

TADEUSZ RUTKOWSKI

## METODY TAKSONOMETRYCZNE W BADANIACH REGIONALNYCH

### I. WSTĘP

W ostatnim okresie nastąpił wzrost badań dotyczących różnych aspektów ekonomiki regionów. Podejmowanie trafnych decyzji gospodarczych stymulujących rozwój badanych zjawisk w układzie regionalnym wymaga uprzedniego poznania ich przestrzennego zróżnicowania. Aby analiza była poprawna, należy z całego obszaru poddanego badaniu wydzielić grupy jednostek terytorialnych wewnątrznie podobnych do siebie po uwzględnieniu wybranych cech. Cel ten można osiągnąć stosując metody taksonometryczne. W literaturze przedmiotu istnieje duża różnorodność metod wykorzystywanych do delimitacji regionów podobnych. Celem niniejszego artykułu jest krótka charakterystyka najczęściej stosowanych metod taksonometrycznych, ukazanie ich zalet i wad oraz możliwości wykorzystania w badaniach ekonomicznych. Chcemy na tym tle zaproponować (kompleksową procedurę regionalizacyjną wykorzystującą najnowsze metody taksonometrii numerycznej opartej na elektronicznej technice obliczeniowej. Zdajemy sobie sprawę, że nie jest to procedura idealna, lecz niewątpliwie jedna z najlepiej nadających się do badań ekonomicznych, ze względu na małą pracochłonność, jednoznaczność wyników i łatwość interpretacji otrzymanych rozwiązań. Procedura ta wykorzystuje trzy programy na EMC: analizę statystyczną XDS2/29 — dla obliczenia głównych składowych, niehierarchiczną analizę skupisk ANO6 oraz nieliniowe mapowanie danych ANO7.

### II. PROBLEMY TEORETYCZNE I METODOLOGICZNE WYSTĘPUJĄCE PRZY STOSOWANIU METOD TAKSONOMETRYCZNYCH

Stosowane metody regionalizacji można najogólniej podzielić na opisowe i ilościowe. W ostatnich latach dominującą pozycję zdobyły metody ilościowe i to zarówno w badaniach prowadzonych w kraju jak i za granicą.

Klasyfikacja metod taksonometrycznych nastęrcza spore trudności

spowodowane dużą liczbą tych metod jak też ich różnorodnością. Metody taksometryczne znajdują szerokie zastosowanie w ekonomicznych badaniach przestrzennych. Służą one do podziału niejednorodnego z punktu widzenia określonych i wymiernych cech diagnostycznych zbioru elementów na podzbiory wewnętrznie bardziej jednorodne. Metody te można podzielić na wzorcowe i bezwzorcowe. W metodach wzorcowych prowadzi się podział badanego zbioru na podzbiory według z góry ustalonych wzorców, a najbardziej znaną (metodą spośród metod wzorcowych jest analiza dyskryminacyjna H. A. Fishera<sup>1</sup>). W badaniach regionalnych delimitację regionów trzeba oprzeć na wielu zmiennych, przy czym trudno jest jednoznacznie a priori określić wzorce jednostek regionalnych, dlatego też powszechnie stosuje się w nich metody bezwzorcowe.

Bezwzorcowe wielowymiarowe metody taksometryczne mają na ogół uniwersalny charakter i zawierają następujące etapy postępowania<sup>2</sup>:

- 1) opracowanie macierzy informacji,
- 2) redukcja przestrzeni wielocechowej,
- 3) estymacja wielocechowego podobieństwa,
- 4) grupowanie jednostek,
- 5) testowanie dokonanej regionalizacji.

Ad 1) Punktem wyjścia w postępowaniu badawczym jest dobór cech diagnostycznych, które w sposób prawidłowy opisywać będą poszczególne elementy badanego zbioru (np. jednostki terytorialne). Najbardziej intuicyjnym, kryterium „doboru” cech diagnostycznych jest wielkość ogólnego zasobu informacji o istotnych właściwościach badanej zbiorowości, dostarczana przez każdą z tych cech. Przy takim podejściu powstaje pytanie, jak wyspecyfikować cechy najlepsze, z danego zespołu cech oraz jak ustalić rangi tych cech ze względu na siłę ich oddziaływania na różnicowanie układów przestrzennych. Zazwyczaj przyjmuje się pewne kryteria, które powinny spełniać cechy diagnostyczne. J. Fierich zakłada, że cechy diagnostyczne powinny<sup>3</sup>:

- ujmować istotne właściwości badanej zbiorowości,
- być proste i logicznie ze sobą powiązane,
- być mierzalne za pomocą wartości bezwzględnych i względnych,
- posiadać wyraźną zmienność przestrzenną i stałość w czasie,
- być niezależne od siebie.

Niektóre z tych kryteriów takie jak: zmienność przestrzenna, niez-

<sup>1</sup> R. A. Fisher, *The procession of discrimint-functions*, Annales of Eugenics, Vol. 10, No 4, 1940.

<sup>2</sup> Z. Chojnicki, T. Czyż, *Metody taksonomii numerycznej w regionalizacji geograficznej*, Warszawa 1973.

<sup>3</sup> Por. J. Fierich, *Próba zastosowania metod taksonomicznych do rejonizacji systemów rolniczych w województwie krakowskim*, Myśl Gospodarcza 1957, nr 1.

leżność są bardzo ważne, inne, jak stałość w czasie, posiadają mniejsze znaczenie. Zjawiska ekonomiczne ulegają dynamicznym zmianom, co wcale nie uniemożliwia stosowania taksonometrycznych metod ich regionalizacji. W tym przypadku za ich pomocą możemy obadać czy dynamika zmian jest jednakowa na całym obszarze czy też różna. Spełnienie warunku pierwszego prowadzi zawsze do uwzględnienia w badaniu znacznej liczby cech. Jednocześnie jednak zbyt duża liczba cech diagnostycznych zwiększa, nieraz niepotrzebnie, pracochłonność obliczeń powodując jednocześnie zaciemnienie ostatecznego obszaru. W dobie obecnej przy wykorzystaniu elektronicznej techniki obliczeniowej traci na ostrości wada pracochłonności, ale w dalszym ciągu czynniki przypadkowe, generowane przez cechy mało istotne, zaciemniają istotę badanego zjawiska. Z tego powodu niezwykle korzystne dla całego procesu badawczego jest stosowanie metod łączenia cech i ich redukcji. Empiryczne wartości cech diagnostycznych wyrażone są w liczbach (bezwzględnych mianowanych. Wadą oczywistą takiego wyrażenia poszczególnych cech jest niesumowalność jednostek różnoimiennych. Wadę tę można wyeliminować poprzez normalizację zmiennych. Normalizacji zmiennych dokonuje się na podstawie wzoru:

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

gdzie:  $x_{ij}$  — zaobserwowane, rzeczywiste wartości zmiennej  $j$  dla każdego  $i$  ( $i=1, 2, \dots, n$ ),  $\bar{x}_j$  — średnia wartość zmiennej  $j$ ,  $s_j$  — odchylenie standardowe zmiennej  $j$ .

Wartości  $t_{ij}$  są znormalizowane, o średniej równej zero i odchyleniu standardowym równym jedności. Dzięki wprowadzeniu normalizacji, wartości zmiennych, wyrażone w różnych jednostkach, są porównywalne, co stwarza możliwość sumowania cech jakościowo różnych. Liczby niemianowane, otrzymane w wyniku normalizacji, pozostają w tym samym stosunku do siebie, w jakim pozostawały odpowiadające im liczby mianowane przed normalizacją. Po dokonaniu normalizacji zmiennych przechodzimy do następnego etapu postępowania, w którym dokonujemy redukcji cech.

Ad 2) Gdy rozpatrzmy zbiór  $N$  punktów (jednostek terytorialnych) za pomocą  $n$  cech diagnostycznych o wartościach rzeczywistych, to elementy tego zbioru można zinterpretować jako punkty przestrzeni euklidesowej  $n$ -wymiarowej. Jednak zbiór punktów, pomimo że jest wyrażony za pomocą  $n$  zmiennych, może w rzeczywistości leżeć, przynajmniej w przybliżeniu, w przestrzeni o mniejszej liczbie  $m$  wymiarów. Trzeba zatem określić  $m$  nowych zmiennych ( $m < n$ ) i dokonać w ten sposób redukcji przestrzeni cechowej. Problem ten można rozwiązać za pomocą metod redukcji przestrzeni wieloczechowej. Istnieje wiele metod redukcji

cech. Z teoretycznego punktu widzenia należy przyjąć taką metodę, która zapewniłaby zrealizowanie kryteriów stawianych cechom diagnostycznym przez nowo otrzymane wyniki. Z tego punktu widzenia najbardziej efektywnymi i przydatnymi są metody analizy czynnikowej. Nazwa analiza czynnikowa jest wieloznaczna i w ogólniejszym zarysie określa się nią wiele metod<sup>4</sup>. Spośród nich najbardziej znane i szeroko stosowane to analiza czynnikowa w dosłownym znaczeniu „factor analysis” i analiza głównych składowych „principal component analysis”.

Wyróżniamy różne techniki postępowania w analizie czynnikowej: centroidalną, osi głównych „principal axes”, największej wiarygodności „maximum likelihood” czy też „image analysis”. Metody te zostały dosyć szeroko opisane w literaturze. Każda z nich ma swoje zalety i wady. Za najpoprawniejszą uważa się na ogół metodę największej wiarygodności D. N. Lawley'a i A. E. Maxwella, według której pracują programy na maszynie cyfrową np. program „Statistical Analysis XDS2/29”. Zasadnicza idea wykorzystania metod analizy czynnikowej w badaniach regionalnych polega na zastosowaniu jej jako metody łączenia cech i sprowadzenia macierzy oryginalnych zmiennych do układu wartości czynników wspólnych w postaci macierzy o mniejszych wymiarach. Czynniki wspólne skupiają informację dotychczas rozproszoną w oryginalnych cechach diagnostycznych. Redukcja liczby cech pociąga za sobą przy tym postępowaniu stosunkowo małą stratę informacji. Wprowadzając w postaci czynników wspólnych niewielką liczbę podstawowych „wymiarów” analiza czynnikowa ma korzystne znaczenie zarówno ze względów praktycznych jak i teoretycznych. Wyznaczone czynniki są uporządkowane według ich wkładu do powodowania zmienności wspólnej. Operując więc ich wartościami własnymi możemy wyznaczyć rangę poszczególnych czynników i ich udział w wyjaśnianiu zmienności wspólnej. Ponadto czynniki te spełniają warunek ortogonalności czyli są niezależne. Dotychczas w badaniach rejonizacji najczęściej stosowano metody analizy czynnikowej jako wstępny krok w procedurze delimitacyjnej, ponieważ wyodrębnione czynniki spełniają wymagania stawiane cechom diagnostycznym<sup>5</sup>.

Analiza czynnikowa posiada również pewne słabe strony do których w szczególności należą:

— arbitralność i niedoskonałość metod szacowania zasobów zmienności wspólnej ( $h^2$ ), co ma wpływ zarówno na liczbę wyodrębnionych czynników, jak i na wartość ładunków czynnikowych,

<sup>4</sup> Por. D. N. Lawley, A. E. Maxwell, *Factor analysis as a statistical method*, London 1963.

<sup>5</sup> T. Czyż, *Zastosowanie metody analizy czynnikowej do badania ekonomicznej struktury regionalnej Polski*, Prace Geograficzne IGPAN 1974, nr 92; T. Rutkowski, *Wyznaczanie czynników określających dochody ludności*, Studia KPZK PAN, t. LII, 1975.

— przyjęcie założenia o liniowej zależności między zmiennymi, co nie zawsze jest sprawdzalne w rzeczywistości.

Spośród wymienionych metod analizy wielowymiarowej za najbardziej poprawną z punktu widzenia teorii statystyki uznawana jest analiza głównych składowych. Szczególne zalety tej metody to:

— jednostkowa wariancja składowych głównych, co eliminuje szacowanie zasobów zmienności wspólnej i daje jednoznaczne wyniki,

— zhierarchizowanie wyznaczonych składowych głównych według ich znaczenia poprzez podanie dla każdej z nich wartości własnej oraz procentu wyjaśniania całkowitej wariancji,

— nieskorelowanie składowych głównych,

— stosunkowo mała strata informacji przy przejściu z przestrzeni  $n$ -wymiarowej do przestrzeni o mniejszej liczbie wymiarów.

Zasadniczo analiza głównych składowych jest ortogonalną transformacją wyjściowego zespołu zmiennych ( $x_i$ ) cech diagnostycznych w zbiór nowych zmiennych ( $y_i$ ) głównych składowych, przy czym liczba nowych zmiennych równa się liczbie zmiennych teoretycznych, a ogólna wariancja zmiennych ( $x_i$ ) równa się wariancji składowych ( $y_i$ ). Jest to więc metoda różna od metod analizy czynnikowej i nie można jej z nimi utożsamiać. W praktyce metodę tę możemy wykorzystać do redukcji wyjściowego zespołu cech, nie tracąc przy tym nic z jej zalet. Posługując się programem „Statistical analysis XDS2/29” oraz podając procent wariancji ogólnej na poziomie niższym niż 100%, np. 95%, uzyskamy liczbę składowych głównych, która wyjaśnia 95% wariancji. Liczba tych składowych jest znacznie mniejsza od wprowadzonej ilości cech diagnostycznych i oscyluje wokół liczby czynników wspólnych, uzyskanych poprzez zastosowanie jednej z metod analizy czynnikowej do redukcji zmiennych.

Reasumując, wykorzystanie analizy głównych składowych do redukcji zmiennych jest bardzo wskazane, a wyniki otrzymane poprzez jej zastosowanie spełniają w pełni warunki stawiane cechom diagnostycznym. Z tego powodu proponujemy zastosowanie tej metody do redukcji cech diagnostycznych.

Ad 3) Następnym etapem w procedurze delimitacyjnej jest estymacja wielocechowego podobieństwa badanych jednostek.

Przy posługiwaniu się cechami diagnostycznymi spełniającymi uprzednio postawione warunki lub zastępującymi je głównymi składowymi, podstawowymi miarami podobieństw są współczynniki odległości i współczynniki korelacji. Współczynniki korelacji występują zasadniczo, gdy dysponujemy empirycznymi wartościami cech diagnostycznych. Do badania współzależności pomiędzy cechami możemy zastosować współczynnik korelacji K. Pearsona lub współczynnik korelacji rang Ch. Spearmana, które występują w taksonomii jako części składowe metody podobieństw

J. Czekanowskiego. Niekiedy wykorzystywane są również współczynniki korelacji wielokrotnej<sup>6</sup>

Zasadniczą grupę miar podobieństwa stanowią współczynniki odległości, oparte na pojęciu odległości w ujęciu geometrii analitycznej pomiędzy punktami w  $n$ -wymiarowej przestrzeni euklidesowej ( $n$  oznacza liczbę cech). Odległość taksonomiczna wyrażona jest wzorem:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2},$$

gdzie:  $d_{ij}$  — odległość między punktami  $i, j$ ,  $x_{ik}$  oraz  $x_{jk}$  —  $k$ -te współrzędne punktów  $i, j$ .

Obliczona macierz odległości przy zastosowaniu odległości taksonomicznej zawiera odległości między każdą jednostką a wszystkimi innymi w przestrzeni cechowej i jest macierzą symetryczną. Należy podkreślić, że miara odległości taksonomicznej może być stosowana tylko w przypadku, gdy przestrzeń cechowa jest ortogonalna. Założenie to jest spełnione w proponowanej metodzie bowiem obliczanie odległości taksonomicznej poprzedza redukcja cech metodą analizy głównych składowych. Obliczone główne składowe są ortogonalne i stanowią dane wejściowe dla obliczenia macierzy współrzędnych wyjściowych punktów (jednostek terytorialnych) w układzie otrzymanych składowych. W polskich badaniach przestrzennych najczęściej stosowaną miarą jest przeciętna różnica J. Czekanowskiego określona wzorem:

$$\bar{d}_{k,p} = \sum_{j=1}^n \frac{x_{kj} - x_{pj}}{n},$$

gdzie:  $\bar{d}_{k,p}$  — przeciętna różnica,  $x_{kj}$  — wartość  $j$  zmiennej w  $k$  elemencie,  $x_{pj}$  — wartość  $j$  zmiennej w  $p$  elemencie,  $n$  — liczba cech diagnostycznych.

Tak określona przeciętna różnica pozwala na porównywalne poszczególnych elementów zbiorowości ze wszystkimi pozostałymi w zakresie wszystkich cech.

W proponowanej przez nas niehierarchicznej analizie skupisk program ANO6 dla obliczenia macierzy odległości wykorzystuje odległość taksonomiczną.

Ad 4) Macierze odległości są podstawą dla grupowania jednostek w jednorodne grupy. Istnieją różne metody postępowania. Do najczęściej stosowanych należą<sup>7</sup>:

<sup>6</sup> A. Wanke, *Metoda badania częstości występowania zespołów cech, czyli metoda stochastycznej korelacji wielorakiej*, Przegląd Antropologiczny 1953 nr 19.

<sup>7</sup> B. Podolec, K. Zajac, *Ekonometryczne metody ustalania rejonów konsumpcji*, Warszawa 1978.

- diagraficzna metoda Czekanowskiego,
- metoda dendrytowa,
- inne metody (analiza skupień, drzewo połączeń B. J. L. Berry'ego, metoda J. H. Wanda).

Najczęściej stosowaną metodą w badaniach polskich jest diagraficzna metoda J. Czekanowskiego. Metoda ta polega na ułożeniu wyjściowego diagramu na podstawie macierzy odległości, a następnie przekształceniu go w taki sposób, aby elementy najbardziej podobne wyodrębniły się najwyraźniej wzdłuż głównej przekątnej. Zasadniczą cechą dodatnią uporządkowanego diagramu Czekanowskiego jest to, że obrazuje on stopień podobieństwa danego elementu (jednostki terytorialnej) z każdym innym. W metodzie tej występuje jednak wiele elementów subiektywnych, które są jej mankamentami<sup>8</sup>. Pierwszy z nich to dowolność w ustalaniu sikali podobieństw. Zarówno liczba stopni podobieństw jak i obszar zmienności przeciętnych różnic w ramach poszczególnych stopni nie są ściśle określone i zależą od przyjętej metody postępowania, których wiele proponuje się w literaturze. Innym nie rozwiązany problem jest takie ustalenie wielkości przeciętnej różnicy między badanymi jednostkami, aby uznać ją za istotną. Ustalona skala podobieństw nie rozwiązuje w pełni tego zagadnienia. Ostatnim dosyć ważnym mankamentem omawianej metody jest brak jednoznacznego sposobu ustalania kolejności porównywanych jednostek terytorialnych. Uporządkowanie wyjściowego diagramu J. Czekanowskiego oparte jest w dużej mierze na intuicji badacza, a przy dużej liczbie jednostek jest to dosyć trudne i pracochłonne. Wymóg maksymalnego skupienia wzdłuż głównej przekątnej jednostek zawierających podobieństwa bliższych stopni jest możliwy do osiągnięcia jedynie metodą kolejnych przybliżeń. W dodatku nigdy nie wiadomo, czy uzyskany w końcu obraz jest rzeczywiście ostateczny. W przypadku istnienia wyraźnych grup jednostek zawierających podobieństwa pierwszego stopnia podział diagramu nie następuje z trudności. Kłopotliwy natomiast problem stanowi zaszeregowanie elementów, które wykazują równie bliskie powiązania z dwoma rejonami. Istnieje wiele propozycji ułatwiających podział uporządkowanego diagramu J. Czekanowskiego takich jak: wykorzystanie do tego celu testu  $\chi^2$  J. Steczkowskiego, propozycja prostego miernika poprawności podziału B. Podolca czy też nowa metoda taksonomiczna Z. Piaseckiego. Wykorzystanie tych metod jest jednak pracochłonne i nie zawsze likwiduje wspomnianą wadę metody diagraficznej.

Jako metodę likwidującą w pewnym stopniu mankamenty metody diagraficznej uważa się metodę dendrytowa, której twórcami są K. Florek, J. Łukaszewicz, J. Perkal, H. Steinhaus i S. Zubrzycki. Dendryt buduje się przez łączenie każdej jednostki z jednostką do niej najbar-

<sup>8</sup> F. Stokowski, *Regionalizacja demograficzna Polski*, Warszawa 1977.

dziej podobną (powiązania pierwszego stopnia). Jeżeli w wyniku takiego postępowania nie wszystkie elementy zbioru zostaną połączone, to poszczególne podzbiory należy łączyć ze sobą w jeden dendryt, przez bezpośrednie połączenie linią elementów różnych podzbiorów najbardziej do siebie podobnych (powiązania drugiego stopnia). Podstawowym założeniem grupowania dendrytowego jest zachowanie warunku, żeby dendryt łączący jednostki w przestrzeni  $n$ -wymiarowej był jak najkrótszy i nie był zamknięty. Ułożenie elementów dendrytu — a tym samym kształt linii łamanej — jest dowolne. Tak zdefiniowane uporządkowanie dendrytowe ma tę wyższość nad diagramem J. Czekanowskiego, że jest jednoznacznie określone. Pewną wadą metody dendrytowej jest to, że dendryt uwzględnia jedynie najmniejsze różnice abstrahując od wszelkich innych. Nie daje on pełnego obrazu przestrzeni, w której znajdują się badane jednostki, a zatem wnioskowanie na podstawie dendrytu oparte jest na mniejszej liczbie informacji niż ta, która jest zawarta w uporządkowanym diagramie J. Czekanowskiego. Dlatego uważa się, że najlepsze rozwiązanie stanowi połączenie metody diafragmicznej i dendrytowej. Jest to o tyle łatwe, że obydwie metody mogą być oparte na tych samych różnicach przeciętnych.

Pośród innych metod grupowania, rzadziej spotykanych w polskiej literaturze ekonomicznej, chcielibyśmy omówić analizę skupisk (cluster analysis) w celu jej szerszego spopularyzowania. Technika analizy skupisk operuje zmieniającym się progiem podobieństwa i zakłada z góry liczbę tworzonych skupisk. Obejmuje ona zwykle dwa etapy:

- identyfikację jąder albo rdzeni skupisk,
- przyporządkowanie pozostałych jednostek do tych jąder.

Wykorzystywany program „Niehierarchicznej analizy skupisk” ANO6 po wprowadzeniu macierzy współrzędnych wyjściowych punktów w układzie otrzymanych głównych składowych oblicza początkowe centra skupisk, współrzędne dla poszczególnych centrów oraz macierz odległości dla centrów skupisk stosując formułę odległości taksonomicznej. Dokonuje też przyporządkowania poszczególnych jednostek terytorialnych do utworzonych centrów z jednoczesnym podaniem wartości odległości danej jednostki od centrum. Dodatkowo dla każdego skupiska podawana jest średnia odległość punktu od centrum co pozwala na wnioskowanie o jednorodności otrzymanych skupisk. Po każdej iteracji (dla poszczególnej liczby skupisk) podaje się wartość odchylenia od centrum dla danego rozwiązania, która informuje w pewnym sensie o „dobroci” uzyskanego rozwiązania.

Niezwykle przydatne jest wykorzystanie programu „Nieliniowego mapowania danych dla analizy struktury danych” ANO7, który na podstawie macierzy współrzędnych wyjściowych punktów w układzie otrzymanych głównych składowych przedstawia w sposób graficzny rozmieszczenie poszczególnych jednostek na płaszczyźnie. Daje to plastyczny obraz

grupowania się poszczególnych jednostek w skupiska i jest doskonałym uzupełnieniem wyników uzyskanych przy niehierarchicznej analizie skupisk. Dodatkowa zaleta sporządzonego wykresu graficznego to ukazanie obrazu wzajemnych relacji wszystkich jednostek terytorialnych. Rozwiązanie otrzymane przez zastosowanie wymienionych programów ANO6 i ANO7 daje wyniki zbliżone do uważanych za „optymalne”, uzyskanych w wyniku połączenia metody diagraficznej J. Czekanowskiego z metodą dendrytową.

Ad 5) Ostatni etap postępowania w procedurze delimitacyjnej to weryfikacja dokonanego podziału. Najlepszym narzędziem dokonania weryfikacji są parametryczne i nieparametryczne testy statystyczne oparte na teorii wnioskowania statystycznego<sup>9</sup>. Z testów parametrycznych możemy stosować testy — Studenta lub Duncana oraz test  $F$  na analizę wariancji. Z testów nieparametrycznych proponuje się przede wszystkim test zgodności rozkładów  $\chi^2$ . Obydwie propozycje mają swoje zalety i wady. Testy parametryczne wymagają założenia o normalności rozkładu badanych zbiorowości, nie określają one ponadto optymalnej regionalizacji tylko weryfikują uzyskany wynik. Mogą więc być użyteczne jedynie w procedurze alokacji jednostek przestrzennych do regionów. W obydwu przypadkach występuje pewien element subiektywizmu związany z wyborem poziomu istotności. Z tych powodów w literaturze można spotkać wiele propozycji sprawdzenia poprawności dokonanych podziałów, które odnoszą się do konkretnych przypadków np. analiza dyskryminacyjna (w przypadku dwu regionów), wskaźnik zagęszczenia S. Kozłowskiego (przy ustalonej skali podobieństw) itp. W programie ANO6 — analizie skupień — jako test oceniający poprawność rozwiązania dla zmieniającej się liczby skupisk wykorzystuje się analizę wariancji. W metodzie tej wariancję ogólną rozbija się na dwa składniki mierzące zmienność pomiędzy regionami i wewnątrz regionów. Porównując obydwie wariancje, dowiadujemy się czy średnie regionalne różnią się istotnie od siebie czy też nie. Przy pomocy testu  $F$  sprawdzamy hipotezę zerową ( $H_0$ ) o braku istotnych różnic między regionami. Sprawdzeniem  $H_0$  jest statystyka  $F$  obliczona według formuły:

$$F = \frac{\text{wariancja międzyregionalna}}{\text{wariancja wewnątrzregionalna}}$$

Porównując obliczane wartości statystyki  $F$  z wartością teoretyczną odczytaną z tablic rozkładu Fischeba-Snedecoira parzy  $k-1$  i  $n-k$  stopniach swobody i założonym poziomie istotności  $H_0$  możemy odrzucać wtedy, gdy wartość obliczona jest większa od teoretycznej. Zastosowanie tego

<sup>9</sup> J. Steczkowski, *Zasady i metody rejonizacji produkcji rolniczej*, Warszawa 1966.

testu w proponowanej metodzie ma pewne zalety. Ponieważ liczba skupisk jest dowolna, to możemy testować każde rozwiązanie od  $k$  do 1. Na tej podstawie stwierdzamy, które rozwiązanie jest najlepsze.

Na zakończenie tych teoretycznych rozważań należy stwierdzić, że żadnej z metod taksonometrycznych nie można stosować mechanicznie. Wynika to ze złożoności związków ekonomiczno-społecznych pomiędzy poszczególnymi jednostkami przestrzennymi. Dlatego niezwykle pomocną w praktyce okazuje się merytoryczna znajomość badanych zjawisk oraz analiza jakościowa.

## TAXONOMIC METHODS IN REGIONAL STUDIES

### Summary

In the article have been taken up theoretical and methodological problems connected with application of taxonomic methods in regional studies. There have been discussed procedure stages in patternless, multidimensional taxonomic methods. In the course of description of particular procedure stages comparisons have been made of such generally applied in hitherto practice taxonomic methods as: Czekanowski method of mean differences, dendritic method, analysis of concentrations. Next there has been proposed certain procedure implementing the following digital computer programmes: factor analysis XDS2/29, nonhierarchical analysis of concentrations ANO6 and nonlinear picture of data structure ANO7.

In the next part of the article have been presented results of application of the proposed study procedure for drawing regionalization of Poland into similar regions in respect of spatial distribution of money incomes of population.

*Tłumaczył Tadeusz Kowalski*