

TADEUSZ GRABIŃSKI, KAZIMIERZ ZAJĄC

NIEKTÓRE NOWE METODY BADANIA STRUKTURY I DYNAMIKI
PROCESÓW DEMOGRAFICZNYCH
NIEMIECKIEJ REPUBLIKI DEMOKRATYCZNEJ

I. WPROWADZENIE

Przystępując do badania jakiegokolwiek zjawiska, a w szczególności procesów demograficznych, należy przede wszystkim poznać jego strukturę, wyjaśnić jakie wewnętrzne zależności wiążą jej poszczególne elementy oraz jakie czynniki zewnętrzne wywołują trwałe lub losowe zmiany całej struktury w czasie. Każda bowiem zmiana w przyrodzie i społeczeństwie jest wynikiem określonych przyczyn. Zasada przyczynowości stanowi z kolei podstawę do poszukiwania prawidłowości w badanej rzeczywistości. Pozwala to na przedstawienie wyników badań naukowych jako mniej lub bardziej wyraźnie ujawniających się prawidłowości¹. Chcąc zatem ocenić je w sposób właściwy należy przede wszystkim zbadać, czy proces obserwowany w czasie ma charakter jednorodny.

Wszelkie więc rozważania modelowe odnoszące się do szeregów czasowych powinny być poprzedzone oceną jednorodności przebiegu zjawiska w czasie. Trzeba stwierdzić, iż wiele błędnych ocen tendencji rozwojowej oraz pseudozależności wykrytych pomiędzy badanymi ocenami miało swe źródło w niedocenianiu tego rodzaju analizy.

Celem pracy jest przedstawienie niektórych metod numerycznych pozwalających na ustalenie w sposób możliwie obiektywny bardziej jednorodnych faz rozwojowych procesów demograficznych w NRD w latach 1949-1973. Chodzi więc o periodyzację, czyli podział badanego okresu czasu na nie ustaloną z góry liczbę rozłącznych, wyczerpujących i możliwie jednorodnych podokresów (faz), zawierających obserwacje względnie bardziej podobne do siebie w obrębie podokresów z punktu widzenia uwzględnionych cech, niż ma to miejsce między podokresami

W niniejszym opracowaniu proponuje się rozwiązywać sformułowane

¹ Por. J. Topolska, *Metodologia historii*, Warszawa 1973, s. 212.

powyżej zagadnienia za pomocą metod taksonomicznych². Jednakże w stosunku do dotychczasowego podejścia, w którym klasyfikowanymi obiektami były jednostki przestrzenne lub cechy charakteryzujące ich własności, dokonuje się istotnej modyfikacji przyjmując, że przedmiotem dyskryminacji są okresy czasu, w których prowadzi się obserwację statystyczną, uzyskując w ten sposób nowe możliwości w zakresie metodologii badań szeregów czasowych oraz znaczne ułatwienie wykrywania ewentualnych prawidłowości rozwojowych, które stanowią podstawę do budowy poprawnych prognoz demograficznych.

Godzi się zauważyć, że w pracy przed przystąpieniem do rozwiązywania właściwego tematu na podstawie rozeznania merytorycznego ustalono pewien zespół cech demograficznych — a z kolei spośród nich na podstawie zaprezentowanych w drugiej części rozprawy metod — dokonano wyboru takiego zespołu cech, który ma istotny wpływ na kształtowanie się procesów demograficznych w NRD. Uzyskany zespół cech stanowił podstawę periodyzacji.

II. METODY BADANIA

Dobór metod taksonomicznych powinien być uzasadniony przede wszystkim specyfiką rozważanego problemu. Jednakże biorąc pod uwagę dużą na ogół pracochłonność związaną z ich stosowaniem, nie mniej ważną rolę odgrywają w tym względzie możliwości przeprowadzenia obliczeń za pomocą elektronicznych maszyn cyfrowych. Dlatego też w niniejszej pracy zastosowane zostały te spośród wielu znanych w literaturze algorytmów, dla których istnieją odpowiednie programy obliczeniowe.

Kilka pierwszych metod zostało zaadaptowanych do rozwiązania problemu periodyzacji bez jakichkolwiek modyfikacji. Dwie ostatnie natomiast zostały opracowane specjalnie z punktu widzenia możliwości wykorzystania ich do wyodrębniania jednorodnych faz rozwojowych. Uwzględnienie wielu różnych algorytmów taksonomicznych ma na celu z jednej strony ugruntowanie analizy merytorycznej, a z drugiej wzajemną weryfikację rezultatów, zwłaszcza w odniesieniu do nowych metod, które, jak dotychczas, nie były sprawdzone w wystarczającym stopniu.

Szczegółowe omówienie wykorzystywanych procedur taksonomicznych znaleźć można w cytowanej literaturze przedmiotu. W pracy przedstawione są tylko zasadnicze idee poszczególnych metod.

² N. Jardine, R. Sibson, *The Mathematical Taxonomy*, London 1972; H. Skarabis, *Mathematische Grundlagen und praktische Aspekte der Diskrimination und Klassifikation*, Würzburg 1970.

Metody G. M. Lance'a i W. T. Williamsa. G. M. Lance oraz W. T. Williams podali³ ogólną metodę taksonomiczną grupowania punktów w przestrzeni wielowymiarowej. Należy ona do grupy procedur aglomeracyjnych, w których poszczególne punkty traktuje się wstępnie jako podzbiory jednoelementowe. Proces klasyfikacyjny polega tu na stopniowej redukcji liczby podzbiorów drogą kolejnego łączenia skupisk położonych najbliżej siebie. Omawiana metoda ma kilka odmian w zależności od sposobu zdefiniowania odległości dwóch zbiorów w przestrzeni wielowymiarowej. Schemat postępowania ująć można w następujące punkty:

- 1) wyznacza się macierz odległości pomiędzy wszystkimi punktami przyjmując w tym celu wybraną metrykę odległości,
- 2) na podstawie minimalnego elementu macierzy odległości łączy się w jedną grupę S_r skupiska punktów S_p i S_q , leżące najbliżej siebie,
- 3) oblicza się odległość nowo utworzonej grupy S_r od wszystkich pozostałych skupisk S_b zgodnie z wzorem:

$$d_{ir} = \alpha d_{ip} + \beta d_{iq} + \gamma d_{pq} + \delta |d_{ip} - d_{iq}|, \quad (1)$$

gdzie: d_{st} — odległość pomiędzy skupiskami S_s oraz S_t ,

$\alpha, \beta, \gamma, \delta$ — parametry przekształcenia,

- 4) kontynuuje się operacje opisane w punktach 2 i 3, aż do momentu uzyskania jednej grupy obejmującej wszystkie klasyfikowane punkty.

Przyjmując odpowiednie wartości dla parametrów $\alpha, \beta, \gamma, \delta$ otrzymuje się pięć różnych wariantów metody. D. Wishart⁴ wykazał, że powyższy algorytm jest identyczny z procedurą zaproponowaną przez J. H. Warda⁵ i podał wzory parametrów przekształcenia właściwe dla

Tabela 1

Wzory parametrów przekształcenia

Nazwa metody	α	β	γ	δ
Najbliższego sąsiedztwa	0,5	0,5	0	-0,5
Najdalszego sąsiedztwa	0,5	0,5	0	0,5
Mediany	0,5	0,5	-0,25	0
Średniej grupowej	$\frac{k_p}{k_r}$	$\frac{k_p}{k_r}$	0	0
	$\frac{k_p}{k_r}$	$\frac{k_q}{k_r}$	$\frac{k_p \cdot k_q}{k_r}$	0
Środka ciężkości	$\frac{k_i + k_p}{k_i + k_r}$	$\frac{k_i + k_q}{k_i + k_r}$	$\frac{k_i}{k_i + k_r}$	0
	$\frac{k_i + k_p}{k_i + k_r}$	$\frac{k_i + k_q}{k_i + k_r}$	$\frac{k_i}{k_i + k_r}$	0

³ G. M. Lance, W. T. Williams, *A General Theory of Classificatory Sorting Strategies*, Computer Journal 1967, t. IX.

⁴ D. Wishart, *An Algorithm for Hierarchical Classifications*, Biometrics 1969, 1.1.

⁵ J. W. Ward, *Hierarchical Grouping to Optimise an Objective Function*, Journal of the American Statistical Association 1963, 58.

tego przypadku. Formuły odpowiadające poszczególnym wariantom zebrane zostały w tabeli 1.

W tabeli tej k_i oznacza liczbę punktów zawartych w grupie S_i . Różnice pomiędzy wyróżnionymi wariantami tkwią w odmiennym pojmowaniu odległości skupisk. W metodzie najbliższego sąsiedztwa (*nearest neighbour*) jest to najmniejsza odległość spośród wszystkich odległości pomiędzy punktami należącymi do różnych grup. W metodzie najdalszego sąsiedztwa (*furthest neighbour*) jest to odległość największa. W metodzie mediany (*median's method*) przyjmuje się w tym celu odległość środkową, w sensie mediany. W metodzie średniej grupowej (*group average*) odległość skupisk definiuje się za pomocą średniej arytmetycznej ze wszystkich odległości punktów należących do różnych grup. W metodzie środka ciężkości (*centroid method*) odległością tą jest różnica pomiędzy średnimi arytmetycznymi z odległości punktów w każdym z porównywanych skupisk. I wreszcie w metodzie J. H. Warda odległość skupień wyrażona jest różnicą sum kwadratów odchyłeń odległości poszczególnych punktów od środków ciężkości grup, do których punkty te należą.

Obszarowa dyskryminacja wrocławska. Metoda ta reprezentuje odrębną grupę algorytmów taksonomicznych, w których całą wielowymiarową przestrzeń dzieli się w pewien ustalony a priori sposób na rozłączne podprzestrzenie, zaś znajdujące się w nich punkty traktuje jako odrębne klasy⁶. Tok postępowania można tu przedstawić następująco:

- 1) rejestruje się przynależność poszczególnych punktów do hiperkul o arbitralnie ustalonym promieniu ϱ ⁷ oraz współrzędnych środka określonych przez zaobserwowane wartości współrzędnych danego punktu,
- 2) punkty zawarte w najliczniejszej hiperkuli uważa się za pierwszą grupę podobieństwa i wyłącza z dalszych rozważań,
- 3) w stosunku do pozostałych punktów powtarza się opisaną wyżej procedurę, aż do wyczerpania wszystkich elementów dyskryminowanego zbioru.

Taksonomia wrocławska. Metoda taksonomii wrocławskiej należy do starszych procedur, jednakże ciągle jest z powodzeniem wykorzystywana w różnorodnych badaniach. Ogólnie biorąc idea jej polega na płaszczyznowym odwzorowaniu punktów leżących w przestrzeni wielowymiarowej, w taki sposób, aby suma odległości rzutów tych punktów

⁶ W. Bukietyński i in., *Uwagi o dyskryminacji zbiorów skończonych*, Prace Naukowe WSE Wrocław, 1969, t. 21.

⁷ Autorzy metody proponują wyznaczać promień ϱ Z następującej relacji: $\varrho = \bar{d} \pm s_d$, gdzie \bar{d} jest średnią arytmetyczną z minimalnych odległości poszczególnych punktów od wszystkich pozostałych, natomiast s_d — odchyleniem standardowym z tychże wartości.

na płaszczyznę była najmniejsza⁸. Kolejne operacje związane z omawianą procedurą przedstawiają się następująco:

1) na podstawie macierzy odległości klasyfikowanych punktów wyznacza się parę punktów znajdujących się najbliżej siebie,

2) szuka się następnego punktu leżącego najbliżej jednego z punktów dotychczas wybranych i przyłącza do niego,

3) postępowanie takie powtarza się dotąd, aż wyczerpane zostaną wszystkie elementy rozpatrywanego zbioru punktów.

W rezultacie otrzymuje się graf spójny, nie zamknięty, który określa najkrótsze spośród wszystkich możliwych uporządkowań płaszczyznowych. Odrzucając kolejne, najdłuższe krawędzie grafu uzyskuje się żądany podział zbioru na dowolną liczbę rozłącznych i wyczerpujących podzbiorów, zawierających punkty maksymalnie podobne do siebie w ramach danego podzbioru oraz maksymalnie niepodobne w stosunku do punktów należących do innych podzbiorów.

Dendrytowa metoda R. C. Prima. Zbliżoną do taksonomii wrocławskiej metodę konstrukcji najkrótszego dendrytu punktów podał R. C. Prim⁹. Zbiór klasyfikowanych punktów dzielony jest na dwa podzbiory, z których pierwszy — A , zawiera punkty należące w danym momencie do tworzonego dendrytu, natomiast drugi — B , punkty pozostałe. Na początku zbiór A jest zbiorem pustym, natomiast B obejmuje wszystkie rozpatrywane punkty. Cały algorytm postępowania przedstawić można¹⁰ w następujących etapach:

1) do zbioru A włącza się dowolny punkt należący do B ,

2) oblicza się dystanse wybranego punktu od pozostałych, tworząc z nich zbiór C ,

3) szuka się najmniejszego elementu w zbiorze C , przyłączając przyporządkowany mu punkt do zbioru A ,

4) oblicza się odległości punktu aktualnie przyłączonego do A od wszystkich punktów znajdujących się w B , porównując je każdorazowo z odpowiadającymi im odległościami zawartymi w zbiorze C ,

5) jeśli odległość danego punktu jest mniejsza od dotychczasowej odległości figurującej w C , to dokonuje się ich zamiany, zostawiając w zbiorze C odległość mniejszą,

6) w przeciwnym przypadku wraca się do punktu 3, kontynuując takie postępowanie, dopóki w zbiorze A nie znajdą się wszystkie klasyfikowane punkty.

⁸ K. Florek i in., *Sur la liaison et la division des pointes d'un ensemble fini*, Colloquium Mathematicum 1951, 2; J. Perkal, *Taksonomia wrocławska*, Przegląd Antropologiczny 1953 t. 19.

⁹ R. C. Prim, *Shortest Connection Networks and Some Generalizations*, The Bell System Technical Journal 1957, t. 6.

¹⁰ J. C. Gower, G. J. S. Ross, *Minimum Spanning Trees and Single Linkage Cluster Analysis*, Applied Statistics 1969, t. 1.

W efekcie numery kolejno pojawiających się w zbiorze A punktów wyznaczają najkrótszy dendryt o długościach krawędzi określonych przez wartości elementów znajdujących się w ostatnim z otrzymanych zbiorów C . Sposób podziału uzyskanego grafu jest identyczny jak w taksonomii wrocławskiej.

Metoda B. Kinga. W odróżnieniu od dotychczas wymienionych metod, procedura B. Kinga¹¹ posłużyła wyłącznie do grupowania cech. Wynikało to stąd, że punktem wyjścia jest tu macierz współczynników korelacji liniowej R pomiędzy wszystkimi klasyfikowanymi cechami.

Algorytm metody ująć można w następujące punkty:

1) szuka się maksymalnego elementu macierzy korelacji r_{pq} i łączy odpowiadające mu cechy X_p oraz X_q w jedną grupę,

2) w miejsce współczynników korelacji pomiędzy cechą X_p a pozostałymi cechami wstawia się do macierzy R nowe wartości zdefiniowane wzorem:

$$r_{pi} = \frac{\text{cov}(x_p + x_q, x_i)}{[\text{var}(x_p + x_q) \cdot \text{var}(x_i)]^{\frac{1}{2}}}, \quad (2)$$

gdzie:

$\text{cov}(x_i, x_j)$ — ocena kowariancji pomiędzy zmiennymi X_i oraz X_j ,
 $\text{var}(x_i)$, $\text{var}(x_j)$ — oceny wariancji zmiennych X_i oraz X_j ,

$$x_i = x_p + x_q,$$

natomiast w miejsce współczynników korelacji pomiędzy zmienną X_q a resztą zmiennych przyjmuje się wartości mniejsze od -1 , eliminując w ten sposób możliwość uwzględnienia tej zmiennej w dalszych rozważaniach,

3) postępowanie opisane powyżej powtarza się, aż w macierzy R wszystkie elementy będą liczbami mniejszymi od -1 , zaś wszystkie sklasyfikowane cechy połączą się w jedną grupę.

Metoda dychotomicznego podziału. Dotychczas omówione metody taksonomiczne mają jedną zasadniczą wadę jeśli chodzi o ich zastosowanie w procesie periodyzacji. Nie uwzględniają one mianowicie warunku liniowego uporządkowania klasyfikowanych punktów. Tymczasem w periodyzacji rolę tych punktów odgrywają kolejne okresy czasu mające ściśle określony porządek. W związku z tym powstaje konieczność stworzenia procedur taksonomicznych omijających tę trudność.

Istota jednej z takich metod¹² polega na dychotomicznym podziale

¹¹ B. King, *Step-wise Clustering Procedures*, Journal of the American Statistical Association 1967, t. 62.

¹² T. Grabiński, *Numeryczne metody periodyzacji rozwoju obiektów gospodarczych*, Przegląd Statystyczny (w druku).

całego rozpatrywanego czasokresu analizy na wszystkie możliwe sposoby i obliczaniu za każdym razem uogólnionej miary odległości pomiędzy otrzymanymi z podziału wielocechowymi podzbiorami obserwacji¹³. W punkcie odpowiadającym maksymalnej wartości metryki odległości proponuje się przeprowadzić pierwszą granicę periodyzacyjną.

Z kolei identyczne postępowanie powtarza się w stosunku do każdego z dwóch aktualnie otrzymanych podokresów, otrzymując w efekcie dal-sze granice podziału. Proces taki kontynuuje się dotąd, dopóki maksymalna wartość miernika odległości nie spadnie poniżej pewnego ustalonego z góry progu podobieństwa¹⁴.

Dychotomiczny charakter procedury podziału może w pewnych sytuacjach prowadzić do wyodrębnienia podokresów nie różniących się istotnie pomiędzy sobą. Dlatego też uzyskane granice faz rozwojowych poddaje się weryfikacji, likwidując to z nich, dla których ponownie obliczony miernik odległości okaże się mniejszy od założonego progu podobieństwa.

Metoda graficzna. We wszystkich dotychczas zaprezentowanych procedurach niezbędne jest, przynajmniej w badaniach prowadzonych na szerszą skalę, korzystanie z elektronicznej techniki cyfrowej. Dlatego też skonstruowano prostą metodę pozwalającą, jak się okazało, na efektywne rozwiązanie zagadnienia periodyzacji, bez konieczności przeprowadzenia żmudnych obliczeń.

Pewien stosunkowo nieduży nakład pracy wymagany jest tylko we wstępnym etapie, polegającym na wyznaczeniu odległości pomiędzy sąsiadującymi w czasie okresami badania. Określają one sumaryczną wielkość zmian, jakie miały miejsce w rozwoju analizowanego obiektu społeczno-ekonomicznego w danym okresie. Szczególnie duże wartości tych mierników sugerują pojawienie się punktów zwrotnych w tendencji rozwojowej badanych procesów.

Odrzucając kilka największych wartości obliczonych mierników od-

¹³ Można w tym celu stosować na przykład uogólnioną odległość P. C. Mahalanobisa, *On the Generalized Distance in Statistics*, Proceedings of the National Institute of Sciences of India 1936, 12:

$$D^2 = \frac{n_1 \cdot n_2}{n_1 + n_2} (\bar{x}_1 - \bar{x}_2)' s^{-1} (\bar{x}_1 - \bar{x}_2), \quad (3)$$

gdzie: n_1, n_2 — ilość okresów czasu w każdym z wyodrębnionych podokresów; \bar{x}_1, \bar{x}_2 — wektory średnich arytmetycznych cech dla pierwszego i drugiego podokresu; S^{-1} — macierz odwrotna do macierzy wariancji i kowariancji wyznaczonej na podstawie łącznej próby.

¹⁴ W przypadku korzystania z mierników (3) oraz przyjęcia założenia, że zaobserwowane wartości cech pochodzą z populacji mającej wielowymiarowy, sferyczny rozkład normalny, odpowiednią wartość można odczytać z tablic rozkładu χ^2 , przy ustalonym poziomie istotności α oraz m (ilość cech) stopniach swobody; por. C. R. Rao, *Advanced Statistical Methods in Biometrie Research*, New York 1952.

ległości otrzymuje się automatycznie granice szukanych faz rozwojowych. Przy podejmowaniu decyzji o liczbie wyróżnionych punktów zwrotnych, a tym samym i ilości faz rozwojowych, wygodnie jest przedstawić mierniki odległości w postaci wykresu punktowego na osi liczbowej lub wykresu wartości mierników w odniesieniu do zmiennej czasowej.

III. PRÓBA USTALENIA FAZ ROZWOJOWYCH PROCESÓW DEMOGRAFICZNYCH NRD W LATACH 1949 - 1973

Podstawę badań stanowią dane liczbowe zaczerpnięte z Roczników Statystycznych NRD. Informacje te obejmują okres 25 lat, od r. 1949 do 1973, przy czym wstępnie przyjęto do analizy następujące cechy: małżeństwa na 1000 osób (X_1), urodzenia na 1000 osób (X_2), zgony na 1000 osób (X_3), przyrost naturalny na 1000 osób (X_4), urodzenia martwe na 1000 urodzeń żywych (X_5), zgony niemowląt na 1000 urodzeń żywych (X_6), udział urodzeń pozamażeńskich w urodzeniach żywych (X_7), roz-

Tabela 2

Wartości cech charakteryzujących procesy demograficzne NRD w latach 1949 - 1973

Lata	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1949	10,1	14,5	13,4	1,1	23,4	78,3	12,0	1,98	125	36,3	56,9	1,08
1950	11,7	16,5	11,9	4,6	21,7	72,2	12,9	2,70	125	36,0	56,0	1,39
1951	10,6	16,9	11,4	5,5	21,7	63,8	13,0	2,08	125	36,1	56,3	1,42
1952	9,6	16,7	12,1	4,6	21,2	59,1	13,1	1,80	124	36,1	56,4	1,38
1953	8,7	16,4	11,7	4,7	19,4	53,5	13,1	1,70	124	36,1	56,4	1,40
1954	8,4	16,3	12,2	4,1	19,1	50,3	13,3	1,60	124	36,0	56,2	1,34
1955	8,7	16,3	11,9	4,4	18,4	48,9	13,1	1,43	124	36,1	56,3	1,37
1956	8,6	15,9	12,0	3,9	18,1	46,5	13,3	1,32	124	36,4	57,1	1,33
1957	8,6	15,6	12,9	2,7	16,3	45,5	13,3	1,33	123	36,8	58,2	1,21
1958	8,9	15,6	12,7	2,9	15,9	44,2	12,4	1,33	123	37,2	59,1	1,23
1959	9,4	16,9	13,3	3,6	16,1	40,7	12,0	1,35	122	37,8	60,3	1,27
1960	9,7	17,0	13,6	3,4	16,0	38,8	11,6	1,42	122	38,7	63,0	1,25
1961	9,9	17,6	13,0	4,6	15,3	33,7	11,2	1,52	122	39,8	65,8	1,35
1962	9,7	17,4	13,7	3,7	14,5	31,6	10,1	1,46	121	40,4	67,8	1,27
1963	8,6	17,6	12,9	4,7	14,0	31,2	9,4	1,44	121	41,0	69,4	1,36
1964	8,0	17,2	13,3	3,9	13,4	28,6	9,5	1,48	119	41,6	71,2	1,29
1965	7,6	16,5	13,5	3,0	12,9	24,8	9,8	1,56	119	41,8	71,8	1,22
1966	7,1	15,7	13,2	2,5	12,5	22,9	10,0	1,64	119	42,0	72,5	1,19
1967	6,9	14,8	13,3	1,5	11,5	21,4	10,7	1,66	118	42,2	72,9	1,11
1968	7,0	14,3	14,2	0,1	11,3	20,2	11,5	1,68	118	42,2	73,0	1,01
1969	7,3	14,0	14,3	-0,3	10,8	20,3	12,4	1,70	118	42,2	73,0	0,98
1970	7,7	13,9	14,1	-0,2	10,4	18,5	13,3	1,65	117	42,1	72,7	0,99
1971	7,6	13,8	13,8	0,0	9,8	18,0	15,1	1,80	117	42,0	72,5	1,00
1972	7,8	11,8	13,8	-2,0	9,2	17,6	14,0	2,00	116	41,8	72,0	0,85
1973	8,1	10,6	13,7	-3,1	9,2	16,0	15,0	2,30	116	41,7	71,3	0,77

wody na 1000 osób (X_8), ilość kobiet na 100 mężczyzn (X_9), ilość osób w wieku nieprodukcyjnym na 100 osób (X_{10}), ilość osób w wieku nieprodukcyjnym na 100 osób w wieku produkcyjnym (X_{11}), współczynnik dynamiki demograficznej (X_{12}). Materiał statystyczny przedstawiono w tabeli 2.

W pierwszym etapie procesu periodyzacji dokonuje się redukcji wyjściowego zbioru cech, wykorzystując w tym celu przedstawione w poprzednim punkcie algorytmy taksonomiczne w ich postaci dualnej, z wyjątkiem dwóch ostatnich metod, które posłużyły wyłącznie do ustalenia faz rozwojowych. Przedmiotem klasyfikacji były tu poszczególne cechy traktowane jako zbiór 12 punktów w 25-wymiarowej przestrzeni lat.

Do obliczeń przyjęto następującą metrykę odległości cech:

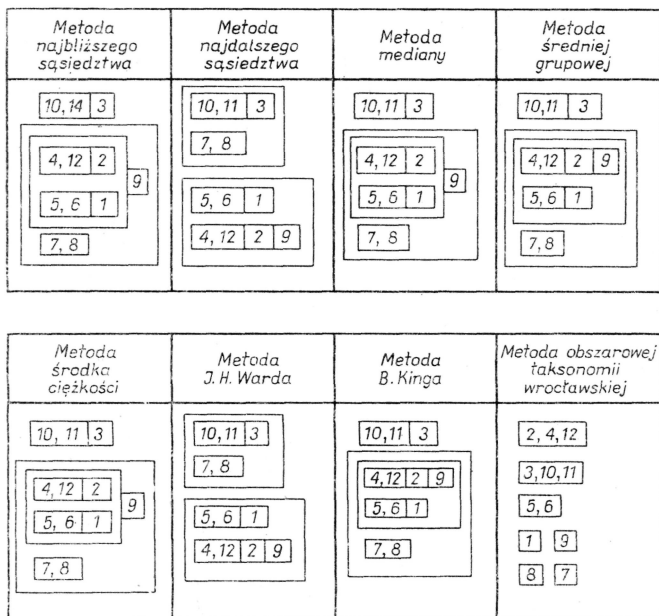
$$d_{ik} = \left[\frac{\sum_{j=1}^n (x'_{ij} - x'_{kj})^2}{n} \right]^{\frac{1}{2}}, \quad (4)$$

gdzie: x'_{ij} , x'_{kj} — standaryzowane wartości i -tej oraz k -tej cechy dla j -tego roku ($i, k=1, 2, \dots, n; i > k$),

n — ilość lat,

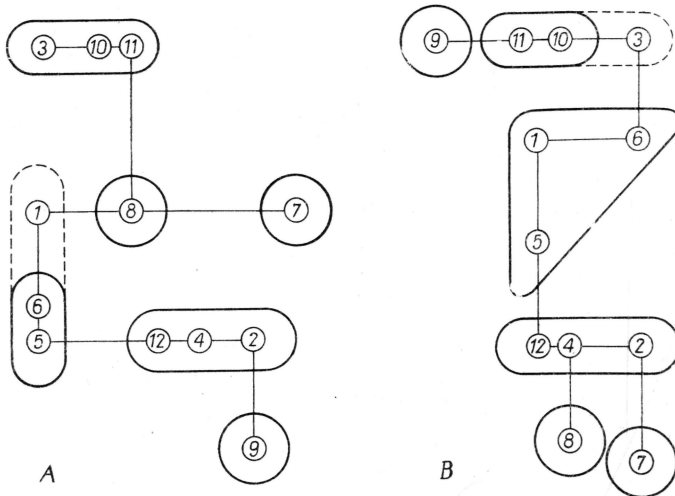
m — ilość cech.

Ostateczne rezultaty przedstawione zostały na ryc. 1 i 2. Ze względu na ograniczone ramy pracy nie zamieszczono wyników pośrednich, to jest macierzy wartości standaryzowanych oraz macierzy odległości cech.



Ryc. 1. Wyniki klasyfikacji 12 cech charakteryzujących rozwój procesów demograficznych w NRD w l. 1949 - 1973

Rycina 1 zawiera wyniki klasyfikacji cech w postaci dwóch grup cech podobnych, otrzymane za pomocą metod aglomeracyjnych oraz obszarowej taksonomii wrocławskiej. Chcąc uzyskać bardziej szczegółowe podziały należy odrzucać stopniowo najbardziej zewnętrzne ramki w poszczególnych schematach. Natomiast na rycinie 2 znajdują się dwa den-



Ryc. 2. Dendryt wrocławski (A) oraz R C Prima (B) wyjściowego zbioru cech. $1 \text{ cm}=0,4$

dryty badanych cech (A — dendryt wrocławski, B — dendryt R. C. Prima), na których zaznaczone są wyróżniające się ugrupowania cech.

Analizując wyniki w przekroju poszczególnych procedur, zauważyć można wysoką zbieżność otrzymanych klasyfikacji. Wszędzie pojawiały się cztery zasadnicze grupy, z których pierwsza zawierała współczynniki dynamiki demograficznej (X_{12}), przyrostu naturalnego (X_4) oraz urodzeń (X_2), druga — wskaźniki struktury ludności według wieku (X_{10} , X_{11}) wraz ze współczynnikiem zgonów (X_3), trzecia — wskaźnik małżeństw (X_1), urodzeń martwych (X_5) i zgonów niemowląt (X_6), natomiast czwarta — ilość rozwodów (X_8) w połączeniu z udziałem urodzeń pozamałżeńskich (X_7). Cecha (X_9) — struktura ludności według płci jest raczej niepodobna do żadnej z wyróżnionych wyżej grup.

Uwzględniając otrzymane rezultaty taksonomii cech, do dalszych badań wybrano zestaw 4 najbardziej typowych cech: małżeństwa na 1000 osób (X_1), urodzenia na 1000 osób (X_2), zgony na 1000 osób (X_3) oraz rozwody na 1000 osób (X_8). Uważa się, że cechy te opisują w wystarczającym stopniu całokształt procesów demograficznych w NRD, a przynajmniej najbardziej charakterystyczne ich elementy. Natomiast pozostałe — powtarzają tylko informacje tkwiące już w wybranych cechach, w związku z czym przyjęcie ich do analizy jest niecelowe, a nawet może zniekształcić wyniki końcowe periodyzacji wskutek zróżnicowanego trak-

towania poszczególnych kryteriów delimitacji. W takiej bowiem sytuacji, wydzielanie faz rozwojowych dokonywane byłoby na podstawie jednego wybijającego się kryterium, a więc w istocie rzeczy byłaby to procedura jednocechowa, a nie wielocechowa.

Na marginesie zaznaczyć należy, iż wybrane cechy zaliczane są w statystyce demograficznej rzeczywiście do najważniejszych, o czym świadczy fakt zamieszczania właśnie tych informacji we wszelkiego rodzaju publikacjach statystycznych. Wyniki analizy taksonomicznej stanowią formalne potwierdzenie i uzasadnienie tego rodzaju postępowania.

Znamienny jest również fakt przyłączenia do stopy urodzeń współczynników dynamiki demograficznej i przyrostu naturalnego, będących funkcją nie tylko wskaźnika urodzeń, ale i wskaźnika zgonów. Oznacza to, że zmiany w kształtowaniu się współczynników dynamiki i przyrostu determinowane były bardziej przez czynnik urodzeń aniżeli zgonów. W związku z tym przyjęcie do analizy jednego z omawianych współczynników jako substytutu stopy urodzeń i stopy zgonów nie jest poprawne, ponieważ zwiększa się w ten sposób wagę urodzeń, a zmniejsza znaczenie zgonów.

Z kolei przeprowadzono periodyzację rozwoju procesów demograficznych w NRD za pomocą wszystkich opisanych w punkcie 2 metod. Wyodrębnianie faz rozwojowych dokonywane było w dwóch wariantach: w wariancie *A*, traktowanym jako zasadniczy, oparto się na zredukowanym zbiorze 4 cech, natomiast w wariancie *B*, jako podstawę przyjęto cały wyjściowy zbiór obserwacji, obejmujący wartości 12 cech. Postępowanie takie miało na celu uwidocznienie skutków związanych z opuszczeniem wstępnego etapu doboru cech.

W trakcie obliczeń korzystano z następującego miernika odległości lat:

$$d_{ik} = \left[\frac{\sum_{j=1}^m (x'_{ij} - x'_{kj})^2}{m} \right]^{\frac{1}{2}}, \quad (5)$$

gdzie:

x'_{ij} , x'_{kj} — standaryzowane wartości j -tej cechy dla i -tego oraz k -tego roku ($i, k = 1, 2, \dots, n; i > k$),
 n, m — jak we wzorze (4).

Z uwagi na ograniczoną objętość pracy nie zamieszcza się macierzy odległości lat, jak również pomija się wykresy schematów klasyfikacyjnych i dendrytów otrzymanych w poszczególnych procedurach taksonomicznych. Szerzej potraktowana zostanie tylko graficzna metoda periodyzacji, ponieważ dotychczas nigdzie jej nie prezentowano.

Obliczane ze wzoru (5) wartości mierników odległości pomiędzy sąsiadującymi okresami $d_{t,t+1}$ ($t = 1, 2, \dots, n-1$) zawarte są w tabeli 3. Wykresy obrazujące kształtowanie się tych wartości w czasie podane są na ryc. 3 oraz 5, odpowiednio dla wariantu *A* i *B*. Rozkłady zaobserwo-

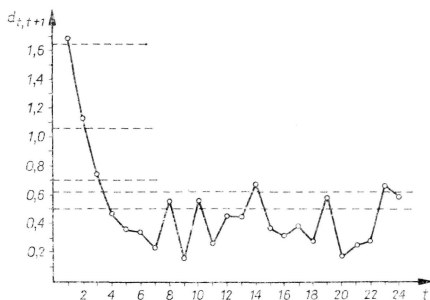
Tabela 3

Wartości mierników odległości $d_{t,t+1}$ pomiędzy sąsiadującymi latami

Lata	Wariant A	Wariant B	Lata	Wariant A	Wariant B
1949/50	1,691	1,208	1961/62	0,448	0,391
1950/51	1,113	0,670	1962/63	0,676	0,465
1951/52	0,742	0,456	1963/64	0,375	0,291
1952/53	0,484	0,322	1964/65	0,315	0,261
1953/54	0,365	0,254	1965/66	0,382	0,244
1954/55	0,344	0,218	1966/67	0,281	0,287
1955/56	0,218	0,172	1967/68	0,569	0,434
1956/57	0,555	0,429	1968/69	0,168	0,207
1957/58	0,176	0,209	1969/70	0,224	0,215
1958/59	0,565	0,368	1970/71	0,299	0,376
1959/60	0,249	0,221	1971/72	0,659	0,559
1960/61	0,441	0,397	1972/73	0,595	1,457

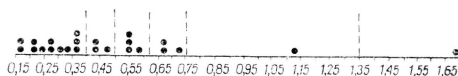
Źródło: Obliczenia własne.

wanych mierników odległości przedstawiają ryc. 4 i 6. Sposób ustalenia punktów zwrotnych uwidoczniony jest we wszystkich wykresach za pomocą przerywanych linii.



Ryc. 3. Wyniki graficznego wyodrębnienia faz rozwojowych według wariantu A.
— — — linie graniczne

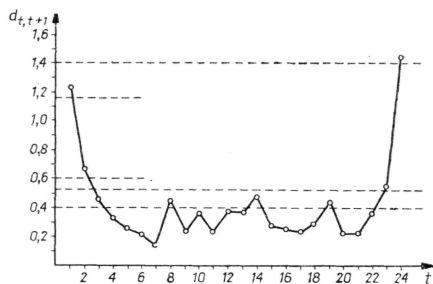
Końcowe rezultaty periodyzacji rozwoju zjawisk ludnościowych w NRD uzyskane za pomocą poszczególnych metod zebrane zostały dla wariantu A analizy na ryc. 7, natomiast dla wariantu B — na ryc. 8. Hierarchię ważności granic wyznaczających kolejne fazy rozwojowe otrzymuje się na podstawie kształtowania się wartości progów podobieństwa, przy których pojawiała się dana granica. Dla porównania, w główkach



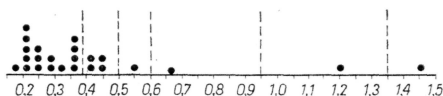
Ryc. 4. Rozkład mierników odległości sąsiednich lat $d_{t,t+1}$ według wariantu A

obydwu tabel zaznaczono fazy w rozwoju demograficznym Polski ustalone na podstawie zbliżonego zestawu cech.

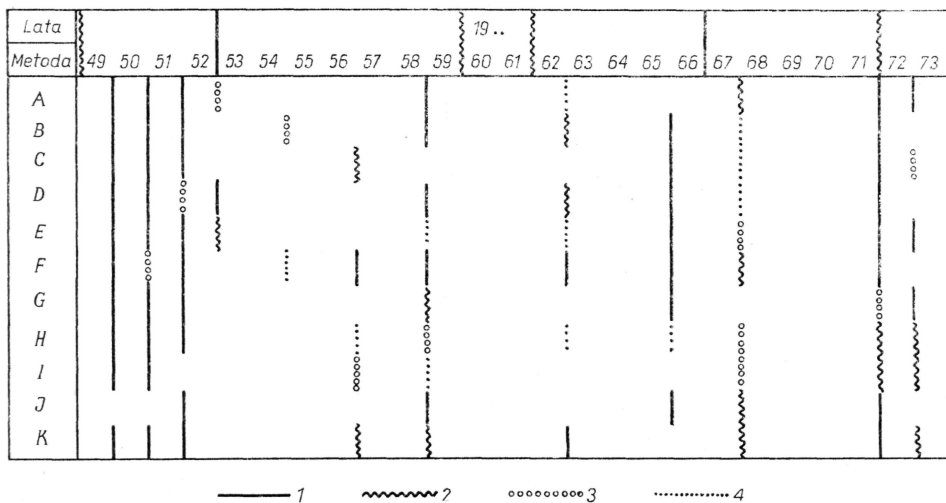
Przełglądając wyniki periodyzacji można generalnie stwierdzić zgodność punktów zwrotnych uzyskanych za pomocą różnych procedur, za-



Ryc. 5. Wyniki graficznego wyodrębnienia faz rozwojowych według wariantu B.
— — — linie graniczne



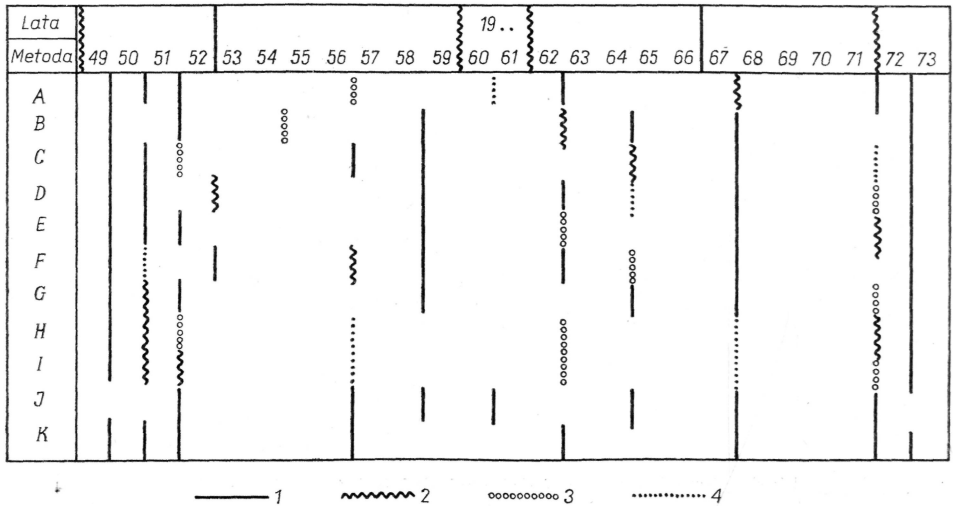
Ryc. 6. Rozkład mierników odległości sąsiednich lat $t, t+1$ według wariantu B.



Ryc. 7. Granice wyróżnionych faz rozwoju demograficznego NRD w l. 1949 - 1973 według Wariantu A.

A — metoda najbliższego sąsiedztwa; B — metoda najdalszego sąsiedztwa; C — metoda mediany; D — metoda średniej grupowej; E — metoda środka ciężkości; F — metoda J. H. Warda; G — obszarowa taksonomia wrocławska; H — taksonomia wrocławska; I — metoda R. C. Prima; J — metoda dychotomicznego podziału; K — metoda graficzna; 1-4 — skala ważności granic według hierarchii ważności (w główce tabelki zaznaczone są fazy rozwojowe dla Polski)

równy co do ich ilości, umiejscowienia w czasie, jak i ważności. Sporadycznie pojawiające się granice (1952/53 i 1954/55) uznać należy raczej za nieistotne. Wypada zauważyć, że ten ostatni wniosek można sformułować tylko dzięki uwzględnieniu w analizie wielu metod.



Ryc. 8. Granice wyróżnionych faz rozwoju demograficznego NRD w l. 1949 - 1973 według wariantu B.

A — metoda najbliższego sąsiedztwa; B — metoda najdalszego sąsiedztwa; C — metoda mediany; D — metoda średniej grupowej, E — metoda środka ciężkości; F — metoda J. H. Warda; G — obszarowa taksonomia wrocławska; H — taksonomia wrocławska; I — metoda R. C. Prima; J — metoda dychotomicznego podziału; K — metoda graficzna; 1-4 — skala ważności granic według hierarchii ważności (w główce tabelki zaznaczone są fazy rozwojowe dla Polski)

Podobnie przedstawia się sprawa jeśli chodzi o rezultaty otrzymane w przekroju obydwu wariantów badań — A i B. Różnią się one tylko w jednym przypadku, a mianowicie granica z roku 1965/66 w wariantcie A przesunęła się w wariantcie B o rok wcześniej. Częstsze są natomiast zmiany ważności poszczególnych granic, co dotyczy szczególnie punktów zwrotnych zaobserwowanych w latach 1967/68, 1971/72 oraz 1972/73.

Przechodząc do szczegółowej analizy łatwo zauważyć, że w początkowym okresie 1949-1951 w rozwoju procesów demograficznych w NRD zachodziły szybkie i istotne zmiany. W latach następnych dostrzega się większą stabilizację w kształtowaniu się badanych zjawisk. Wyróżnić tu można trzy zasadnicze fazy rozwojowe, przy czym pierwsza z nich obejmowała lata 1952-1958, druga 1959-1965 i trzecia — 1966-1971. Granice o mniejszym znaczeniu pojawiają się w latach 1956/57, 1962/63, 1967/68 oraz 1972/73. Wydaje się, że są one sygnałem pewnych zmian strukturalnych, które nieco później dają znać o sobie w postaci istotnych

załamań w kierunku tendencji rozwojowej rozważanych procesów ludnościowych.

Porównując fazy rozwoju demograficznego dla Polski i NRD trudno jest dopatrywać się dużego ich podobieństwa. Świadczyć to może za hipotezą, że procesy demograficzne w obydwóch krajach kształtują się pod wpływem różnego splotu przyczyn głównych lub jeśli nawet uwarunkowane są tymi samymi czynnikami systematycznymi, to czynniki te wykazują istotny wpływ w jednym kraju w innych okresach czasu, niż w drugim.

IV. ZAKOŃCZENIE

Z przeprowadzonych badań nasuwają się następujące wnioski ogólne:

1) metody taksonomii numerycznej okazały się przydatne w zagadnieniach związanych z periodyzacją, w przypadku gdy za jednostkę grupowania przyjęto okresy czasu, co stanowi novum w stosunku do klasycznego podejścia, spotykanego w dotychczasowych badaniach,

2) za pomocą zaproponowanych metod wydzielono w rozwoju zjawisk demograficznych w NRD istotnie zróżnicowane fazy rozwojowe, potwierdzając tym samym hipotezę, że obserwacje pochodzące z dłuższych okresów czasu nie stanowią jednorodnej zbiorowości,

3) ważnym problemem pozwalającym na uzyskanie bardziej poprawnej periodyzacji jest optymalny dobór cech do analizy, które to zagadnienie można rozwiązać również przy użyciu procedur taksonomicznych w ich postaci dualnej,

4) w ślad za przedstawionymi w opracowaniu badaniami powinna iść analiza merytoryczna, mająca na celu szczegółowe uzasadnienie otrzymanych rezultatów, co jednak stanowi zadanie przedstawicieli nauk przedmiotowych.

SOME NEW METHODS IN RESEARCH OF THE STRUCTURE AND DYNAMICS OF DEMOGRAPHIC PROCESSES IN GDR

Summary

The purpose of this paper is to present some numerical methods for objective determination of more homogeneous phases in demographic processes in GDR from 1947 to 1973.

Periods described by 12 attributes are the subject of discrimination. Such an approach opens new possibilities in time series analyses. Several taxonomic algorithms were used, along with two completely new numerical methods. Using these methods, the authors mark out substantially different phases in the development of demographic processes in GDR.