staty-

¹ Artykuł opracowany na podstawie referatu wygłoszonego podczas Ogólnopolskiej Konferencji *Pomiar ubóstwa i wykluczenia społecznego w układach regionalnych i lokalnych*, Poznań, 11 i 12 czerwca 2015 r.

stycznych. Bogaty zestaw wskaźników „laekenowskich” jest tego najlepszym przykładem. Często nawet to samo pojęcie może być z analitycznego punktu widzenia interpretowane rozmaicie. I właśnie zagrożenie ubóstwem jest tego przykładem. Z jednej strony bowiem ocenia się je pojedynczym wskaźnikiem ARPR (*at-risk-of-poverty rate*) jako odsetek osób, których dochód ekwiwalentny do dyspozycji (po uwzględnieniu w dochodach transferów społecznych) jest niższy od granicy ubóstwa ustalonej na poziomie 60% krajowej mediany ekwiwalentnych dochodów do dyspozycji gospodarstwa domowego, z drugiej — można wszak potraktować je jako zjawisko wielowymiarowe i opisać odpowiednimi wskaźnikami ilustrującymi poszczególne czynniki nań wpływające.

W opracowaniu zbadano zatem zróżnicowanie zagrożenia ubóstwem w podregionach Polski w 2011 r., analizując różnorodne wskaźniki z zakresu rynku pracy, opieki społecznej i wynagrodzeń. Zastosowano przy tym analizę skupień według metody Warda z odległością euklidesową. Następnie porównano wyniki analizy skupień z klasyfikacją podregionów według wielkości wskaźnika zagrożenia ubóstwem ARPR, uzyskaną za pomocą równych zakresowo jego klas wielkościowych.

Artykuł składa się z czterech części. W części pierwszej przedstawiono założenia przeprowadzanej analizy, w drugiej zaprezentowano wyniki przeprowadzonej analizy skupień, natomiast część trzecia zawiera porównanie rezultatów analizy skupień z klasami uzyskanymi na podstawie wartości wskaźnika zagrożenia ubóstwem ARPR. Na zakończenie sformułowano wnioski dotyczące zagrożenia ubóstwem w podregionach Polski w 2011 r.

ZAŁOŻENIA ANALIZY

Punkt wyjścia przeprowadzonej analizy stanowiły dane wskaźnikowe dotyczące rynku pracy, wynagrodzeń i opieki społecznej zaczerpnięte z Banku Danych Lokalnych (BDL). W początkowej fazie analizy przyjęto 52 wskaźniki uzyskane z BDL oraz 177 wskaźników obliczonych na podstawie danych BDL — łącznie było to 229 wskaźników.

Zastosowano analizę skupień według metody Warda z odległością euklidesową, która jest podejściem hierarchicznym i aglomeracyjnym. Jej celem było wyznaczenie grup podregionów podobnych ze względu na zagrożenie ubóstwem.

W pierwszym kroku pobrano dane źródłowe z BDL dotyczące ludności, bezrobocia, opieki społecznej oraz wynagrodzenia brutto w 2011 r. w poszczególnych podregionach Polski. Na podstawie danych źródłowych obliczono odpowiednie wskaźniki. Następnie przeprowadzono weryfikację zmiennościową cech:

- obliczono średnią i nieobciążone odchylenie standardowe, a przy ich pomocy wyznaczono współczynnik zmienności,
- wyeliminowano zmienne, które wskazywały na małe zróżnicowanie podregionów, czyli dla których wartość bezwzględna współczynnika zmienności

była poniżej 10%. Zmienne takie charakteryzują się bowiem zbyt niską zdolnością różnicującą badane obiekty.

Po odrzuceniu zmiennych o małym współczynniku zmienności, przy pomocy pakietu Statistica utworzono macierz korelacji Pearsona dla pozostałych zmiennych. Następnie wyznaczono macierz odwrotną do macierzy korelacji. Jej elementy diagonalne winny należeć do przedziału $[1, \infty)$ (Neter i in., 1985; Malina, Zeliaś, 1998). Jeśli tak nie było lub były one zbyt duże (większe od 10), to oznaczało, że pewne zmienne są nadmiernie skorelowane z pozostałymi. Podjęto więc próbę takiej eliminacji, aby usunąć jak najmniej zmiennych powodujących wadliwą postać owej macierzy, a równocześnie uzyskać zestaw, w przypadku którego macierz odwrotna do macierzy korelacji będzie miała na diagonalu tylko elementy większe od 1 i jednocześnie mniejsze niż 10. Ustalano zatem kolejno wartość krytyczną elementów diagonalnych macierzy na poziomie większym od: 100000; 10000; 1000; 100; 50; 40 i 10, redukując poszczególne kolumny i wiersze, a następnie wyznaczono ponownie macierze odwrotne. Uzyskano macierz korelacji zmiennych diagnostycznych, których pozostało 18.

W kolejnym etapie określono wpływ zmiennych diagnostycznych na zjawisko ubóstwa w podregionach w 2011 r. Wytypowano stymulanty — zmienne, których wysoka wartość pogłębiała zjawisko ubóstwa w podregionach oraz destymulanty — zmienne, których wysoka wartość zmniejszała natężenie tego zjawiska (Młodak, 2006). Do stymulant zaliczono 12 zmiennych:

- X_1 — udział ludności w wieku poprodukcyjnym w ogólnej jej liczbie w %;
- X_3 — udział osób bezrobotnych zarejestrowanych w miastach w liczbie osób bezrobotnych zarejestrowanych ogółem w %;
- X_6 — udział bezrobotnych mężczyzn w wieku 25—34 lata w ogólnej liczbie bezrobotnych w %;
- X_7 — udział bezrobotnych kobiet z wykształceniem policealnym lub średnim zawodowym w ogólnej liczbie bezrobotnych w %;
- X_8 — udział bezrobotnych kobiet z wykształceniem średnim ogólnokształcącym w ogólnej liczbie bezrobotnych w %;
- X_9 — udział bezrobotnych mężczyzn z wykształceniem zasadniczym zawodowym w ogólnej liczbie bezrobotnych w %;
- X_{10} — udział bezrobotnych kobiet z wykształceniem gimnazjalnym i poniżej w ogólnej liczbie bezrobotnych kobiet w %;
- X_{11} — udział bezrobotnych mężczyzn ze stażem pracy 1 rok i mniej w ogólnej liczbie bezrobotnych w %;
- X_{12} — udział bezrobotnych kobiet ze stażem pracy 1 rok i mniej w ogólnej liczbie bezrobotnych kobiet w %;
- X_{13} — udział bezrobotnych mężczyzn ze stażem pracy 5—10 lat w ogólnej liczbie bezrobotnych w %;
- X_{14} — udział bezrobotnych kobiet ze stażem pracy powyżej 30 lat w ogólnej liczbie bezrobotnych ze stażem pracy powyżej 30 lat w %;
- X_{16} — udział bezrobotnych kobiet pozostających bez pracy 6—12 miesięcy w ogólnej liczbie bezrobotnych w %.

- Za destymulanty uznano 6 pozostałych zmiennych diagnostycznych:
- X_2 — przeciętne miesięczne wynagrodzenie brutto ogółem w zł;
 - X_4 — udział osób bezrobotnych zarejestrowanych z prawem do zasiłku w liczbie bezrobotnych zarejestrowanych ogółem w %;
 - X_5 — udział osób bezrobotnych wyrejestrowanych z tytułu podjęcia pracy w liczbie wyrejestrowanych bezrobotnych ogółem w %;
 - X_{15} — udział bezrobotnych kobiet pozostających bez pracy 3 miesiące i mniej w ogólnej liczbie bezrobotnych kobiet w %;
 - X_{17} — kwota świadczeń rodzinnych na 100 mieszkańców w tys. zł;
 - X_{18} — udział gospodarstw domowych korzystających z pomocy społecznej powyżej kryterium dochodowego w ogólnej ich liczbie w %.

Następnie w celu unifikacji charakteru cech diagnostycznych zmieniono znaki wartości destymulant na przeciwne, po czym dla otrzymania ujednoczonych co do zakresu i miana ich wartości dokonano normalizacji owych cech. Zastosowano tutaj metodę unitaryzacji zerowanej daną wzorem (Kukuła, 2000):

$$z_{ij} = \frac{x_{ij} - \min_{k=1,2,\dots,n} x_{kj}}{\max_{k=1,2,\dots,n} x_{kj} - \min_{k=1,2,\dots,n} x_{kj}}$$

gdzie:

$i=1, 2, \dots, n,$

$j=1, 2, \dots, m,$

n — liczba obiektów,

m — liczba cech diagnostycznych.

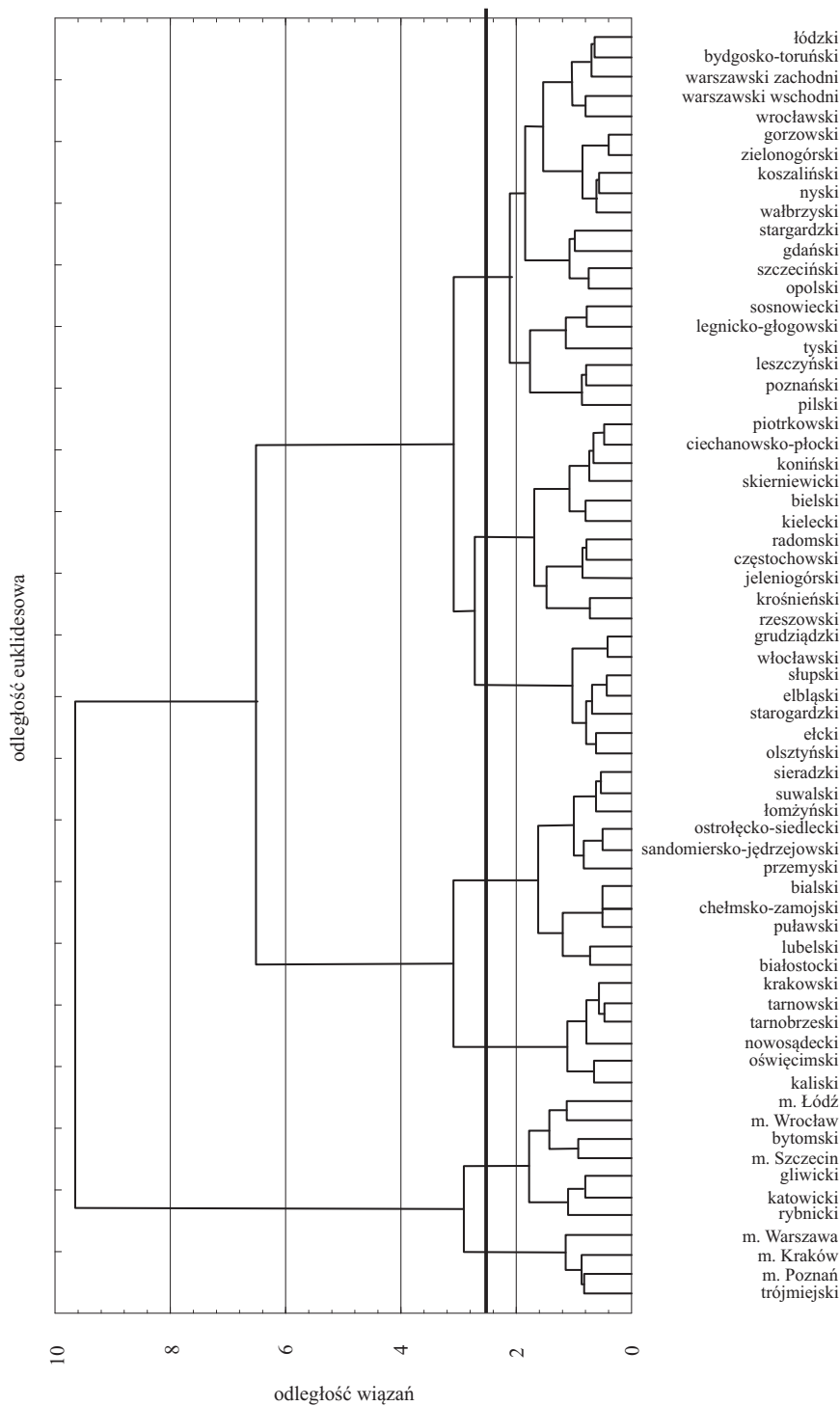
Zastosowanie tego wzoru spowodowało, że wszystkie wartości normalizowanych zmiennych należą do przedziału $[0, 1]$. Zakres znormalizowanych zmiennych stał się identyczny, co umożliwiło sporządzenie wykresu średnich wartości median w obu klasyfikacjach — wskaźnika zagrożenia ubóstwem oraz analizy skupień.

Metodę unitaryzacji zerowanej wybrano w celach porównawczych w grupach syntetycznych wskaźników o podobnym zakresie, jak wskaźnik zagrożenia ubóstwem (ARPR) analizowany przez Zespół Ośrodka Statystyki Małych Obszarów (OSMO) (Mapy..., 2013).

WYNIKI ANALIZY SKUPIEŃ

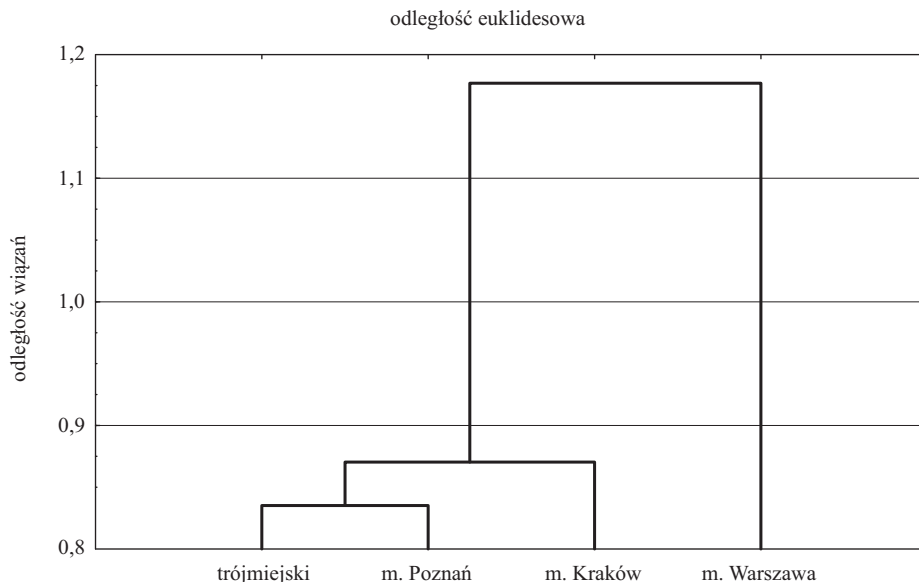
Jak już wspomniano, do wyznaczania podregionów podobnych ze względu na poziom zagrożenia ubóstwem zastosowano metodę analizy skupień Warda (Ward, 1963). Po dokonaniu aglomeracji przecięto dendrogram na wysokości 2,5, aby otrzymać 7 skupień, podobnie jak w przytoczonej pracy Mapy... (2013). Pokazano to na wyk. 1, natomiast wykresy 2—8 ilustrują wewnętrzne różnicowanie grup uzyskanych w wyniku tej klasyfikacji.

Wykr. 1. DENDROGRAM CAŁOŚCIOWEGO GRUPOWANIA PODREGIONÓW



Źródło: opracowanie własne przy użyciu pakietu Statistica.

WYKR. 2. SUBDENDROGRAM DLA GRUPY 1 PODREGIONÓW



Źródło: jak przy wykry. 1.

Grupa 1 charakteryzuje się: najwyższym udziałem ludności w wieku poprodukcyjnym; największymi dochodami ludności wyrażonymi przeciętnym wynagrodzeniem miesięcznym brutto w gospodarce narodowej; najlepszą sytuacją bezrobotnych, czyli największą liczbą osób bezrobotnych mających prawo do zasiłku; najwyższym udziałem gospodarstw domowych korzystających z pomocy społecznej powyżej kryterium dochodowego; najniższą kwotą świadczeń rodzinnych w tys. zł na 100 mieszkańców, a także najwyższym udziałem kobiet pozostających bez pracy do 3 miesięcy w ogólnej ich liczbie.

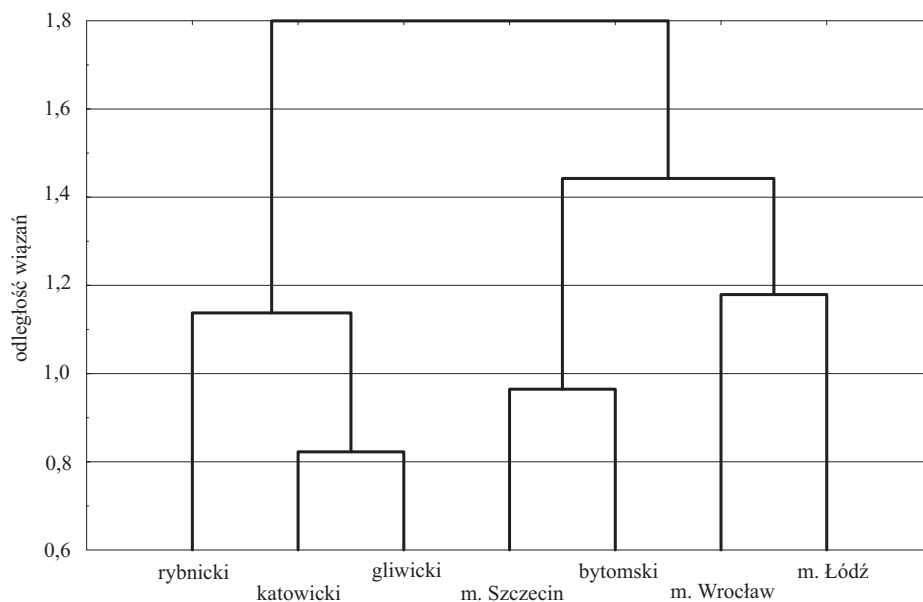
Grupę 2 cechuje: najmniejszy udział bezrobotnych mężczyzn w wieku 25—34 lata w ogólnej liczbie bezrobotnych; szczególnie wyraźne zróżnicowanie odsetka bezrobotnych wyrejestrowanych z tytułu podjęcia pracy; znaczny udział bezrobotnych kobiet ze stażem pracy do 1 roku w ogólnej ich liczbie oraz maksymalne odsetki bezrobotnych kobiet z wykształceniem gimnazjalnym i niższym.

Grupa 3 wyróżnia się z kolei największym odsetkiem bezrobotnych kobiet z wykształceniem policealnym i średnim zawodowym w ogólnej liczbie bezrobotnych, a zarazem najmniejszym udziałem kobiet z wykształceniem gimnazjalnym i niższym w ogólnej ich liczbie. Ponadto obserwuje się w tej grupie minimalny udział bezrobotnych mężczyzn ze stażem pracy 1 rok i mniej w ogólnej liczbie bezrobotnych.

Charakterystyczne właściwości podregionów w skupieniu 4 to najmniejsza skala bezrobocia kobiet pozostających bez pracy od 6 do 12 miesięcy; największy udział bezrobotnych młodych mężczyzn w ogólnej liczbie bezrobotnych oraz stosunkowo mała feminizacja bezrobotnych o najdłuższym stażu pracy (30 lat i więcej).

Wykr. 3. SUBDENDROGRAM DLA GRUPY 2 PODREGIONÓW

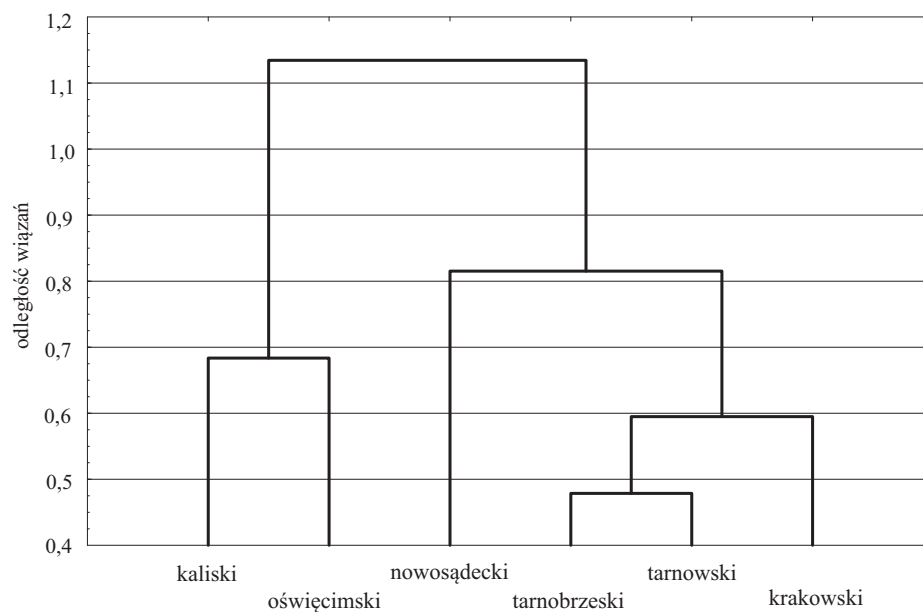
odległość euklidesowa



Źródło: jak przy wykr. 1.

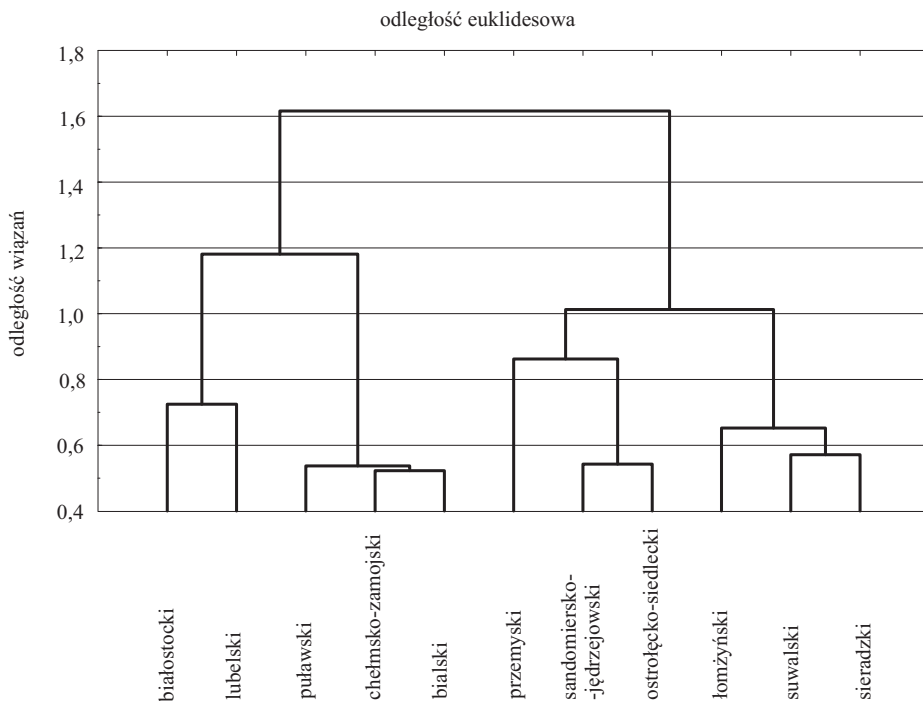
Wykr. 4. SUBDENDROGRAM DLA GRUPY 3 PODREGIONÓW

odległość euklidesowa



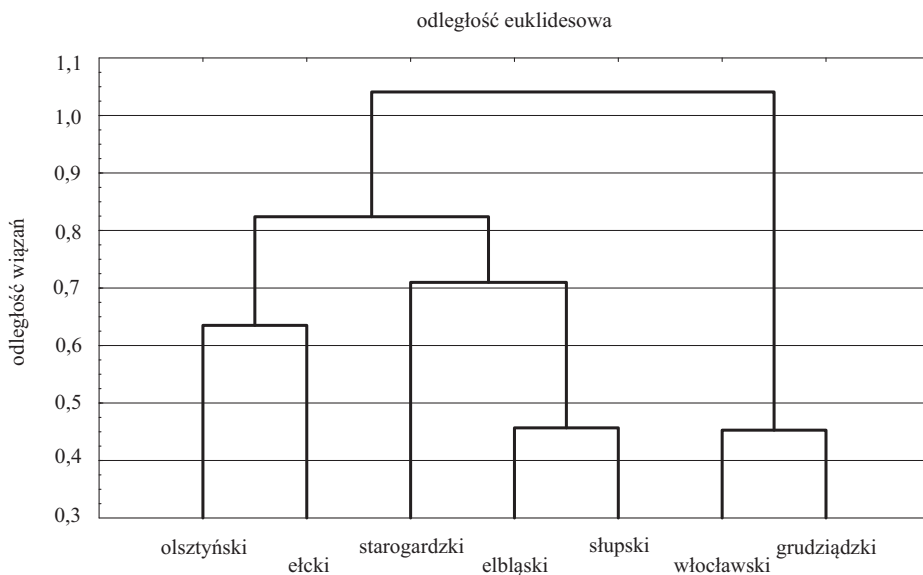
Źródło: jak przy wykr. 1.

Wykr. 5. SUBDENDROGRAM DLA GRUPY 4 PODREGIONÓW



Źródło: jak przy wykr. 1.

Wykr. 6. SUBDENDROGRAM DLA GRUPY 5 PODREGIONÓW



Źródło: jak przy wykr. 1.

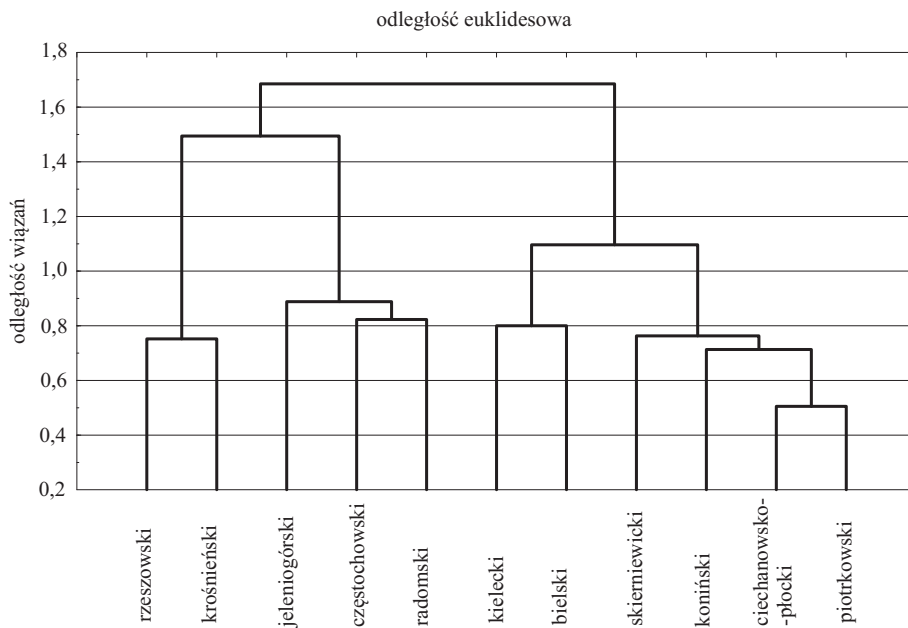
Skupienie 5 cechuje: najmniejszy udział ludności w wieku poprodukcyjnym w ogólnej jej liczbie; niskie przeciętne wynagrodzenia miesięczne; najmniejszy udział bezrobotnych kobiet z wykształceniem policealnym i średnim zawodowym w ogólnej liczbie bezrobotnych; najmniejszy udział bezrobotnych mężczyzn ze stażem pracy 1 rok i mniej w ogólnej liczbie bezrobotnych oraz stosunkowo wysoki udział bezrobotnych mężczyzn ze stażem pracy 5—10 lat.

Skupienie 6 charakteryzuje się najmniejszym odsetkiem bezrobotnych z prawem do zasiłku; najmniejszym udziałem bezrobotnych kobiet z wykształceniem średnim ogólnokształcącym w ogólnej liczbie bezrobotnych oraz najmniejszym udziałem bezrobotnych kobiet pozostających bez pracy do 3 miesięcy w ogólnej ich liczbie.

Grupa 7 to przede wszystkim najmniejsze udziały bezrobotnych kobiet ze stażem pracy do 1 roku w ogólnej ich liczbie; duże udziały bezrobotnych kobiet o najkrótszym okresie pozostawania bez pracy (do 3 miesięcy) oraz relatywnie spory udział kobiet wśród bezrobotnych z najdłuższym stażem pracy (powyżej 30 lat).

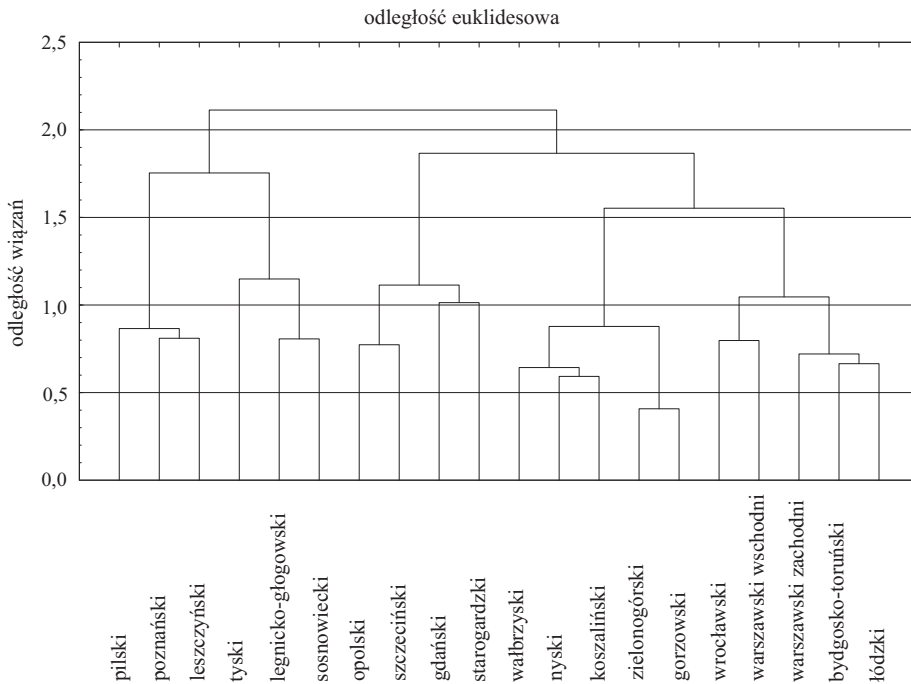
W wyniku przeprowadzonej analizy taksonomicznej można sformułować następujące spostrzeżenia: największy odsetek ludności w wieku poprodukcyjnym wystąpił w dużych aglomeracjach zaliczonych do grupy 1, tj.: m. Warszawie, m. Krakowie, m. Poznaniu i podregionie trójmiejskim, w których wynosił ok. 20%, natomiast najmniejszy w grupie 5 obejmującej podregiony: elbląski, ełcki, olsztyński, słupski, starogardzki, gdzie wynosił średnio 15%. Przeciętne miesięczne wynagrodzenie brutto w przeliczeniu na mieszkańca było najwyższe w dużych miastach należących do grupy 1 (wymienionych wcześniej) i kształtowało się w przedziale od 4 tys. zł do 5 tys. zł, a najmniejsze notowano wśród podregionów grupy 5 — od 2,9 tys. zł do 3,2 tys. zł. Największy odsetek zarejestrowanych bezrobotnych z prawem do zasiłku wystąpił w dużych miastach zaliczanych do grupy 1 i wyniósł ok. 20%, najmniejszy zaś wystąpił w grupie 6 i stanowił ok. 16%. Należy zaznaczyć, że w podregionach grupy 7 znalazły się takie, dla których wartość tego wskaźnika wyniosła ponad 22%, jak np. poznański czy wałbrzyski. Największy odsetek — ok. 50% bezrobotnych — znalazł pracę w grupie 3, tj. w podregionach: krakowskim, nowosądeckim, oświęcimskim, tarnowskim, tarnobrzeskim i kaliskim. Szczególnie wyraźne zróżnicowanie tego wskaźnika obserwowano w skupieniu 2 — od 32,4% w podregionie gliwickim (minimum) do 48,2% w m. Łodzi. Najmniejszy odsetek bezrobotnych mężczyzn w wieku 25—34 lata wystąpił na obszarach zaliczonych do grupy 2, tj. w podregionach: m. Łodzi, bytomskim, gliwickim, katowickim, rybnickim, m. Szczecinie oraz m. Wrocławiu i wyniósł ok. 10%. Największa wartość tego wskaźnika wystąpiła w podregionach: sieradzki, ostrołęcko-siedleckim, białskim, chełmsko-zamojskim, lubelskim, puławskim, przemyskim, białostockim, łomżyńskim, suwalskim i sandomiersko-jędrzejowskim, zaliczonych do grupy 4 i była — przeciętnie rzecz ujmując — o 4 p.proc. wyższa niż w grupie 2.

Wykr. 7. SUBDENDROGRAM DLA GRUPY 6 PODREGIONÓW



Źródło: jak przy wykr. 1.

Wykr. 8. SUBDENDROGRAM DLA GRUPY 7 PODREGIONÓW



Źródło: jak przy wykr. 1.

W podregionach należących do grupy 3, tj.: krakowskim, nowosądeckim, oświęcimskim, tarnowskim, tarnobrzeskim i kaliskim wystąpił największy odsetek bezrobotnych kobiet z wykształceniem policealnym i średnim zawodowym (ok. 16%) oraz z wykształceniem ogólnokształcącym (ok. 9%) w ogólnej liczbie bezrobotnych. Najmniejszy udział kobiet bezrobotnych z wykształceniem policealnym i średnim zawodowym notowano w podregionach: grudziądzkim, włocławskim, słupskim, starogardzkim, elbląskim, elckim i olsztyńskim, należących do grupy 5, gdzie stanowił ok. 13%, natomiast z wykształceniem średnim ogólnokształcącym — w grupie 6 obejmującej podregiony: piotrowski, skierniewicki, ciechanowsko-płocki, radomski, bielski, częstochowski, krośnieński, rzeszowski, kielecki, koniński oraz jeleniogórski i wyniósł od 5,3% do 7,9% ogólnej liczby bezrobotnych.

Wśród bezrobotnych kobiet największy ich odsetek w grupie z wykształceniem gimnazjalnym i niższym obserwowano w regionach zaliczanych do grupy 2, tj.: m. Łodzi, bytomskim, gliwickim, katowickim, rybnickim, m. Szczecinie oraz m. Wrocławiu i wyniósł przeciętnie ok. 30%, natomiast najmniejszy — niższy prawie o połowę, wynoszący przeciętnie ok. 16% — w podregionach: krakowskim, nowosądeckim, oświęcimskim, tarnowskim, tarnobrzeskim, kaliskim, które utworzyły skupienie 3.

Analizując udział bezrobotnych mężczyzn z wykształceniem zasadniczym zawodowym w ogólnej liczbie bezrobotnych widoczne jest, że pod tym względem w Polsce wyodrębniły się dwa obszary. Do pierwszego zaliczyć można podregiony z grup 1 i 2, gdzie udział bezrobotnych mężczyzn z wykształceniem zasadniczym zawodowym w ogólnej liczbie bezrobotnych wynosił ok. 11—12%. W drugim obszarze udział ten był większy przeciętnie o ok. 5—6 p.proc. Udział bezrobotnych mężczyzn ze stażem pracy 1 rok i mniej w ogólnej liczbie bezrobotnych okazał się najmniejszy w podregionach z grupy 5 (ok. 5%, z wyjątkiem podregionu elckiego, gdzie przekroczył 6,8%) oraz w podregionach grupy 3 (5—6%). W pozostałych podregionach był wyższy przeciętnie o 1—2 p.proc.

Największy udział bezrobotnych kobiet ze stażem pracy 1 rok i mniej w ogólnej liczbie kobiet bezrobotnych wystąpił w grupie 2 podregionów i wyniósł przeciętnie 20%, natomiast najmniejszy obserwowano w grupie 7, gdzie był średnio o ok. 4—9 p.proc. niższy. Z kolei udział bezrobotnych kobiet ze stażem pracy powyżej 30 lat w ogólnej liczbie bezrobotnych ze stażem powyżej 30 lat wyniósł w Polsce ok. 30%, przy czym udział mężczyzn w tej grupie był znacznie wyższy. Szczególnie duże wartości tego wskaźnika notowane są w podregionach skupienia 2 (w niektórych przypadkach przekraczając nawet 40%). Największy udział bezrobotnych kobiet pozostających bez pracy 3 miesiące i mniej w ogólnej liczbie bezrobotnych kobiet wystąpił w grupie 1, obejmującej duże aglomeracje miejskie, gdzie wyniósł powyżej 30%, natomiast najmniejszy był w grupie 6 i ukształtował się na poziomie przeciętnie o 10 p.proc. mniej-

szym. Największy udział kobiet pozostających bez pracy 6—12 miesięcy w ogólnej liczbie bezrobotnych obserwowano w grupie 1 i wyniósł przeciętnie ok. 11%, natomiast najmniejszy był w grupie 4, gdzie ukształtował się na poziomie ok. 8—9%.

Jak wynika z analizy, udział bezrobotnych mężczyzn pozostających bez pracy przez 6—12 miesięcy w ogólnej liczbie bezrobotnych był znacznie wyższy niż w przypadku kobiet. Z kolei udział bezrobotnych mężczyzn ze stażem pracy 5—10 lat w ogólnej liczbie bezrobotnych we wszystkich podregionach Polski wyniósł ok. 6—7%. Wyższe wartości tych wskaźników obserwowano w skupieniu 5.

Największa kwota świadczeń rodzinnych na 100 mieszkańców wystąpiła w grupie 5 i wyniosła przeciętnie ok. 30 tys. zł na 100 mieszkańców, natomiast najmniejszą obserwowano w grupie 1, gdzie była przeciętnie o ok. 21 tys. zł niższa.

Największy udział gospodarstw domowych korzystających z pomocy społecznej powyżej kryterium dochodowego dał się zaobserwować w grupie 1, gdzie wyniósł nawet powyżej 50%, a w pozostałych podregionach udział ten był średnio o ok. 8—10 p.proc. mniejszy, bez wyraźnej koncentracji obszarów o najmniejszych wartościach tego wskaźnika.

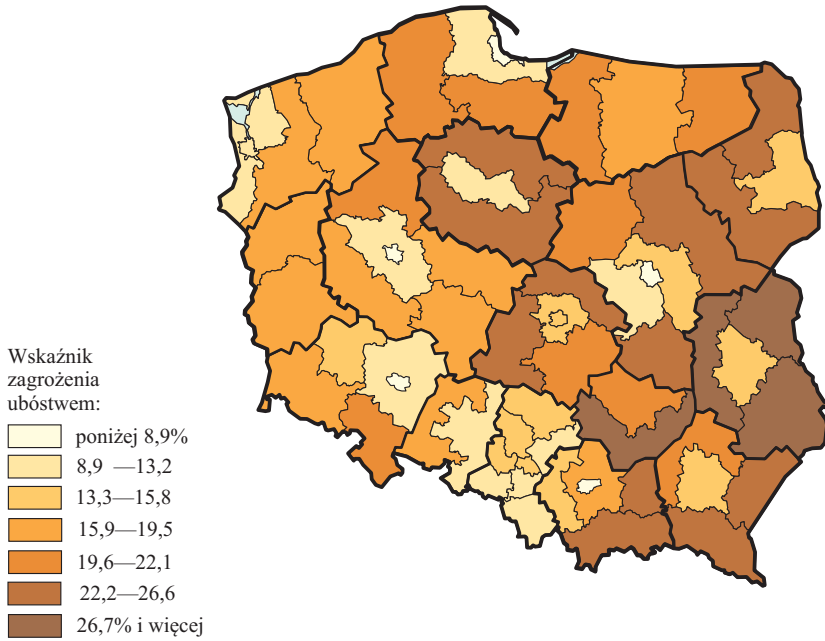
PORÓWNANIE REZULTATÓW ANALIZY SKUPIEŃ Z KLASAMI WIELKOŚCI WSKAŹNIKA ZAGROŻENIA UBÓSTWEM ARPR

W tej części rozważań rezultaty wyznaczenia skupień podregionów podobnych ze względu na poziom zagrożenia ubóstwem porównano z wynikami prac studialnych w zakresie przestrzennej dywersyfikacji oszacowań wskaźnika zagrożenia ubóstwem (ARPR) w 2011 r. przeprowadzonych przez OSMO we współpracy z ekspertami Banku Światowego (*Mapy...*, 2013) (wykr. 9 i 10). Oszacowany przez OSMO, na podstawie badania EU-SILC (Europejskie Badanie Warunków Życia Ludności), wskaźnik zagrożenia ubóstwem definiuje się jako odsetek osób, których dochód ekwiwalentny do dyspozycji (po uwzględnieniu w dochodach transferów społecznych) jest niższy od granicy ubóstwa ustalonej na poziomie 60% krajowej mediany ekwiwalentnych dochodów do dyspozycji gospodarstwa domowego.

Rozpatrywano wówczas wielkości wskaźnika ARPR według 7 klas:

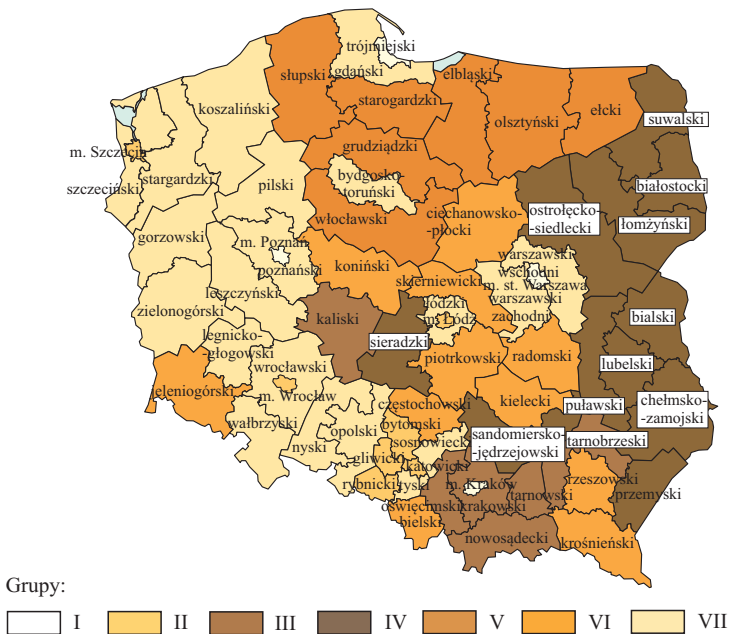
- I — poniżej 8,9%,
- II — 8,9—13,2%,
- III — 13,3—15,8%,
- IV — 15,9—19,5%,
- V — 19,6—22,1%,
- VI — 22,2—26,6%,
- VII — 26,7% i więcej.

Wykr. 9. KLASY WIELKOŚCI WSKAŹNIKA ARPR



Źródło: *Mapy...* (2013).

Wykr. 10. GRUPY ZAGROŻENIA UBÓSTWEM WEDŁUG ANALIZY SKUPIEŃ



Źródło: opracowanie własne z wykorzystaniem pakietu SAS Enterprise Guide 4.3.

Do celów porównawczych — w przypadku wyk. 10 — przypisano (nadano) kolor klasy wartości wskaźnika ARPR:

- grupie 1 — poniżej 8,9%, ponieważ udział w tej klasie wskaźnika wyniósł 100% podregionów tej grupy;
- grupie 7 — na poziomie 8,9—13,2%, ze względu na 45% udział podregionów 7 grupy skupienia;
- grupie 2 — na poziomie 13,3—15,8%, z uwagi na 57,1% udział podregionów z 2 grupy skupienia;
- grupie 6 — z przedziału 15,9—19,5%, ze względu na 27,2% udział podregionów tego skupienia;
- grupie 5 — na poziomie 19,6—22,1%, z uwagi na 57,1% udział podregionów należących do tego skupienia;
- grupie 3 — z przedziału 22,2—26,6%, z uwagi na to, że 33,3% podregionów tej grupy weszło w skład owej klasy ARPR;
- grupie 4 — na poziomie 26,7% i więcej, ponieważ zawiera tylko podregiony skupienia 4 (36,4% jego składu).

Jak widać, w obu przypadkach zagrożenie ubóstwem koncentruje się na wschodzie i południu kraju, choć w przypadku wskaźnika ARPR koncentracja ta jest nieco mniej widoczna. W tablicy pokazano szczegółowe porównanie składu grup uzyskanych drogą analizy skupień oraz klas określonych na podstawie wartości wskaźnika ARPR.

**PORÓWNANIE ZAWARTOŚCI KLAS WEDŁUG ARPR
ORAZ GRUP OTRZYMANYCH W WYNIKU ANALIZY SKUPIEŃ**

Podregiony	Grupy według analizy skupień	Odsetek podregionów grupy należących do danej klasy ARPR
ARPR poniżej 8,9%		
M. Poznań	1	100,0
M. Kraków	1	
M. Warszawa	1	
Trójmiejski	1	
M. Wrocław	2	14,3
ARPR 8,9—13,2%		
Szczeciński	7	45,0
Gdański	7	
Bydgosko-toruński	7	
Poznański	7	
Warszawski zachodni	7	
Opolski	7	
Tyski	7	
Wrocławski	7	
Sosnowiecki	7	
M. Szczecin	2	
Rybnicki	2	28,6
Bielski	6	9,1

**PORÓWNANIE ZAWARTOŚCI KLAS WEDŁUG ARPR
ORAZ GRUP OTRZYMANÝCH W WYNIKU ANALIZY SKUPIEŃ (cd.)**

Podregiony	Grupy według analizy skupień	Odsetek podregionów grupy naleŹących do danej klasy ARPR
------------	------------------------------	--

ARPR 13,3—15,8%

Legnicko-głogowski	7	15,0
Łódzki	7	
Warszawski wschodni	7	
Gliwicki	2	57,1
Katowicki	2	
Bytomski	2	
M. Łódź	2	
Oświęcimski	3	16,7
Białostocki	4	9,1
Częstochowski	6	9,1

ARPR 15,9—19,5%

Koszaliński	7	30,0
Gorzowski	7	
Zielonogórski	7	
Leszczyński	7	
Nyski	7	
Stargardzki	7	27,2
Jeleniogórski	6	
Koniński	6	
Rzeszowski	6	14,3
Olsztyński	5	
Lubelski	4	9,1
Kaliski	3	33,3
Krakowski	3	

ARPR 19,6—22,1%

Starogardzki	5	57,1
Elbląski	5	
Ełcki	5	
Słupski	5	
Ciechanowsko-płocki	6	27,3
Piotrkowski	6	
Kielecki	6	
Piłski	7	10,0
Wałbrzyski	7	
Tarnobrzeski	3	16,7

ARPR 22,2—26,6%

Sieradzki	4	45,5
Ostrołęcko-siedlecki	4	
Łomżyński	4	
Suwalski	4	
Przemyski	4	
Skierniewicki	6	27,3
Radomski	6	

**PORÓWNANIE ZAWARTOŚCI KLAS WEDŁUG ARPR
ORAZ GRUP OTRZYMANÝCH W WYNIKU ANALIZY SKUPIEŃ (dok.)**

Podregiony	Grupy według analizy skupień	Odsetek podregionów grupy należących do danej klasy ARPR
------------	------------------------------	--

ARPR 22,2—26,6% (dok.)

Krośnieński	6	27,3
Tarnowski	3	33,3
Nowosądecki	3	
Grudziądzki	5	28,6
Wrocławski	5	

ARPR 26,7% i więcej

Sandomiersko-jędrzejowski	4	
Puławski	4	36,4
Białski	4	
Chełmsko-zamojski	4	

Źródło: opracowanie własne z wykorzystaniem wyników prac OSMO (Mapy..., 2013).

Klasa wskaźnika zagrożenia ubóstwem na poziomie poniżej 8,9% pokryła w 100% podregiony grupy 1, wyznaczonej na podstawie analizy skupień (tablica). Warto jednak zauważyć, że okazała się ona nieco szersza niż owo skupienie. Tym różnicującym elementem jest Wrocław, który nie znalazł się w owym skupieniu m.in. z uwagi na znaczny udział bezrobotnych kobiet ze stażem pracy 1 rok i mniej. Wskaźnik zagrożenia ubóstwem wyznaczony przez OSMO na poziomie 8,9—13,2% osiągnęło 45% podregionów, które zaliczono w tej analizie skupień do grupy 7, gdzie obserwuje się m.in. niską wysokość kwoty świadczeń rodzinnych na 100 mieszkańców oraz niewysoki odsetek bezrobotnych kobiet ze stażem pracy 1 rok i mniej. Ponadto do rozpatrywanej klasy według wskaźnika ARPR wchodziły też podregiony, które w analizie skupień zaliczono do grupy 2 oraz 6. Rozpatrywane cechy diagnostyczne w tym przypadku noszą zatem ze sobą dodatkowe informacje, których nie obejmuje ARPR. W owych dwóch podregionach skupienia 2 można dostrzec m.in. niskie udziały bezrobotnych mężczyzn ze stażem pracy 5—10 lat w ogólnej liczbie bezrobotnych, jak też niewielką kwotę świadczeń rodzinnych na 100 mieszkańców. W przypadku podregionu bielskiego, należącego do skupienia 6, obok relatywnie niskiego udziału bezrobotnych kobiet z wykształceniem gimnazjalnym i niższym można zauważyć również niską kwotę świadczeń rodzinnych na 100 mieszkańców, co sugeruje, że ma ona dość wyraźny wpływ na ARPR. Jest to o tyle logiczne, iż zarówno kryteria wypłacania świadczeń, jak i wskaźnik ARPR oparte są na dochodach gospodarstw domowych.

Klasa podregionów, których wskaźnik zagrożenia ubóstwem ukształtował się na poziomie 13,3—15,8% również zawierała podregiony, które w wyniku anali-

zy skupień zaliczono do grupy 2. Cechowały się one wysokim udziałem ludności w wieku poprodukcyjnym i wysokim udziałem bezrobotnych kobiet z wykształceniem gimnazjalnym i niższym. Znalazły się tam trzy podregiony z grupy 7 charakteryzujące się wysokim odsetkiem bezrobotnych kobiet ze stażem pracy 30 lat i więcej w ogólnej liczbie bezrobotnych z takim stażem, podobnie jak jedyny podregion ze skupienia 3 (oświęcimski) czy też podregion białostocki należący do grupy 4, która dodatkowo charakteryzuje się najniższymi udziałami bezrobotnych kobiet z wykształceniem policealnym i średnim zawodowym oraz bezrobotnych mężczyzn z wykształceniem zasadniczym zawodowym w ogólnej ich liczbie. Skupienie 6 reprezentowane jest także przez jeden podregion (częstochowski) charakteryzujący się m.in. niskim udziałem bezrobotnych kobiet z wykształceniem średnim ogólnokształcącym oraz bezrobotnych kobiet z wykształceniem gimnazjalnym i niższym. Przemieszczenie się części grupy 7 do kolejnego przedziału wskaźnika zagrożenia ubóstwem mogło być wynikiem zaokrągleń oszacowań. Warto w tym miejscu przypomnieć, że grupa 7, ustalona na podstawie analizy skupień, była grupą najbardziej liczną, obejmującą 20 podregionów.

W przedziale zagrożenia ubóstwem według ARPR na poziomie 15,9—19,5% ponownie wystąpiły podregiony z grupy 7 (głównie z zachodniej i południowej części kraju — niemal 1/3 składu tej grupy), jednak w mniejszej liczbie niż w przedziale 8,9—13,3% (do którego należało — jak pamiętamy — 9). To potwierdza, że grupa 7 jest bardzo liczna i wewnętrznie zróżnicowana. Znajdują się tutaj również trzy podregiony z grupy 6, m.in. z niskim odsetkiem bezrobotnych kobiet z wykształceniem średnim ogólnokształcącym, dwa z grupy 3, które charakteryzuje wysoki udział bezrobotnych wyrejestrowanych z tytułu podjęcia pracy czy bezrobotnych kobiet z wykształceniem policealnym i średnim zawodowym, ale też z relatywnie niskim udziałem bezrobotnych młodych mężczyzn w wieku 25—34 lata oraz jeden z grupy 5, który ma względnie niewysoki udział ludności w wieku poprodukcyjnym oraz gospodarstw domowych korzystających z pomocy społecznej, jak również dość znaczny odsetek bezrobotnych kobiet z wykształceniem gimnazjalnym i niższym.

Kolejny przedział wskaźnika zagrożenia ubóstwem to 19,6—22,1%. Warto zauważyć, że należące tu podregiony ze skupienia 5 położone są na północy Polski, zaś ze skupienia 6 — w środkowej. Takiej regularności nie ma np. w przypadku należących tutaj podregionów grupy 7 oraz 3. Zaliczone do tej klasy wskaźnika ARPR podregiony grupy 5 charakteryzują się wysokim udziałem bezrobotnych z prawem do zasiłku oraz względnie niskim udziałem młodych mężczyzn (w wieku 25—34 lata) wśród bezrobotnych. Podregiony z grupy 6 wyróżniają się m.in. wysokim udziałem bezrobotnych kobiet ze stażem pracy powyżej 30 lat w ogólnej ich liczbie oraz względnie niską kwotą świadczeń rodzinnych na 100 mieszkańców. Z kolei podregion tarnobrzeski ze skupienia 3 charakteryzuje się niskim udziałem bezrobotnych kobiet z wykształceniem gimnazjalnym i niższym oraz

wysokim udziałem gospodarstw domowych korzystających z pomocy społecznej w ogólnej ich liczbie.

Wskaźnik zagrożenia ubóstwem na poziomie 22,2—26,6% wyznaczył z kolei klasę obejmującą niemal połowę podregionów zaliczanych do grupy 4, ponad 1/4 podregionów z grupy 6, w których występuje stosunkowo niski udział bezrobotnych kobiet pozostających bez pracy 6—12 miesięcy oraz po dwa podregiony z grupy 3 (charakteryzujące się wysokim udziałem bezrobotnych kobiet z wykształceniem średnim ogólnokształcącym) i grupy 5 (w których zauważa się — co z racjonalnego punktu widzenia dość nietypowe — wysoki udział bezrobotnych wyrejestrowanych z tytułu podjęcia pracy i znaczny poziom kwoty świadczeń rodzinnych na 100 mieszkańców w tys. zł).

Ostatnia klasa wielkości zagrożenia ubóstwem (26,7% i więcej) zawierała cztery podregiony, które zaliczono w wyniku analizy skupień do grupy 4, cechujące się m.in. niskim udziałem bezrobotnych mężczyzn ze stażem pracy 5—10 lat w ogólnej liczbie bezrobotnych. Rozlokowanie grupy 4 w przedziale ostatnim i przedostatnim może być spowodowane zaokrągleniami oszacowań.

Wyniki grupowania podregionów według podobieństwa w zakresie zagrożenia ubóstwem są zbliżone do efektów prac poznańskiego OSMO. Występują jednakże pewne różnice, co widać szczególnie w rozczłonkowaniu grupy 7 między bardzo różnorodnymi klasami wielkości wskaźnika zagrożenia ubóstwem badanego przez OSMO. Może to mieć źródło w fakcie, że wskaźnik ARPR — w przeciwieństwie do analizy skupień — nie uwzględnia oddziaływania wielu różnorodnych czynników decydujących o wystąpieniu ryzyka ubóstwa.

Wnioski

Z przeprowadzonych analiz wynika, że zarówno według wskaźnika ARPR, jak i analizy skupień południowa część wschodniej Polski charakteryzowała się największym zagrożeniem ubóstwem. Do skupienia wskazującego obszary o największych problemach w tym zakresie należą wszakże podregiony północnej części wschodu naszego kraju. Najmniejszy stopień zagrożenia ubóstwem objął natomiast w obu metodach podregiony zurbanizowane.

Jak można zauważyć, zaobserwowany w skupieniu 1 najniższy stopień zagrożenia ubóstwem wynikał głównie z faktu, że podregiony tego skupienia charakteryzuje nie tylko najkorzystniejsza sytuacja dochodowa, ale również niskie bezrobocie. Z kolei w grupie 4 dostrzeżono korzystny niski odsetek bezrobotnych mężczyzn ze stażem pracy 5—10 lat (co jednak może mieć związek z rolniczym charakterem tych obszarów, a zatem szerszymi możliwościami pracy — także dorywczej — w tym właśnie sektorze), a w grupie 5 — największą kwotę świadczeń rodzinnych na 100 mieszkańców, co pośrednio może stanowić konsekwencję niskiego udziału ludności w wieku poprodukcyjnym, gdyż nie korzysta już ona raczej z tego rodzaju świadczeń. Zdecydowane różnice widać na

poziomie wskaźnika ARPR 8,9—13,2% (najbardziej podobnym do grupy 7). Według ARPR podregiony tej klasy są rozrzucone w całej Polsce, natomiast według analizy skupień obejmują głównie Polskę zachodnią. Polska północna to praktycznie analogiczne wyniki obu metod, w przeważającej części tego regionu wskaźnik zagrożenia ubóstwem wynosił 19,6—22,1%. Południowa część Polski jest bardzo zróżnicowana pod względem zagrożenia ubóstwem, od najniższego wskaźnika w m. Krakowie do najwyższego w podregionie sandomiersko-jędrzejowskim — jednak i tutaj widać podobne wyniki obu metod.

Średnich wartości median znormalizowanych cech nie można przedstawić na jednym wykresie z uwagi na to, że obie klasyfikacje są różne (wskaźnik ARPR — według prac OSMO, w analizie skupień rozpatrywane są zupełnie inne i bardziej różnorodne wskaźniki). Jednak można zauważyć, że zróżnicowanie średnich wartości median znormalizowanych cech w obu metodach obserwacji przestrzennego rozkładu zagrożenia ubóstwem jest bardzo zbliżone, aczkolwiek nieco większe dla efektów analizy skupień.

mgr Beata Kraszewska — *Urząd Statystyczny w Poznaniu*

LITERATURA

- Golinowska S. (2012), *O biedzie i polityce jej zwalczania współcześnie. Podejście porównawcze w świetle Europejskiego Roku Zwalczania Ubóstwa i Wykluczenia Społecznego*, [w:] *Polska bieda w świetle Europejskiego Roku Walki z Ubóstwem i Wykluczeniem Społecznym*, red. H. E. Kubiak, Krakowskie Wydawnictwo Edukacyjne sp. z o.o. — Oficyna Wydawnicza AFM, Kraków: s. 93—115.
- Kordos J., Ochocki A. (1993), *Problemy pomiaru ubóstwa w krajach EWG i w Polsce*, „Wiadomości Statystyczne”, nr 1: s. 3—8.
- Kukuła K. (2000), *Metoda unitaryzacji zerowanej*, Wydawnictwo Naukowe PWN, Warszawa.
- Luszniewicz A. (1982), *Statystyka Społeczna. Podstawowe problemy i metody*, Państwowe Wydawnictwo Ekonomiczne, Warszawa.
- Malina A., Zeliaś A. (1998), *On Building Taxonomic Measures on Living Conditions*, „Statistics in Transition”, Vol. 3, No. 3: s. 523—544.
- Mapy ubóstwa na poziomie podregionów w Polsce z wykorzystaniem estymacji pośredniej* (2013), Urząd Statystyczny w Poznaniu, Ośrodek Statystyki Małych Obszarów, GUS, Poznań, http://stat.gov.pl/download/gfx/portalinformacyjny/pl/defaultaktualnosci/5732/4/1/1/mapowanie_ubostwa_prace_studialne.pdf (dostęp 06.11.2013 r.).
- Meliczek H. (1991), *Metody pomiaru ubóstwa*, „Wiadomości Statystyczne”, nr 12: s. 1—6.
- Młodak A. (2006), *Analiza taksonomiczna w statystyce regionalnej*, Centrum Doradztwa i Informacji DIFIN, Warszawa.
- Neter J., Wasserman W., Kutner M. H. (1985), *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs*, nr 469, Homewood, IL: Richard D. Irwin, Inc.
- Panek T., Podgórski J., Szulc A. (1999), *Ubóstwo: teoria i praktyka pomiaru*, SGH, Warszawa.
- Panek T., Zwierzchowski J. (2013), *Porównawcza analiza sfery ubóstwa w krajach UE w ujęciu regionalnym*, Zeszyty Naukowe, Instytut Statystyki i Demografii SGH, Warszawa.
- Ward J. H. (1963), *Hierarchical Grouping to Optimize an Objective Function*, „Journal of the American Statistical Association”, Vol. 58: s. 236—244.

Summary. *The article discusses the use of cluster analysis methods to assess the differentiation of risk of poverty in the Polish subregion. On the basis of data on the labor market, wages and social care, developed on the basis of the resources of the Local Data Bank of the CSO and their variable-correlation verification, the author has defined a set of diagnostic features used to determine the cluster of sub-regions similar in terms of risk of poverty. Results were compared with the results of study work in the field of spatial diversification of estimating at-risk-of-poverty rate (ARPR) in 2011 conducted by the Small Areas Statistics Centre of the Statistical Office in Poznań in cooperation with experts from the World Bank.*

Keywords: cluster analysis, Ward's method, Euclidean distance, at-risk-of-poverty rate.

Резюме. *Статья была посвящена вопросу использования методов кластерного анализа для оценки дифференциации опасности выступления бедности в субрегионах. На основе данных по рынке труда, вознаграждений и социального обеспечения, разработанных на основе фондов Банка локальных данных ЦСУ и непостоянно-корреляционной проверки, был определен набор диагностических признаков, который использовался в определении кластеров субрегионов аналогичных в отношении к опасности выступления бедности.*

Полученные результаты были сопоставлены с результатами научных работ в области пространственной диверсификации оценок показателя опасности выступления бедности ARPR (at-risk-of-poverty rate) в 2011 г. проведенных Центром статистики малых домен Статистического управления в Познани в сотрудничестве с экспертами Всемирного банка.

Ключевые слова: кластерный анализ, метод Уорда, расстояние Евклида, показатель опасности выступления бедности.