

Podstawowe rodzaje tablic statystycznych w kontekście racjonalnego wykorzystania danych przestrzennych

Jerzy J. Parysek^a 

Streszczenie. Celem artykułu jest prezentacja odpowiednio zestawionych tablic danych przestrzennych oraz sposobu ich wykorzystania w badaniach przestrzenno-strukturalnych i funkcjonalnych. Badania dynamiczne i strukturalne nawiązują do tego, co jest istotą rzeczywistości, w jakiej żyjemy, cechującej się zmiennością zdarzeń i procesów w czasie i przestrzeni. Zebrane, niekiedy z trudem, informacje liczbowe nie zawsze są optymalnie wykorzystywane w badaniu coraz bardziej złożonej rzeczywistości. Uporządkowane w tablicach dane stanowią punkt wyjścia do prowadzenia badań dynamiczno-strukturalnych przy wykorzystaniu znanych metod analizy statystycznej. W opracowaniu przedstawiono trzy rodzaje tablic (macierzy) danych: strukturalne (jedno- i wielocechowe), funkcjonalne i relacyjne (interakcyjne). Omówiono ich zastosowanie oraz wskazano podstawowe, możliwe do wykonania procedury obliczeniowe.

Słowa kluczowe: dane przestrzenne, badania dynamiczne, badania strukturalne, tablice strukturalne, tablice funkcjonalne, tablice relacyjne

Basic types of statistical tables as regards rational usage of spatial data

Summary. The article attempts to present a set of spatial data in appropriate tables with the ways they might be used in both spatial and structural research as well as in functional one. Dynamic and structural research allows for investigating a constantly changing reality. Collected numerical information is often ineffectively used in the research in a more complex reality. Organising collected spatial data in such tables makes it possible to apply various kinds of quantitative methods for the analyses. The article also presents certain principles of variable selection allowing not only for correct application of specific statistical methods, but also for the achievement of reliable results. Uni- and multivariate structural tables, functional tables and relational ones (interactive) are presented in the article, indicating simultaneously operations that can be carried out on the data collected in these tables with specific methods of quantitative analysis.

Keywords: spatial data, dynamic research, structural studies, structural tables, functional tables, relational tables

JEL: C000, C650

^a Uniwersytet im. A. Mickiewicza w Poznaniu – profesor emerytowany.

Rzeczywistość, w jakiej żyjemy, jawi się jako coraz bardziej złożony system, o którego rozwoju i funkcjonowaniu decyduje coraz więcej czynników i uwarunkowań. Dla jej zbadania, opisanie i wyjaśnienia potrzebne są zatem coraz pojemniejsze zbiory różnego rodzaju informacji, w tym także danych liczbowych. Wobec takiego wyzwania stają przedstawiciele wielu dyscyplin naukowych, także ci, których przedmiotem badania jest rzeczywistość społeczno-gospodarcza. Praktyką staje się więc, często mozolne, poszukiwanie i zdobywanie konkretnych danych liczbowych, pozwalających w możliwie największym stopniu uchwycić złożoność przedmiotu badań. Ważne, aby były one wiarygodne, ponieważ jest to gwarancją poprawnej interpretacji wyników, i aby umożliwiały zastosowanie najważniejszych metod analizy ilościowej. Po spełnieniu tych warunków, czyli po zebraniu odpowiednich danych liczbowych, można przystąpić do prowadzenia badań, których zakres rzeczowy będzie wyznaczony przez zebrane informacje.

Niestety częste są sytuacje nieefektywnego wykorzystania zgromadzonych danych, a przez to ograniczenia zakresu rzeczowego badań. Poziomemu rozwojowi technik informacyjno-komunikacyjnych pozwala na coraz łatwiejsze zdobycie potrzebnych informacji, a zarazem na poszerzenie zakresu rzeczowego prowadzonych badań. Dane, jakie są dostępne, umożliwiają coraz pełniejsze zbadanie i opisanie złożonej rzeczywistości. Zastanawiają zatem sytuacje, kiedy zebrane dane liczbowe wykorzystuje się w ograniczonym stopniu.

Celem artykułu jest prezentacja właściwie zestawionych tablic danych przestrzennych oraz sposobu ich wykorzystania w badaniach przestrzenno-strukturalnych i funkcjonalnych. W opracowaniu wskazuje się, jak zestawienie posiadanych danych liczbowych w odpowiednich tablicach może prowadzić nie tylko do ich racjonalnego uporządkowania, lecz także zwiększenia możliwości szerszego ujęcia podejmowanej problematyki badawczej, co w pewnym stopniu jest wymuszone przez coraz większą złożoność rzeczy, zdarzeń i procesów będących przedmiotem prowadzonych badań naukowych.

Opracowanie w sposób szczególny uwzględni społeczno-ekonomiczne badania przestrzenne, czyli takie, których przedmiot stanowią zbiory jednostek przestrzennych, przede wszystkim jednostek terytorialnego podziału kraju, takie jak: gmina, powiat, województwo, kraj, a także różnego rodzaju subregiony i makroregiony (np. aglomeracje oraz jednostki NUTS statystyki regionalnej UE). Nie oznacza to jednak wykluczenia przedstawionych propozycji z badań dotyczących innych jednostek badawczych, np. miast. W wierszach proponowanych tablic danych przestrzennych brane są pod uwagę jednostki przestrzenne (i -te jednostki; $i = 1, 2, \dots, m$), a w kolumnach – charakteryzujące je cechy opisane za pomocą odpowiednio dobranych zmiennych (j -te cechy; $j = 1, 2, \dots, n$), choć nie dotyczy to wszystkich tablic.

W artykule przedstawiono wybrane kategorie możliwych do zestawienia tablic danych przestrzennych, wskazano przykłady ich zastosowania i podano możli-

wości dokonania, na podstawie zawartości tych tablic, odpowiednich obliczeń¹. Otwiera to duże, często niedostrzegane lub pomijane, możliwości interpretacyjne uzyskanych wyników. Pewne kategorie proponowanych tablic noszą w literaturze (zwłaszcza geograficznej) nazwy macierzy geograficznych lub macierzy Berry'ego, który prawdopodobnie jako pierwszy wskazał na korzyści płynące z tego rodzaju uporządkowania danych przestrzennych, mając na uwadze ich wykorzystanie w analizie statystycznej (Chojnicki i Czyż, 1973).

Przedstawione propozycje pokazują, oprócz znanych i stosowanych, nowe podejścia badawcze, pozwalające sprostać współczesnym wyzwaniom i zmieniającej się rzeczywistości. Artykuł nie pretenduje jednak do bycia wyczerpującym omówieniem problematyki ani prezentacją najnowszych metod analizy ilościowej możliwych do zastosowania w odniesieniu do danych zestawianych w tablicach. Nie jest też krytyczną oceną metod stosowanych w badaniach przestrzenno-ekonomicznych. Ma przede wszystkim charakter praktyczny i edukacyjny.

BADANE OBIEKTY I ICH CECHY

Podjęta problematyka badawcza uwzględnia zbiór badanych obiektów i charakteryzujące je cechy. Przedmiotem badań przestrzennych mogą być zarówno obiekty pseudonaturalne, do których należą np. jednostki terytorialnego podziału kraju², jak i sztuczne jednostki przestrzenne, np. o kształcie kwadratu czy sześcioboku foremnego, wyznaczone w ramach sieci przyjętej dla danego badania. Zbiory jednostek pseudonaturalnych są przyjmowane powszechnie, natomiast jednostki sztuczne, mimo wielu zalet, stosuje się dość rzadko, także z uwagi na trudności związane z zebraniem odpowiednich danych liczbowych. Niezależnie od rodzaju obiektów ważne jest jednak to, aby dla jednego badania zestawiać obiekty tej samej rangi czy kategorii (gminy, powiaty, województwa itp. lub kwadraty czy inne wieloboki foremne) i aby tworzyły one zbiór adekwatny i rozłączny, co gwarantuje poprawność analizy ilościowej i interpretacji wyników (Parysek, 1982). Zakłada się bowiem, że w przypadku zbiorów uporządkowanych danych mogą być stosowane odpowiednie metody analizy ilościowej, od metod prostych do bardziej zaawansowanych.

Właściwości badanych obiektów są opisywane przez adekwatne do problemu cechy, a właściwie zmienne odnoszące się do tych cech. Charakter i rodzaj zmiennych określa jednak nie tylko podejmowana problematyka badawcza, lecz

¹ Szczególnym punktem odniesienia są tablice danych społeczno-ekonomicznych, które mogą mieć zastosowanie, gdy zbiór badanych obiektów charakteryzowany jest przez większą liczbę cech.

² Gminy, powiaty, województwa oraz kraje trudno nazwać jednostkami naturalnymi, ponieważ nie mają takiego charakteru jak krainy geograficzne, zlewnie, łańcuchy górskie czy jeziora. Są jednak pewnymi całościami, wyróżnionymi arbitralnie poprzez konkretny przebieg ich granic.

także konkretnie sformułowany cel badań, ich przedmiot, jak również przewidziane do wykorzystania metody analizy ilościowej. Powadzi to do konkretyzacji zarówno cech, które bierze się pod uwagę, jak i zmiennych opisujących te cechy. W pierwszej kolejności należy więc dokonać wyboru obiektów najodpowiedniejszych do planowanych badań (o czym już wspomiano), a następnie odpowiednich do celu badań cech charakteryzujących te obiekty i zmiennych, które możliwie najlepiej opisują brane pod uwagę cechy i jednocześnie są odpowiednie do zaplanowanych do wykorzystania metod ilościowych.

Spełnienie opisanych wymagań może zagwarantować jedynie dobra znajomość przedmiotu badań (gospodarki, rozwoju społecznego, rolnictwa, usług, infrastruktury technicznej itp.) oraz metod. Aby uzyskać wiarygodne wyniki, należy, oprócz znajomości przedmiotu badań, brać pod uwagę właściwości cech opisujących badane obiekty (zmiennych) jakie są wymagane przy wykorzystaniu konkretnej metody, zwłaszcza wielozmiennej analizy statystycznej (Anderson, 1958; Morrison, 1967).

Zmienne wyrażające cechy w tablicy danych przestrzennych powinny zatem (Orloci, 1967; Parysek, 1982):

- być wyrażone we właściwych skalach pomiaru (ilorazowa, porządkowa, nominalna) za pomocą odpowiednich rodzajów zmiennych (zmienne ciągłe, dyskretne, binarne);
- być sumowalne, co np. może zapewnić standaryzacja cech;
- być niezależne, czyli nieskorelowane ze sobą lub skorelowane na niskim poziomie istotności współczynnika korelacji, co jest możliwe do uzyskania przy wykorzystaniu np. analizy składowych głównych;
- nie być kombinacjami liniowymi innych uwzględnianych w badaniach zmiennych;
- być relatywne, czyli odniesione do innych właściwości badanych obiektów, np. liczby ludności, powierzchni itp.;
- mieć znany rozkład statystyczny.

Ważne jest także, aby cechy wyraźnie różnicowały badane obiekty, czyli aby charakteryzowała je dość znaczna (wysoka) wariancja³.

Dane liczbowe wykorzystywane w badaniach przestrzennych dość rzadko spełniają wszystkie wymienione warunki, zwłaszcza pod względem sumowalności i niezależności cech, a także znanego rozkładu statystycznego. Sumowalność można zapewnić, dokonując standaryzacji cech oryginalnych przez odpowiednie ich przekształcenie. Niezależność cech uzyskuje się natomiast dzięki transformacji cech oryginalnych w składowe główne. Zależność lub współzależność można także ograniczyć, obliczając współczynniki korelacji pomiędzy parami zmiennych, a następnie odrzucając te silnie ze sobą skorelowane, mniej

³ W literaturze można znaleźć wiele metod doboru właściwego zbioru cech (np. Młodak, Józefowski i Wawrowski, 2016).

znaczące z punktu widzenia prowadzonych badań. Co się tyczy wariancji, to do eliminacji zmiennych mało różnicujących zbiorów badanych obiektów mogą posłużyć wyliczone współczynniki zmienności. Podejmując badania, warto zatem pamiętać, że poprawny dobór cech decyduje o jakości i wiarygodności uzyskanych wyników (Parysek, 1980, 1982; Parysek i Ratajczak, 1978, 2002).

Zbierane dane liczbowe można zestawiać w różnych tablicach danych. O ich rodzaju i charakterze przesądzają cel prowadzonych badań oraz metody ilościowe, jakie zamierza się zastosować. Różne rodzajowo tablice dają możliwości określania, z jednej strony dynamiki, a z drugiej strony – struktury zbioru badanych obiektów, co ułatwia zastosowanie metod ilościowych. Jest to szczególnie ważne, kiedy przyjmuje się, że do podstawowych cech badanej rzeczywistości należą jej zmienność czasowa i zróżnicowanie przestrzenne (Chojnicki i Czyż, 1973; Parysek, 1982).

TABLICE STRUKTURALNE

Najczęściej tworzoną tablicą danych przestrzennych jest tablica strukturalna⁴ – jednocechowa lub wielocechowa. Tablica jednocechowa ma zastosowanie w badaniu dynamiki, umożliwia zatem badanie zmian w czasie (tj. zdarzeń i procesów), a także stanów, natomiast tablica wielocechowa jest wykorzystywana do określania przestrzennego zróżnicowania badanego obszaru i pozwala na badanie struktur przestrzennych.

Ogólny model obu rodzajów tablic jest taki sam; w istocie rzeczy różnią się one zawartością kolumn.

Tablice jednocechowe

Tablica jednocechowa ma wymiary $m \times n$, gdzie m jest liczbą branych pod uwagę obiektów ($i = 1, 2, 3, \dots, m$), a n – liczbą uwzględnianych cech (zmiennych opisujących cechy). Element y_{ij} tej tablicy jest zatem wartością j -tej zmiennej (cechy) dla i -tego obiektu.

W kolumnach tablicy strukturalnej jednocechowej (dynamicznej) znajdują się wartości tylko jednej zmiennej uwzględnianej w badaniu (np.: gęstość zaludnienia, odsetek ludności miejskiej, liczba podmiotów gospodarczych na 1000 mieszkańców, zbiory zbóż z 1 ha, przeciętne wynagrodzenie, stopa bezrobocia itp.), zmierzone dla badanych obiektów (elementów zbioru badanych obiektów) w kolejno po sobie następujących latach, natomiast w wierszach – wartości danej zmiennej zmierzone w kolejno po sobie następujących latach dla poszczegól-

⁴ Nazwę „tablica strukturalna” (w odróżnieniu od tablicy funkcjonalnej) przyjęto, uznając za strukturę zbiór obiektów i relacji pomiędzy nimi.

gólnych obiektów badanego zbioru. Jednocechowa tablica dynamiczna pozwala więc, oprócz podania ogólnej informacji, zbadać i opisać dokonujące się zmiany, tj. zdarzenia i procesy, w odniesieniu do obiektów badanego zbioru.

Postać, jaką przyjmuje tablica strukturalna jednocechowa, przedstawia tabl. 1.

TABL. 1. MODEL JEDNOCECHOWEJ TABLICY STRUKTURALNEJ

Obiekty	Cecha (zmienna) zmierzona dla kolejnych lat				
	1	2	3	...	<i>n</i>
1	y_{11}	y_{12}	y_{13}	...	y_{1n}
2	y_{21}	y_{22}	y_{23}	...	y_{2n}
3	y_{31}	y_{32}	y_{33}	...	y_{3n}
...
<i>m</i>	y_{m1}	y_{m2}	y_{m3}	...	y_{mn}

Źródło: opracowanie własne.

Za pomocą tablicy jednocechowej (dynamicznej) można badać:

- zdarzenia (zmiany w analizowanym okresie) dotyczące konkretnego obiektu poprzez porównanie ze sobą dwóch kolejnych elementów (zmiennych) danego wiersza tablicy, tj. przyjętego za początkowy i końcowy moment wybranego okresu, w skrajnym przypadku – pierwszego i ostatniego roku danego szeregu czasowego. Najczęściej oblicza się albo wskaźniki dynamiki zmian (W), dzieląc wartość cechy w roku uznanym za końcowy przez wartość cechy w roku uznanym za początkowy i mnożąc uzyskany wynik przez 100 (otrzymuje się wskaźnik wyrażony w %), albo stopę wzrostu (S), również wyrażoną w %, odejmując 100 od wskaźnika dynamiki ($S = W - 100$);
- zróżnicowanie (wariancję) branej pod uwagę zmiennej (cechy) obiektów badanego zbioru w konkretnym roku, także w kolejnych latach, obliczając standaryzowany wskaźnik wariancji, czyli współczynnik zmienności, będący ilorazem odchylenia standardowego i średniej pomnożonym przez 100%;
- podobieństwo stanów opisujących wartości zmiennych (cech) dla poszczególnych obiektów w analizowanych dwóch momentach czasu, które wyraża współczynnik korelacji wyliczony dla wartości z dwóch kolumn tablicy odpowiadających okresowi badań;
- podobieństwo procesów zmian w analizowanych parach obiektów, obliczając współczynniki korelacji dla zmiennych w wierszach zestawionego szeregu czasowego (wierszach tablicy);
- kierunki i zakres zmian strukturalnych w zbiorze badanych obiektów w kolejnych latach, na co wskazują współczynniki zmienności wyliczone dla kolejnych lat (dla kolumn).

Z obliczonych wskaźników i miar statystycznych można tworzyć kolejne tablice strukturalne pozwalające określić podobieństwo zachodzących zdarzeń i procesów. Będą to zarówno tablice $m \times n$ (np. dla dynamiki zmian z roku na rok), jak i tablice $m \times m$ dla podobieństwa obiektów, ze względu na procesy, jakie zaszły w uwzględnionym okresie badań (korelacje między wartościami poszczególnych wierszy) oraz $n \times n$ dla określenia podobieństwa stanów w branych pod uwagę cięciach czasowych (korelacje między wartościami poszczególnych kolumn).

Uzyskane wyniki otwierają szerokie pole wnioskowania w przedmiocie prowadzonych badań. Często badania koncentrują się jedynie na opisie rzeczywistości z pominięciem tak ważnego wyjaśniania opisanych sytuacji.

Tego rodzaju tablice strukturalne można tworzyć dla szeregów czasowych konkretnych cech (zmiennych) zbioru badanych obiektów (jednostek przestrzennych). Istnieje wiele przykładów zestawień konstruowanych dla konkretnych badań prowadzonych w przekroju jednostek terytorialnego podziału kraju (gminy, powiaty, województwa) oraz w wymiarze międzynarodowym. Są to tablice dokumentujące zmiany w kolejnych latach wybranych wskaźników rozwoju społeczno-gospodarczego jednostek terytorialnych, a zatem obrazujące rozwój i tendencje rozwojowe, np. tablice zmian w kolejnych latach: liczby ludności, poziomu PKB, stopy bezrobocia, poziomu cen towarów i usług, skupu produkcji rolnej z 1 ha, przeciętnych wynagrodzeń, liczby lekarzy na 1000 mieszkańców itp.⁵

Należy dodać, że dane zawarte w tablicach oraz przeprowadzone obliczenia mogą być przedstawione w formie graficznej (zwłaszcza dane dotyczące badanych procesów) i jako takie stanowić przedmiot wnioskowania, a nie jedynie ilustrację tekstu. Będą to głównie wykresy liniowe, np. wykres obrazujący trend rozwojowy danego obiektu w kategoriach analizowanej cechy. Tego rodzaju wykresy mogą być wykreślane dla każdego obiektu i dla każdej cechy. Można także podjąć próbę dopasowania odpowiedniej funkcji opisującej toczący się proces. Bardziej interesujący wydaje się jednak liniowy wykres współczynników zmienności, obrazujący proces upodabniania się (konwergencji) względnie różnicowania się (dywergencji) badanych obiektów ze względu na cechę braną pod uwagę. Jeśli zaś obiektami takiej tablicy są np. miasta, zawarte dane mogą posłużyć do wykreślenia map potencjału czy trendu powierzchniowego, a także do zastosowania modeli grawitacji. Wymaga to jednak uprzedniej konstrukcji tablicy (macierzy) wzajemnych odległości o wymiarach $m \times m$, gdzie m jest liczbą obiektów (Chojnicki, 1966; Chojnicki, Czyż i Ratajczak, 2011; Dawidowski, Ratajczak i Stankowski, 1985).

⁵ Tablice takie, podobnie jak inne opisane w artykule, można zestawiać z danych zawartych w publikacjach lub na portalach GUS, choć w niektórych przypadkach wymaga to relatywizacji danych surowych.

Tablice wielocechowe

Wielocechowa tablica strukturalna przyjmuje postać przedstawioną w tabl. 2.

TABL. 2. MODEL WIELOCECHOWEJ TABLICY STRUKTURALNEJ

Obiekty	Cechy (zmienne)				
	1	2	3	...	<i>n</i>
1	y_{11}	y_{12}	y_{13}	...	y_{1n}
2	y_{21}	y_{22}	y_{23}	...	y_{2n}
3	y_{31}	y_{32}	y_{33}	...	y_{3n}
...
<i>m</i>	y_{m1}	y_{m2}	y_{m3}	...	y_{mn}

Źródło: jak przy tabl. 1.

Elementami wielocechowej tablicy strukturalnej są cechy (zmienne) opisujące zbiór badanych obiektów, np.: gęstość zaludnienia, odsetek ludności miejskiej, liczba podmiotów gospodarczych na 1000 mieszkańców, zbiory zbóż z 1 ha, przeciętne wynagrodzenie, stopa bezrobocia itp. Wiersze tablicy zawierają więc kolejne cechy charakteryzujące poszczególne uwzględniane w badaniu obiekty, natomiast kolumny – wartości kolejnych cech w przekroju badanych obiektów. Są to cechy zmierzone dla jednego przekroju (cięcia) czasowego, tj. dla konkretnej daty, zazwyczaj roku.

Tego rodzaju tablice pozwalają badać i opisywać ukształtowanie struktury przestrzennej, a także mogą być punktem wyjścia do klasyfikacji i regionalizacji (Parysek, 1982).

Często stosowanym w badaniach zabiegiem jest transformacja wielocechowej tablicy danych surowych, czyli zmiennych opisujących cechy badanych obiektów, w tablice danych standaryzowanych. Jest to szczególnie ważne przy zastosowaniu metody analizy wielocechowej – od prostych wskaźników syntetycznych poczynawszy, na wielozmiennej analizie statystycznej skończywszy.

Najbardziej poprawna, z punktu widzenia teorii statystyki, jest standaryzacja dokonana za pomocą formuły:

$$y_{ij}^* = 1/S_j(y_{ij} - y_j') \quad (1)$$

gdzie:

y_{ij}^* – standaryzowana *j*-ta zmienna (cecha) *i*-tego obiektu,

S_j – odchylenie standardowe *j*-tej cechy,

y_{ij} – oryginalna *j*-ta zmienna (cecha) *i*-tego obiektu,

y_j' – średnia wartość *j*-tej cechy.

Wiele dostępnych programów obliczeniowych (zarówno programów do obliczania składowych głównych, jak i programów klasyfikacyjnych) obejmuje automatyczną standaryzację cech.

Za pomocą wielocехowej tablicy strukturalnej można m.in.:

- opisywać stan rzeczy poszczególnych obiektów w kategoriach uwzględnianych cech, a także porównywać te stany, analizując odpowiednie wiersze lub elementy tablicy;
- oceniać wariancję obiektów w kategoriach poszczególnych cech, wyliczając dla kolumn tablicy (cech) statystyczne współczynniki zmienności⁶;
- określić stopień wzajemnego skorelowania zmiennych (cech) w przekroju badanych obiektów poprzez obliczenie współczynników korelacji dla branych pod uwagę par cech (korelacja dla kolumn). Utworzenie macierzy korelacji o wymiarach $n \times n$ (gdzie n oznacza liczbę uwzględnianych cech), której elementami są współczynniki korelacji r_{jk} ($j \neq k = 1, 2, \dots, n$ – uwzględniane cechy), umożliwia wydzielenie podzbiorów powiązanych ze sobą cech (Karoński i Caliński, 1974a). Miarą związku jest wtedy współczynnik korelacji, a strukturę podobieństwa (powiązania) odwzorowuje wykreślony dendryt dualny (Caliński i Harabasz, 1974). Wykreślenie dendrytu wymaga jednak przekształcenia współczynników korelacji w odległości taksonomiczne za pomocą przekształcenia ($d_{ik} = 1 - r_{ik}$);
- określić wielocехową strukturę podobieństwa badanych obiektów przy wykorzystaniu metody składowych głównych, metod analizy skupień lub innych metod klasyfikacji. W obu przypadkach należy jednak dokonać uprzedniej standaryzacji cech (zmiennych) lub mieć pewność, że zastosowany program uwzględnia automatyczną standaryzację, na co już zwracano uwagę⁷.

W przypadku zastosowania klasycznych modeli analizy skupień strukturę podobieństwa odwzorowuje macierz współczynników podobieństwa (d_{ij}), gdzie $i \neq j = 1, 2, 3, \dots, m$ – badane obiekty, o wymiarach $m \times m$, gdzie m – liczba badanych obiektów. Współczynnikami podobieństwa (d_{ij}) są taksonomiczne odległości euklidesowe⁸. Odległości te wylicza się, stosując formułę:

⁶ Współczynniki zmienności, jako standaryzowane miary wariancji, umożliwiają porównanie różnicowania zbioru obiektów w kategoriach poszczególnych cech.

⁷ W bibliografii zamieszczono wybrane pozycje literaturowe, tradycyjne i późniejsze, opisujące metody wielozmiennej analizy statystycznej, w tym wielozmienną analizę wariancji, analizę składowych głównych i analizę kanoniczną, a także metody klasyfikacji, w tym analizę skupień. Oprócz przywołanych w artykule w innych miejscach są to: Everitt, Landau i Leese (2001), Gnanadesikan (2011), Hardle i Simar (2007), Jajuga, Sokołowski i Bock (2002) oraz Parysek (1977).

⁸ Większość metod taksonomicznych w procedurze klasyfikacyjnej uwzględnia odległości euklidesowe mierzone w przestrzeni n -wymiarowej, niemniej jednak są i takie metody, które wykorzystują inne odległości, np. Mahalanobisa, Minkowskiego, „Manhattan” (*taxi cab*) i łukową (*chord distance*) (Anderberg, 1973; Młodak i in., 2016; Parysek, 1978, 1982), lub w analizie skupień wykorzystuje się analizę gęstości. To ostatnie podejście ma zastosowanie zwłaszcza w odniesieniu do big data, szczególnie w astronomii i kosmologii (Murtagh i Heck, 1987; Silverman, 2018).

$$d_{ik} = \left[\sum_{j=1}^p (y_{ij} - y_{kj})^2 \right]^{\frac{1}{2}} \quad (2)$$

gdzie:

d_{ik} – euklidesowa odległość taksonomiczna (funkcja podobieństwa) obiektu i -tego i k -tego,

y_{ij} – wartość j -tej cechy i -tego obiektu,

y_{kj} – wartość j -tej cechy k -tego obiektu ($i \neq k = 1, 2, 3, \dots, m$ – brane pod uwagę obiekty, $j = 1, 2, 3, \dots, n$ – uwzględniane cechy).

Funkcją podobieństwa obiektów może być także współczynnik korelacji wyliczony dla par wartości w wierszach macierzy standaryzowanej.

Wielocechowa tablica danych ma również zastosowanie w przypadku syntetycznej charakterystyki struktury zbioru obiektów w badanym okresie, co umożliwia wykorzystanie analizy składowych głównych oraz dokonanie klasyfikacji wielocechowej na podstawie wartości składowych głównych (Anderberg, 1973; Anderson, 1958; Caliński, 1969; Chatfield, 2018; Gower, 1967; Kaczmarek i Parysek, 1977; Karoński i Caliński, 1974b; Kaufman i Rousseeuw, 2005; Lance i Williams, 1966; Mackiewicz i Ratajczak, 1993; Murtagh i Heck, 1987; Parysek, 1980, 1982; Parysek i Ratajczak, 1978, 2002; Romesburg, 2004; Silverman, 2018). Uzyskane klasyfikacje mogą w dalszej kolejności być podstawą realizacji procedur regionalizacyjnych (Chojnicki i Czyż, 1973; Parysek, 1980, 1982)⁹.

Dokonywanie operacji na zbiorach danych zawartych w tablicach umożliwiają zarówno proponowane w literaturze zaawansowane procedury obliczeniowe, jak i liczne pakiety programowe obliczeń statystycznych. Zupełnie nowe perspektywy otwiera analiza dużych zbiorów danych (big data).

W tego rodzaju tablicach w przekroju jednostek podziału terytorialnego kraju lub innych jednostek przestrzennych (np. przekroje międzynarodowe) zawarte są dane charakteryzujące poziom rozwoju społeczno-gospodarczego, społecznego lub gospodarczego w konkretnym momencie (najczęściej na koniec roku lub przeciętnie w roku). Są to zazwyczaj dane relatywizowane (stosownie do przyjmowanych standardów, np.: na 1 mieszkańca, 1000 lub 10000 mieszkańców, na 1 lub 100 km², na 100 ha użytków rolnych, na 1 pracującego itp.), obrazujące poziom rozwoju danej dziedziny życia społecznego (sytuacja demograficzna, rynek pracy, dochody i wydatki ludności, struktura spożycia, poziom wykształcenia itp.), gospodarczego (ogólny poziom rozwoju gospodarczego, PKB według np. sekcji gospodarki, struktura dochodów i wydatków budżetowych, efekty ekonomiczne prowadzonej działalności gospodarczej według sekcji, działów czy grup itp.) lub poziom rozwoju społeczno-gospodarczego jednostek tery-

⁹ Przytoczona literatura umożliwia lepsze poznanie jedynie zasygnalizowanych lub skrótowo opisanych procedur obliczeniowych.

torialnych uwzględnianego w badaniu ich zbioru (cechy rozwoju społecznego i gospodarczego)¹⁰.

Na podstawie tablic strukturalnych można prowadzić przede wszystkim różnego rodzaju badania struktur, w tym struktur przestrzennych, a także wykorzystać zawarte w nich dane do budowy modeli regresji wielozmiennej.

TABLICE FUNKCJONALNE

Dysponując odpowiednimi danymi, można tworzyć także tablice funkcjonalne. Funkcjonalność oznacza w tym przypadku pełnienie określonych funkcji¹¹. W badaniach społeczno-ekonomicznych częstym problemem jest klasyfikacja obiektów ze względu na pełnione funkcje (np. klasyfikacja miast, funkcjonalna klasyfikacja jednostek terytorialnych). Dane statystyczne potrzebne do przeprowadzenia takiej klasyfikacji mają specyficzny charakter. Pełnienie lub niepełnienie konkretnej funkcji zaznacza się przy wykorzystaniu zmiennych binarnych, zmierzonych w skali nominalnej, tj. zero-jedynkowej, gdzie 1 oznacza pełnienie danej funkcji (obecność placówek pełniących te funkcje), a 0 – jej niepełnienie (nieobecność).

Ogólny model tablicy funkcjonalnej jest zatem identyczny z modelem tablicy strukturalnej, ale zawiera inne elementy (zmiennne ciągłe są zastąpione zmiennymi binarnymi).

Tablica funkcjonalna (zawsze wielo cechowa) umożliwia m.in.:

- wskazanie, które z badanych obiektów pełnią brane pod uwagę funkcje (zmienna 1), a które nie (zmienna 0);
- wskazanie liczby funkcji charakterystycznych dla danego obiektu (pełnionych przez ten obiekt), o czym świadczy suma wartości równych 1 w wierszach tablicy;
- stwierdzenie, ile i które elementy badanego zbioru (obiekty) pełnią konkretne funkcje, co świadczy o stopniu ich powszechności;
- określenie wielo cechowej struktury funkcjonalnej zbioru badanych obiektów, do czego prowadzi wyliczenie współczynników podobieństwa i zestawienie ich w macierz (macierz $m \times m$);
- dokonanie, na podstawie macierzy współczynników podobieństwa, wielo cechowej klasyfikacji funkcjonalnej¹².

¹⁰ Zestawiane dane (cechy), w celu zapewnienia wiarygodności wykonywanych obliczeń, powinny spełniać warunki omówione w części *Badane obiekty i ich cechy*.

¹¹ Pełnienie określonych funkcji w pewnym sensie może oznaczać także funkcjonowanie.

¹² Klasyfikacje funkcjonalne, a właściwie rangowanie obiektów ze względu na liczbę pełnionych funkcji, można także przeprowadzać na podstawie sum elementów w wierszach (wartości równych 1), ale taka klasyfikacja nie ma charakteru jakościowego (także strukturalnego), lecz jedynie ilościowy. Klasyfikację jakościową, a zarazem strukturalną, uzyskuje się, obliczając wspomniane współczynniki podobieństwa, zestawiając je w macierz struktury podobieństwa, a następnie wydzielając odpowiednie klasy.

Procedura określania struktury podobieństwa i klasyfikacji funkcjonalnej (przy uwzględnieniu zmiennych binarnych) jest jednak nieco inna od klasycznej klasyfikacji wielocechowej, w której funkcją podobieństwa jest geometryczna odległość euklidesowa d_{ij} (ewentualnie współczynnik korelacji r_{ij}). Inny charakter ma funkcja podobieństwa, którą jest współczynnik podobieństwa S_{ik} .

Podstawę wyliczania współczynników podobieństwa stanowi dwudzielcza tablica kontyngencyjna w postaci przedstawionej w tabl. 3.

TABL. 3. MODEL DWUDZIELCZEJ TABLICY KONTYNGENCYJNEJ

Obiekt i -ty	Obiekt k -ty		
	1	0	suma
1	a	b	$a+b$
0	c	d	$c+d$
Suma	$a+c$	$b+d$	$a+b+c+d=N$

U w a g a. a – liczba funkcji pełniących jednocześnie przez obiekt i -ty i k -ty, b – pełnienie danej funkcji przez obiekt i -ty, a niepełnienie przez k -ty, c – pełnienie danej funkcji przez obiekt k -ty, a niepełnienie przez i -ty, d – liczba jednocześnie niepełnionych funkcji przez obiekty i -ty i k -ty.

Ź r ó d ł o: jak przy tabl. 1.

Literatura statystyczna (także geograficzna) zawiera liczne sformułowania funkcji podobieństwa, będących różnymi kombinacjami elementów a , b , c i d (N) dwudzielczej tablicy kontyngencyjnej (Parysek, 1982). Do najbardziej popularnych należą współczynniki Russella i Rao ($S_{ik} = a/N$), Jaccarda ($S_{ik} = a/a+b+c$), Dice'a ($S_{ik} = 2a/2a+b+c$), Rogersa-Tanimoto [$S_{ik} = a+d/a+d+2(b+c)$], Kulczyńskiego ($S_{ik} = a/b+c$) i Yule'a [$S_{ik} = (\frac{1}{2}ad - \frac{1}{2}bc)/(\frac{1}{2}ad + \frac{1}{2}bc)$]. Steinhausen i Langer (1977) opracowali także ogólny model współczynników podobieństwa obiektów opisywanych zmiennymi binarnymi skali nominalnej w postaci:

$$S_{ik} = a + \sigma d/a + \sigma d + \lambda(b + c) \quad (3)$$

gdzie:

- S_{ik} – współczynnik podobieństwa obiektu i -tego i k -tego,
- a, b, c, d – elementy uzyskane z dwudzielczej tablicy kontyngencyjnej,
- σ, λ – parametry przekształceń, gdzie $\sigma = 1$ lub 0 , a $\lambda = \frac{1}{2}, 1$ lub 2 (Parysek, 1982; Steinhausen i Langer, 1977).

Obliczone współczynniki podobieństwa zestawia się w macierz odwzorowującą strukturę podobieństwa obiektów ze względu na pełnione funkcje, na podstawie której dokonuje się wydzielenia klas obiektów podobnych, np. konstruuje dendryt lub dendrogram. W przypadku dendrytu lub dendrogramu problemem wymagającym rozwiązania jest przekształcenie współczynnika podobieństwa

S_{ik} w odległość taksonomiczną d_{ik} . W tym celu można wykorzystać formułę przekształcenia zaproponowaną przez Sokala i Michenera (1958) w postaci:

$$d_{ik} = [2(1 - S_{ik})]^{1/2} \quad (4)$$

Należy podkreślić, że proponowane współczynniki podobieństwa poddawane są ocenie i modyfikacji (Albatineh, Niewiadomska-Bugaj i Mihalko, 2006; Anderberg, 1973; Parysek, 1982). Za współczynniki podobieństwa obiektów charakteryzowanych zmiennymi binarnymi uznaje się także m.in.: współczynnik korelacji, statystykę chi-kwadrat, kosinus kąta pomiędzy wektorami, współczynniki koligacji i asocjacji oraz wiele innych (Anderberg, 1973; Parysek, 1982).

Przykładem tego rodzaju tablic może być tablica, w której zbiór tej samej rangi jednostek osadniczych (np. wszystkich miast danej jednostki terytorialnej, miast o liczbie ludności powyżej 100 tys. mieszkańców, ośrodków gminnych, wsi w danym powiecie itp.) charakteryzują dane binarne, wskazujące na pełnienie określonych funkcji. Liczba 1 oznacza wtedy pełnienie przez daną jednostkę osadniczą konkretnej funkcji, natomiast 0 – niepełnienie tej funkcji. Dopiero w nawiązaniu do tak skonstruowanej tablicy, przy zastosowaniu dwudzielczej tablicy kontyngencyjnej, oblicza się podane wyżej współczynniki podobieństwa badanych jednostek osadniczych. Pełnienie funkcji dotyczy przede wszystkim istnienia w danej miejscowości jednostki społeczno-gospodarczej, a także działania elementu infrastruktury technicznej lub społecznej, np.: szkoły ponadpodstawowej, liceum, szkoły wyższej, sklepu wielobranżowego, galerii handlowej, placówki bankowej, przychodni lekarskiej, szpitala, stacji kolejowej, kościoła, biblioteki publicznej, teatru, muzeum, hali sportowej, krytej pływalni, stacji benzynowej, stacji obsługi samochodów itp. W szczególnych przypadkach tablica w zapisie zero-jedynkowym może także uwzględniać odbywające się w danych ośrodkach (lub nie) imprezy, np.: wystawy, festiwale, imprezy folklorystyczne, koncerty plenerowe, gościnne występy artystów, dożynki, biegi, zawody sportowe, pokazy, zloty itp.

Przedstawione podejście znajduje zastosowanie przede wszystkim w analizie funkcjonalnej układów osadniczych, w tym w badaniach poziomu centralności poszczególnych jednostek osadniczych. Może być użyteczne w kształtowaniu racjonalnych podziałów terytorialnych tworzonych ze względu na funkcje jednostek osadniczych tego samego poziomu centralności (ośrodków pełniących podobne funkcje) oraz ich wielkości i położenia geograficznego.

TABLICE RELACYJNE (INTERAKCYJNE)

Tablice relacyjne, zwane też interakcyjnymi, to zestawienia danych liczbowych o wymiarach $m \times m$, gdzie m jest liczbą uwzględnianych w badaniu obiektów. Ogólny model tego rodzaju tablicy przedstawiono w tabl. 4.

TABL. 4. MODEL TABLICY RELACYJNEJ

Obiekty	Obiekty				
	1	2	3	...	m
1	f_{11}	f_{12}	f_{13}	...	f_{1n}
2	f_{21}	f_{22}	f_{23}	...	f_{2n}
3	f_{31}	f_{32}	f_{33}	...	f_{3n}
...
m	f_{m1}	f_{m2}	f_{m3}	...	f_{mn}

U w a g a. f_{ik} – relacja między obiektem i -tym a k -tym ($i \neq k = 1, 2, 3, \dots, m$).

Ź r ó d ł o: jak przy tabl. 1.

Występujące na głównej przekątnej tablicy elementy f_{ii} przyjmują wartość zerową, bowiem oznaczają brak wzajemnego oddziaływania (obiekt i -ty nie oddziałuje na siebie samego). Należy podkreślić, że tablica (macierz interakcji) jest niesymetryczna, choć główna przekątna dzieli ją na dwie części. Niesymetryczność takich macierzy oznacza, że $f_{ik} \neq f_{ki}$. Dolny trójkąt macierzy odwzorowuje relacje czynne (takie jak wyjazdy, nadania towarów, oddziaływanie czynne itp.) jednostek i -tych względem k -tych, natomiast górny – relacje bierne jednostek k -tych w stosunku do i -tych. Konsekwentnie wiersze macierzy wskazują na relacje czynne (aktywne) obiektów i -tych względem k -tych, natomiast kolumny na relacje bierne (pasywne) elementów k -tych w stosunku do i -tych. Zakładając, że macierz relacyjna dotyczy dojazdów do pracy, w wierszach takiej macierzy znajdują się wyjazdy z miejscowości (obszarów) i -tych do k -tych, a w kolumnach – przyjazdy do miejscowości (obszarów) k -tych z i -tych.

W przeciwieństwie do tablic funkcjonalnych, których elementami są cechy skalarnie zapisane za pomocą zmiennych binarnych, elementy tablicy relacyjnej to cechy wektorowe wskazujące na interakcje, tj. wzajemne oddziaływanie na siebie względnie powiązanie dwóch obiektów (f_{ik}). Do tego rodzaju cech zalicza się np. dane dotyczące: dojazdów do pracy, przepływu osób związanych z korzystaniem z różnego rodzaju usług, migracji, przepływu towarów, pieniędzy itp.

Za pomocą tablic relacyjnych można m.in.:

- określać łączną sumę oddziaływania czynnego (przepływu: osób, towarów, pieniędzy itp.) jednostki i -tej w stosunku do pozostałych k -tych, sumując elementy f_{ik} w poszczególnych wierszach tablicy;
- określać łączną sumę relacji biernych (przyjazdów osób, dostaw towarów, dopływu pieniędzy itp.) danej jednostki k -tej ze wszystkimi jednostkami i -tymi, sumując elementy f_{ik} w poszczególnych kolumnach tablicy;
- określać charakter danej jednostki w zakresie powiązań z inną jednostką (jednostka aktywna czy pasywna), odejmując od siebie elementy, f_{ik} i f_{ki} , czyli relacje czynne i bierne;

- określać siłę wzajemnych powiązań, sumując odpowiednie elementy tablicy wskazujące na wzajemne oddziaływanie dwóch jednostek, tj. elementów $f_{ik} + f_{ki}$;
- określać struktury (np. przestrzenne) powiązań, dzieląc poszczególne elementy f_{ik} przez sumy tych elementów dla wierszy lub kolumn (w wymiarze procentowym mnożąc przez 100);
- wydzielać układy silnych powiązań, stosując odpowiednio zmodyfikowane procedury wykorzystywane w konstrukcji dendrytów lub dendrogramów.

Tablice relacyjne mogą zawierać zarówno zmienne ciągłe, jak i binarne. W każdym przypadku (jednak przede wszystkim w tym drugim) mogą być podstawą zastosowania analizy grafowej odnoszonej do wyznaczania układów powiązań (podsystemów sieci osadniczej).

Przykładami tablic relacyjnych są różnego rodzaju zestawienia dotyczące głównie przepływu osób i towarów, a w mniejszym stopniu (z uwagi na trudną dostępność danych) – pieniędzy, technologii, informacji itp. Dotyczą przede wszystkim migracji, zarówno wahałowych (dojazdy do pracy, szkół, urzędów, placówek handlowych i usługowych, wyjazdy wakacyjne, dla załatwiania innych spraw), jak i stałych (napływ ludności, odpływ ludności). Tego rodzaju tablice mogą uwzględniać zarówno relacje obustronne jednostek, których dotyczy badanie, jak i relacje jednostronne¹³.

PODSUMOWANIE

Problematyka przedstawiona w niniejszym opracowaniu jest podejmowana od lat, ponieważ tablice od dawna służą jako narzędzie porządkowania zebranych danych liczbowych. Można stosować także inne rozwiązania, pozwalające badać dynamikę i określać struktury zbioru elementów, a także opisywać charakter wzajemnych relacji między elementami zbioru badanych obiektów, biorąc za podstawę konkretne tablice danych. Możliwości w tym zakresie wyznaczają, z jednej strony, matematyka i statystyka oraz ich narzędzia, a z drugiej – inwencja osoby prowadzącej badanie. Natomiast barierą jest niemożność pozyskania potrzebnych danych liczbowych.

Stosując uporządkowane zestawy danych oraz wykorzystując możliwości analityczne, jakie tworzą odpowiednio zestawione tablice danych, można w sposób pełniejszy i bardziej różnorodny (niekiedy odmienny od ujęć standardowych) badać struktury, zdarzenia i procesy, a jednocześnie racjonalnie, czyli efektywnie, wykorzystać zebrane, niekiedy z wielkim trudem, zestawy danych liczbowych.

¹³ Jest rzeczą oczywistą, że relacje obustronne nie mogą wystąpić pomiędzy wszystkimi branżami pod uwagę jednostkami przestrzennymi (także jednostkami osadniczymi). Przykładowo z jednostki A do B może dojeżdżać do pracy pewna liczba osób i odwrotnie (relacja obustronna), jednak z jednostki A do B mogą dojeżdżać jakieś osoby do szkół, a z B do A nie (relacja jednostronna).

BIBLIOGRAFIA

- Albatineh, A. N., Niewiadomska-Bugaj, M., Mihalko, D. (2006). Correcting Jaccard and other similarity indices for chance agreement in cluster analysis. *Advances in Data Analysis and Classification*, 5(3), 179–200.
- Anderberg, M. R. (1973). *Cluster analysis for application*. New York: Academic Press.
- Anderson, T. W. (1958). *An introduction to multivariate analysis*. New York: Wiley.
- Caliński, T. (1969). On the application of cluster analysis to experimental results. *Bulletin of the International Statistical Institute*, (42), 101–103.
- Caliński, T., Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics – Theory and Methods*, 41(12), 2279–2280.
- Chatfield, B. (2018). *Introduction to multivariate analysis*. New York: Routledge.
- Chojnicki, Z. (1966). *Zastosowanie modeli grawitacji i potencjału w badaniach przestrzenno-ekonomicznych*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Chojnicki, Z., Czyż, T. (1973). *Metody taksonomii numerycznej w regionalizacji geograficznej*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Chojnicki, Z., Czyż, T., Ratajczak, W. (2011). *Model potencjału. Podstawy teoretyczne i zastosowania w badaniach przestrzenno-ekonomicznych i regionalnych*. Poznań: Bogucki Wydawnictwo Naukowe.
- Dawidowski, J., Ratajczak, W., Stankowski, W. (1985). Trend powierzchniowy w zastosowaniu do określenia rzeźby i powierzchni złóż. *Przegląd Geologiczny*, 33(8), 425–431.
- Everitt, B. S., Landau, S., Leese, M. (2001). *Cluster analysis*. London: Arnold.
- Gnanadesikan, R. (2011). *Methods for statistical data analysis of multivariate observations*. New York: Wiley & Sons Inc.
- Gower, J. C. (1967). A comparison of some methods of cluster analysis. *Biometrics*, 23(4), 623–637.
- Hardle, W. K., Simar, L. (2007). *Applied multivariate statistical analysis*. New York: Springer.
- Jajuga, K., Sokołowski, A., Bock, H. H. (2002). *Classification, clustering and data analysis*. New York: Springer.
- Kaczmarek, Z., Parysek, J. (1977). Zastosowanie analizy wielowymiarowej w badaniach geograficzno-ekonomicznych. W: Z. Chojnicki (red.), *Metody ilościowe i modele w geografii* (s. 94–127). Warszawa: Państwowe Wydawnictwo Naukowe.
- Karoński, M., Caliński, T. (1974a). Grupowanie cech na podstawie współczynnika korelacji. *Roczniki Akademii Rolniczej w Poznaniu*, (64), 95–99.
- Karoński, M., Caliński, T. (1974b). Grupowanie obiektów wielocechowych na podstawie odległości euklidesowych. *Roczniki Akademii Rolniczej w Poznaniu*, (64), 117–122.
- Kaufman, L., Rosseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Wiley-Interscience: New York, John Wiley & Sons Inc.
- Lance, G. N., Williams, W. T. (1966). Computer programs for hierarchical polythetic classification. *The Computer Journal*, 10, 373–380.
- Maćkiewicz, A., Ratajczak, W. (1993). Principal components analysis. *Computers and Geosciences*, 19(3), 303–342.
- Młodak, A., Józefowski, T., Wawrowski, Ł. (2016). Zastosowanie metod taksonomicznych w estymacji wskaźników ubóstwa. *Wiadomości Statystyczne*, (2), 1–24.
- Morrison, D. F. (1967). *Multivariate Statistical Methods*. New York: McGraw-Hill.
- Murtagh, F., Heck, A. (1987). *Multivariate data analysis*. Dordrecht: Springer Netherlands.

- Orloci, L. (1967). An agglomerative method for classification of plant communities. *Journal Ecological*, 55(1), 193–205.
- Parysek, J. (1977). The application of principal component analysis and canonical analysis to the study of socio-economic spatial structure and its changes. *Quaestiones Geographicae*, (4), 131–148.
- Parysek, J. (1978). Zastosowanie taksonomicznej odległości Mahalanobisa w dynamicznych badaniach strukturalno-przestrzennych. *Przegląd Geograficzny*, 50(2), 293–308.
- Parysek, J. (1980). Analiza skupień jako metoda klasyfikacji w geografii. W: Z. Chojnicki (red.), *Metody taksonomiczne w geografii* (s. 87–99). Warszawa–Poznań: Państwowe Wydawnictwo Naukowe.
- Parysek, J. (1982). *Modele klasyfikacji w geografii*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. A. Mickiewicza w Poznaniu.
- Parysek, J., Ratajczak, W. (1978). Analiza głównych składowych cech charakteryzujących właściwości społeczno-ekonomiczne i środowisko geograficzne Polski w 1970 roku. W: Z. Chojnicki, T. Czyż, J. Parysek, W. Ratajczak (red.), *Badania społeczno-ekonomicznej struktury przestrzennej Polski metodami czynnikowymi* (s. 83–106). Warszawa–Poznań: Państwowe Wydawnictwo Naukowe.
- Parysek, J., Ratajczak, W. (2002). Analiza składowych głównych – jej korzyści i ograniczenia z punktu widzenia badań geograficznych. W: H. Rogacki (red.), *Możliwości i ograniczenia zastosowań metod badawczych w geografii społeczno-ekonomicznej i gospodarce przestrzennej* (s. 61–73). Poznań: Bogucki Wydawnictwo Naukowe.
- Romesburg, H. C. (2004). *Cluster analysis for researchers*. North Carolina: Lulu Press. Wodsworth Inc.
- Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge: New York.
- Sokal, R. R., Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *The University of Kansas Science Bulletin*, 38(22), 1409–1438.
- Steinausen, D., Langer, K. (1977). *Clusteranalyse. Einführung in methoden und Verfahren der automatischen Klassifikation*. Berlin–New York: Walter de Gruyter.