ANALIZA SYSTEMÓW PRZESTRZENNYCH. WYBRANE ZAGADNIENIA Jan W. Owsiński (Red.)

POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

ANALIZA SYSTEMÓW PRZESTRZENNYCH

WYBRANE ZAGADNIENIA

Redakcja

Jan W. Owsiński

Warszawa 2010

ANALIZA SYSTEMÓW PRZESTRZENNYCH

WYBRANE ZAGADNIENIA

Polska Akademia Nauk • Instytut Badań Systemowych
Seria: BADANIA SYSTEMOWE
tom 67

Redaktor naukowy:
Prof. dr hab. inż. Jakub Gutenbaum

Warszawa 2010

INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK

ANALIZA SYSTEMÓW PRZESTRZENNYCH

WYBRANE ZAGADNIENIA

Redakcja
Jan W. Owsiński

Warszawa 2010

Autorzy:
Jan W. Owsiński, redaktor
Instytut Badań Systemowych PAN
Pracownia Zastosowań Metod Badań Systemowych
Tel. (48 22) 3810 213
e-mail: Jan.Owsinski@ibspan.waw.pl

Jan Gadomski
Jerzy W. Hołubiec
Barbara Mażbic-Kulma
Michał Milczewski
Jan W. Owsiński
Grażyna Petriczek
Aneta M. Pielak
Henryk Potrzebowski
Krzysztof Sęp
Eugeniusz Sobczak
Jarosław Stańczak

Recenzenci:
Prof. dr hab. inż. Jacek Mercik
Prof. dr hab. Tadeusz Trzaskalik

ISBN 9788389475251
ISSN 0208-8029

Redakcja i opracowanie techniczne: Jan W. Owsiński i Aneta M. Pielak

IV. Badanie jednorodności zbioru danych w analizie przestrzennej

Jerzy Hołubiec, Grażyna Petriczek

IV.1. Wprowadzenie

Algorytm podziału zbioru na jednorodne, rozłączne grupy zastosowano do analizy struktur na poziomie powiatów. W niniejszej pracy algorytm ten został zastosowany do analizy następujących konkretnych zadań, związanych ze zbiorami powiatów: a) podział powiatów Województwa Mazowieckiego na grupy jednorodne (w sensie ilościowym) ze względu na wybrane cechy opisujące rozwój społeczno-ekonomiczny tych powiatów, b) porównanie podziałów powiatów par województw: Mazowieckiego i Śląskiego, Mazowieckiego i Łódzkiego, Mazowieckiego i Świętokrzyskiego, oraz Mazowieckiego i Pomorskiego. Dodajmy, że wymienione województwa należały do różnych klas jednorodności przy podziale zbioru województw.

Do testowania hipotezy o jednorodności zbioru danych wykorzystuje się statystykę U o rozkładzie χ^2. Metoda polega na iteracyjnym podziale niejednorodnego zbioru na dwie jednorodne części. Jeżeli liczba tych podziałów (iteracji) wzrasta, to mogą pojawić się statystycznie niestabilne (nieistotne) granice między sąsiednimi podzbiorami. Agregacja takich podzbiorów oparta jest na badaniu odpowiednio skonstruowanej hipotezy dotyczącej stabilności granic między dwoma podzbiorami. Tak więc, algorytm podziału zbioru danych składa się z dwóch etapów:

- Podział zbioru na jednorodne, rozłączne podzbiory - pierwotny podział zbioru
- Badanie stabilności granic między tymi podzbiorami.

IV.2. Algorytm podziału zbioru na rozłączne jednorodne podzbiory

Omawiany w pracy model jednorodności zbioru danych oparty jest na zasadzie równoważności zmiennych losowych o jednakowych rozkładach. Przedstawiamy w zarysie jego istotę.

Niech S={s_1, s_2, ,s_n} będzie analizowanym zbiorem danych. Załóżmy, że każdemu elementowi $s \in S$ opowiada zmienna losowa ξ_S o dystrybuancie $F_S(x)$. Oznaczmy zbiór zmiennych losowych ξ_S przez E^S, natomiast zbiór wartości x zmiennych losowych przez R ($x \in R$).

Definicja 1. *Zmienne losowe* ξ_{s_1}, ξ_{s_2} *nazywamy zmiennymi losowymi równoważnymi jeżeli dla dowolnych dwóch elementów* $s_1, s_2 \in S$ *zachodzi:*

$$F_{s_1}(x) - F_{s_2}(x) = 0 \quad \text{dla dowolnego } x \in R. \qquad \text{(IV.1)}$$

Oznacza to, że zbiór zmiennych losowych E^S można rozbić na klasy równoważności, tzn. na grupy zmiennych równoważnych.

W ten sposób zbiór S może być przedstawiony w postaci sumy mnogościowej rozłącznych podzbiorów

$$S = S_1 \cup S_2 \cup .... \cup S_k, \quad k \geq 1. \qquad \text{(IV.2)}$$

Jeżeli k=1, to zbiór S jest zbiorem jednorodnym. Jeżeli k>1 , to zbiór S jest niejednorodny i zależność (IV.2) odzwierciedla tą niejednorodność. W oparciu o pojęcie równoważności zmiennych losowych możemy podać następującą definicję jednorodności:

Definicja 2. *Zbiór zmiennych losowych* $E^{S_1} \subset E^S$ *jest zbiorem jednorodnym, jeżeli spełniony jest warunek:*

$$F_{s'}(x) - F_{s''}(x) = 0 \quad \text{dla każdego } s', s'' \in S_1 \text{ oraz } x \in R. \qquad \text{(IV.3)}$$

Tak więc, jeżeli dla zbioru E^{S_1} spełniony jest warunek (IV.3) i zbiór ten pokrywa się z całą przestrzenią, to cała przestrzeń (zbiór) E^S jest zbiorem jednorodnym.

Poprzez zaprzeczenie warunkowi jednorodności, (IV.3), otrzymujemy definicję niejednorodności, czyli:

Definicja 3. *Jeżeli w zbiorze* $E^{S_1} \subset E^S$ *istnieje para* $s', s'' \in S_1$, *dla której*

$$F_{s'}(x) - F_{s''}(x) \neq 0 \quad \text{dla jakiegokolwiek } x \in R \qquad \text{(IV.4)}$$

to zbiór jest zbiorem niejednorodnym.

W powyższych definicjach nie nakłada się żadnych warunków na postać rozkładu badanej zmiennej losowej.

W celu skonstruowania kryteriów dla testowania hipotez o jednorodności, przyjmuje się dodatkowo, że zmienne losowe ξ_S są niezależne i mają rozkłady normalne z funkcją gęstości w postaci:

$$f(x) = \frac{1}{\sqrt{(2\pi)^k}} \left|\Sigma_s\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(x - m_s\right)^T \Sigma_s^{-1}\left(x - m_s\right)\right) \quad \text{(IV.5)}$$

gdzie:

m_S - wektor wierszowy, którego elementami są wartości oczekiwane zmiennej losowej ξ_S

Σ_S - macierz kowariancji o wymiarach [k x k]

$\left|\Sigma_S\right|$ - wyznacznik macierzy kowariancji

$$\Sigma_S = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots\cdots & \gamma_{1k} \\ \gamma_{21} & \gamma_{22} & \cdots\cdots & \gamma_{2k} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ \gamma_{k1} & \gamma_{k2} & \cdots\cdots & \gamma_{kk} \end{bmatrix}$$

$\gamma_{ij} = C\left(\xi_{si}, \xi_{sj}\right)$ - kowariancja zmiennych losowych

$\gamma_{ii} = V\left(\xi_{si}\right)$ - wariancja zmiennej losowej.

Jeśli założymy, że rozpatrywane zmienne losowe mają jednakowe macierze kowariancji, to wówczas warunek jednorodności (IV.3) jest równoważny równości wartości oczekiwanych i ma postać (hipoteza H_0):

$$\begin{aligned} H_0 : \quad & m_{S'} = m_{S''} \quad \text{dla wszystkich } s', s'' \in S \\ & \text{pod warunkiem:} \\ & \Sigma_{S'} = \Sigma_{S''} \end{aligned} \quad \text{(IV.6)}$$

Analogicznie, warunek niejednorodności przybiera postać: (hipoteza H1):

$$\begin{aligned} H_1 : \quad & m_{S'} \neq m_{S''} \quad \text{dla conajmniej jednej pary } s', s'' \in S \\ & \text{pod warunkiem :} \\ & \Sigma_{S'} = \Sigma_{S''} \end{aligned} \quad \text{(IV.6.1)}$$

Łatwo zauważyć, że jeżeli E^S jest zbiorem jednorodnym, to dla dowolnej pary zbiorów E^{S_1}, E^{S_2} zachodzi:

$$\frac{1}{n_1}\sum_{s\in S_1} m_s - \frac{1}{n_2}\sum_{s\in S_2} m_s = 0 \qquad \text{dla wszystkich } S_1, S_2 \in S$$

gdzie: n_1 – liczba elementów zbioru S_1
n_2 – liczba elementów zbioru S_2.

Definiując na zbiorze wszystkich rozbić zbioru S na dwa podzbiory, S_1, S_2 funkcję:

$$\delta(S_1, S_2) = \frac{1}{n_1}\sum_{s\in S_1} m_s - \frac{1}{n_2}\sum_{s\in S_2} m_s \tag{IV.7}$$

otrzymujemy wskaźnik jednorodności k-wymiarowego zbioru zmiennych.

Stosując wskaźnik (IV.7), hipotezę zerową o jednorodności można sformułować następująco:

H_0: $$\delta(S_1, S_2) = 0 \tag{IV.8}$$
dla dowolnej pary (S_1,S_2) należącej do zbioru wszystkich rozbić zbioru S na dwa podzbiory.

Założenia potrzebne do wprowadzenia kryterium (IV.8) w praktyce mogą być przyjęte bez większych przeszkód.

Niech n będzie liczbą obserwacji, zaś k – liczbą rozpatrywanych cech charakteryzujących analizowane zjawisko. Wtedy rezultat jednej obserwacji (o numerze s) zmiennych losowych ξ_{s_j} można zapisać w postaci:

$X_s = \{x_{s1}, x_{s2}, \quad , x_{sj}, \quad , x_{sk}\}$, gdzie: s – numer obserwacji, s=1,.........,n.

Zbiór wszystkich obserwacji k-wymiarowej zmiennej losowej jest macierzą o wymiarach [n x k], o postaci:

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} x_{11}, & x_{12}, & \cdots, & x_{1k} \\ x_{21}, & x_{22}, & \cdots, & x_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_{n1}, & x_{n2}, & \cdots & x_{nk} \end{bmatrix}. \tag{IV.9}$$

Dane przedstawione w macierzy (IV.9) rozpatruje się jak realizację k-wymiarowych zmiennych losowych ξ_s o rozkładzie normalnym, ze średnimi m_s i jednakowymi diagonalnymi macierzami kowariancji.

Kryterium badania hipotezy zerowej H_0 sprawdza się przez porównanie dwóch próbek. Pociąga to za sobą rozbicie macierzy (IV.9) na dwie różne, rozłączne części zawierające, odpowiednio, n_1 i n_2 wierszy.

Statystycznej oceny k-wymiarowego rozbicia dostarcza zmienna losowa $\tilde{\xi}$ o postaci:

$$\tilde{\xi} = \frac{1}{n_1} \sum_{s \in S_1} \xi_s - \frac{1}{n_2} \sum_{s \in S_2} \xi_s \tag{IV.10.1}$$

gdzie:

$$\tilde{\xi} = \left[\tilde{\xi}_1, \tilde{\xi}_2, \dots\dots\dots, \tilde{\xi}_k\right]$$

$$\tilde{\xi}_j = \frac{1}{n_1} \sum_{s \in S_1} \xi_{sj} - \frac{1}{n_2} \sum_{s \in S_2} \xi_{sj}\,. \tag{IV.10.2}$$

Każda ze składowych, występujących w zależności (IV.10.1) jest zmienną losową z odpowiednimi parametrami rozkładu: wartościami oczekiwanymi $(m_{j'}, m_j)$ oraz wariancjami $(\tilde{\sigma}_j^2 / n_1, \tilde{\sigma}_j^2 / n_2)$.

Konstrukcję funkcji kryterialnej do weryfikacji hipotezy zerowej przeprowadza się wykorzystując metodę największej wiarygodności.

Jeżeli zakłada się słuszność hipotezy H_0, to zgodnie z warunkiem (IV.6) musi zachodzić równość $m_{j'} - m_j = 0$ i funkcja wiarygodności przybiera postać:

$$L(x,m)=\frac{1}{\sqrt{(2\pi)^k}}\left(\prod_{j=1}^{k}c_j^2\right)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\sum_{j=1}^{k}\frac{\widetilde{x}_j^2}{c_j^2}\right) \qquad \text{(IV.11)}$$

gdzie: c_j^2 - wariancja zmiennej losowej $\widetilde{\xi}$, przedstawiona za pomocą zależności:

$$c_j^2=\frac{\widetilde{\sigma}_j^2(n_1+n_2)}{n_1 n_2}, \qquad \text{(IV.12a)}$$

$\widetilde{\sigma}_j^2$ - wartość z próby wariancji zmiennej losowej $\widetilde{\xi}_j$,

$\widetilde{x}_j$ - wartość z próby j-tej składowej zmiennej losowej $\widetilde{\xi}$ określona następująco:

$$\widetilde{x}_j=\frac{1}{n_1}\sum_{s\in S_1}x_{sj}-\frac{1}{n_2}\sum_{s\in S_2}x_{sj}. \qquad \text{(IV.12b)}$$

Wartość z próby wariancji $\widetilde{\sigma}_j^2$ wyznacza się z zależności:

$$\widetilde{\sigma}_j^2=\frac{1}{n_1+n_2-1}\left(\sum_{s\in S_1}x_{sj}^2+\sum_{s\in S_2}x_{sj}^2-\frac{1}{n_1+n_2}\left(\sum_{s\in S_1}x_{sj}+\sum_{s\in S_2}x_{sj}\right)^2\right) \text{(IV.12c)}$$

Z postaci funkcji (IV.11) wynika, że jej przebieg zależy od wykładnika potęgowego $\sum_{j=1}^{k}\frac{\widetilde{x}_j^2}{c_j^2}$ i zachodzi:

$$L^*(x;m)\geq L_{pr}(\eta) \quad dla \sum_{j=1}^{k}\frac{\widetilde{x}_j^2}{c_j^2}\leq\eta$$

gdzie: η - pewna zadana wartość

Podstawiając (IV.12a)-(IV.12c) do wykładnika otrzymujemy funkcję kryterialną o postaci:

$$U(S_1,S_2)=\frac{\frac{n_1+n_2-1}{(n_1+n_2)n_1n_2}\sum_{j=1}^{k}\left(n_2\sum_{s\in S_1}x_{sj}-n_1\sum_{s\in S_2}x_{sj}\right)^2}{\sum_{j=1}^{k}\left(\sum_{s\in S}x_{sj}^2-\frac{1}{n_1+n_2}\left(\sum_{s\in S}x_{sj}\right)^2\right)}, \quad \text{(IV.13)}$$

gdzie: $S = S_1 \cup S_2$.

Z postaci (IV.13) wynika, że przy spełnieniu hipotezy zerowej statystyka $U(S_1,S_2)$ ma rozkład χ^2 o k stopniach swobody. Zgodnie z zasadą największej wiarygodności oraz właściwościami funkcji L(,) hipoteza zerowa o jednorodności dwóch próbek może być przyjęta jeżeli zachodzi:

$$U(S_1,S_2)\le \chi^2_{\alpha,k} \quad \text{(W1)}$$

dla dowolnej pary (S_1,S_2) należącej do zbioru wszystkich rozbić zbioru S

gdzie: α - oznacza przyjęty poziom istotności
k – oznacza liczbę stopni swobody i równe jest liczbie rozpatrywanych cech.

Jeżeli warunek (W1) nie jest spełniony, to hipotezę H_0 należy odrzucić: zbiór S nie jest zbiorem jednorodnym i może być rozbity na dwa rozłączne podzbiory, S_1 i S_2.

Wyznaczanie statystyk o postaci (IV.13) dla wszystkich par (S_1, S_2) oraz sprawdzanie nierówności (W1) jest skomplikowane, zwłaszcza gdy n jest duże.

Dlatego też w proponowanej w metodzie przyjmuje się dwa założenia ułatwiające badanie jednorodności zbioru:

1) obserwacje $\{X_1, X_2, ..., X_n\}$ są uszeregowane względem najbardziej istotnej cechy w porządku rosnącym tzn. od wartości najmniejszej do największej.

2) nie można zmieniać zadanego tym porządkiem rozkładu elementów obserwacji względem k cech.

Powyższe założenia nie powodują ani utraty pierwotnej informacji, ani też nie mają wpływu na samą ideę metody.

Przy przyjętych powyżej założeniach weryfikację hipotezy H_0 wystarczy przeprowadzić tylko dla takich par (S_1, S_2), w których do zbioru S_1 należy p pierwszych elementów zbioru $\{X_1, X_2, \ldots, X_n\}$, natomiast do zbioru S_2: n – p pozostałych elementów, gdzie p=1,2,.......,n-1.

Wówczas statystyka (IV.13) przyjmuje następującą postać:

$$U(p, n-p) = \frac{\frac{n-1}{n(n-p)p} \sum_{j=1}^{k} \left((n-p) \sum_{i=1}^{p} x_{ij} - 1 \sum_{i=p+1}^{n} x_{ij} \right)^2}{\sum_{j=1}^{k} \left(\sum_{i=1}^{n} x_{ij}^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_{ij} \right)^2 \right)} \qquad \text{(IV.14)}$$

dla p=1,2,........,n-1,

gdzie: n – liczba obserwacji
k – liczba obserwowanych cech.

Zgodnie z wcześniejszymi rozważaniami, hipotezę H_0 o jednorodności przyjmujemy, jeżeli spełniona jest nierówność:

$$U(p, n-p) \leq \chi^2_{\alpha,k} \qquad \text{dla p=1,2,..........,n-1.} \qquad \text{(W2)}$$

Jeżeli chociaż dla jednego pl (jednego wiersza) nierówność (W2) nie jest spełniona, to hipotezę H_0 odrzucamy: zbiór nie jest jednorodny.

W takim przypadku przyjmujemy hipotezę alternatywną:

$$H_1: \quad \begin{array}{l} \delta(S_1, S_2) \neq 0 \\ U(p, n-p) > \chi^2_{\alpha,k} \end{array} \qquad \text{(W3)}$$

i zbiór S należy podzielić na podzbiory jednorodne.

Podział (rozbicie) niejednorodnego zbioru danych na jednorodne, rozłączne podzbiory jest oparty na przyjęciu hipotezy alternatywnej postaci (W3).

Zgodnie z metodą największej wiarygodności przyjęcie hipotezy H_1 (o niejednorodności) wymaga osiągnięcia przez funkcję L(,) maksymalnej wartości, co z kolei jest równoważne minimalizacji wykładnika potęgowego występującego w postaci tej funkcji. Po odpowiednim przekształceniu problem minimalizacji wykładnika sprowadza się do problemu maksymalizacji statystyki o postaci:

$$\max_{p} U(p, n-p). \tag{W4}$$

Z warunku (W4) wynika, że funkcja wiarygodności osiąga maksimum przy takim rozbiciu niejednorodnego zbioru danych na dwie części, przy którym statystyka U(p,n-p) ma maksymalną wartość.

Warunki (kryteria) (W3) i (W4) stanowią teoretyczną podstawę metody podziału niejednorodnego zbioru na dwa rozłączne, jednorodne podzbiory.

Ogólnie, zatem, algorytm podziału na grupy jednorodne można przedstawić następująco:

1) dla uszeregowanego względem najbardziej charakterystycznych cech zbioru danych $X=\{X_i\}$, i=1,....,n, obliczamy statystyki:

U(1,n-1), U(2,n-2), ..., U(p,n-p), ..., U(n-1,1).

2) wybieramy największą wartość statystyki U; niech to będzie, powiedzmy, $U(l_1, n-l_1)$ – odpowiadającą wierszowi o numerze l_1 w macierzy danych.

3) dzielimy zbiór S na dwa podzbiory, S_1, S_2, zawierające, odpowiednio, l_1 pierwszych elementów i $n-l_1$ pozostałych elementów.

4) dla każdego z otrzymanych podzbiorów testujemy następnie hipotezę H_0 o jego jednorodności; jeżeli którykolwiek z nich nie jest zbiorem jednorodnym, to dzielimy go na dwie części, zgodnie z największą wartością statystyki U.

5) proces ten powtarzamy dopóty, dopóki wszystkie otrzymane w wyniku kolejnych podziałów podzbiory nie będą spełniały hipotezy jednorodności H_0.

W ten sposób w skończonej liczbie iteracji otrzymuje się podział zbioru S na rozłączne, jednorodne podzbiory. Należy zauważyć, że liczba iteracji nie zależy od liczby rozpatrywanych cech.

Otrzymane w wyniku kolejnych podziałów podzbiory (grupy) spełniają podstawowe warunki ilościowego grupowania – tzn. zasadę równoważności elementów w grupie i rozłączności grup.

Zasada równoważności elementów wynika z łączenia w jedną grupę obserwacji na podstawie kryterium (W2). Natomiast rozłączność jednorodnych grup elementów wynika ze sformułowania zadania podziału zbioru według kryterium maksymalnej rozłączności między grupami – kryterium (W4) maksymalnej wartości statystyki U.

IV.3. Agregacja grup: hipoteza o niestabilności granic

Opisana metoda podziału zbioru polegała na iteracyjnym rozbijaniu niejednorodnego zbioru na dwie części. Jeżeli liczba tych rozbić (liczba iteracji) wzrasta, to proces kolejnych podziałów może doprowadzić do pojawienia się statystycznie niestabilnych (nieistotnych) granic między sąsiednimi, jednorodnymi podzbiorami. Znalezienie takich międzygrupowych granic i ich usunięcie z otrzymanego wcześniej podziału prowadzi w wyniku do otrzymania istotnego podziału zbioru populacji na jednorodne podzbiory.

Ogólnie biorąc, statystyczną stabilność granic między podzbiorami (grupami) populacji można badać porównując średnie wielowymiarowe. Jeżeli wielowymiarowe średnie dwóch porównywalnych grup są statystycznie równoważne, to można założyć, że istnieje statystycznie niestabilna granica i grupy, które ona rozdziela, można połączyć w jedną grupę, bez naruszenia jednorodności.

Jeżeli jednak, w wyniku porównania, otrzymuje się istotną różnicę między wielowymiarowymi średnimi, to granica miedzy dwoma jednorodnymi podzbiorami istnieje i łączenie podzbiorów nie ma sensu.

Poniżej podamy metodę badania stabilności granic między grupami opartą na weryfikacji odpowiednio sformułowanej hipotezy.

Załóżmy, że w toku pierwotnego grupowania populacja (zbiór obiektów) została rozbita na M grup jednorodnych. Niech m_i oznacza wielowymiarową średnią i-tego podzbioru (grupy). Przy przyjętych założeniach, hipoteza zerowa o tym, że granica między zbiorami E^{S_i} i $E^{S_{i+1}}$ jest statystycznie niestabilna, zapisana jest w postaci relacji:

$$H_0:\ m_i - m_{i+1} = \{0,0,.........,0\}. \qquad \text{(W5)}$$

Zaś hipoteza alternatywna ma postać:

$$H_1:\ m_i - m_{i+1} \neq \{0,0,.........,0\}. \qquad \text{(W6)}$$

Przyjęcie hipotezy H_0 oznacza, że granica między dwoma zbiorami jest statystycznie niestabilna i zbiory te można połączyć.

Odrzucenie hipotezy zerowej (W5) powoduje przyjęcie jej alternatywy (W6) i uznanie istotności granic między grupami tzn. istnienia stabilnych granic.

Badanie granic przeprowadza się kolejno dla wszystkich podzbiorów i bądź to łączy się sąsiednie podzbiory w jedną grupę (hipoteza H_0), bądź tez uznaje się istnienie istotnych granic między tymi podzbiorami (H_1).

Statystyczne kryterium do weryfikacji hipotez (W5) – (W6) konstruuje się analogicznie jak dla podziału zbioru na jednorodne podzbiory. A więc zakłada się, że elementy rozpatrywanych dwóch podzbiorów stanowią realizacje k-wymiarowych zmiennych losowych. O zmiennych zakłada się, że mają rozkład normalny i są niezależne.

Definiując analogicznie jak w podrozdziale IV.2 funkcję:

$$\delta(S_i, S_{i+1}) = \frac{1}{n_i} \sum_{s \in S_i} m_s - \frac{1}{n_{i+1}} \sum_{s \in S_{i+1}} m_s \qquad \text{(IV.16)}$$

gdzie:

S_i, S_{i+1} - dwa sąsiednie zbiory

n_i, n_{i+1} - liczebność tych zbiorów

hipotezę zerową H_0 postaci (W5) o istnieniu statystycznie niestabilnej granicy między dwoma zbiorami można sformułować następująco:

$$H_0: \qquad \delta(S_i, S_{i+1}) = 0 \qquad \text{(IV.17)}$$

Funkcję kryterialną do badania hipotezy (H_0) konstruuje się, wykorzystując metodę największej wiarygodności. W tym celu określa się k – wymiarową zmienną losową $\tilde{\xi}$ o postaci:

$$\tilde{\xi} = \frac{1}{n_i} \sum_{s \in S_i} \xi_s - \frac{1}{n_{i+1}} \sum_{s \in S_{i+1}} \xi_s \qquad \text{(IV.18.1)}$$

gdzie: $\tilde{\xi} = \left[\tilde{\xi}_1, \tilde{\xi}_2, \ldots\ldots\ldots, \tilde{\xi}_k\right]$

$$\tilde{\xi}_j = \frac{1}{n_i} \sum_{s \in S_i} \xi_{sj} - \frac{1}{n_{i+1}} \sum_{s \in S_{i+1}} \xi_{sj}\,. \qquad \text{(IV.18.2)}$$

Każda ze składowych występujących w zależności (IV.18.2) jest zmienną losową z odpowiednimi parametrami rozkładu: wartościami oczekiwanymi (m_i, m_{i+1}) oraz wariancjami $(\tilde{\sigma}_j^2 / n_i, \tilde{\sigma}_j^2 / n_{i+1})$. Wartość z próby j-tej składowej zmiennej losowej $\tilde{\xi}_j$ ma postać:

$$\tilde{x}_j = \frac{1}{n_i} \sum_{s \in S_i} x_{sj} - \frac{1}{n_{i+1}} \sum_{s \in S_{i+1}} x_{sj} \qquad \text{(IV.18.3)}$$

gdzie:

x_{sj} – wartości zmiennej losowej, będące elementami tablicy obserwacji.

Jeżeli zakłada się słuszność hipotezy zerowej (W5) ($m_i - m_{i+1} = 0$), to odpowiednia funkcja wiarygodności przybiera postać:

$$L(x:m) = \frac{1}{\sqrt{(2\pi)^k}} \left(\prod_{j=1}^{k} c_j^2 \right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \sum_{j=1}^{k} \frac{\tilde{x}_j^2}{c_j^2} \right) \qquad \text{(IV.19.1)}$$

gdzie: c_j^2 - wariancja zmiennej losowej $\tilde{\xi}$, przedstawiona za pomocą zależności:

$$c_j^2 = \frac{\tilde{\sigma}_j^2 (n_i + n_{i+1})}{n_i n_{i+1}}, \qquad \text{(IV.19.2)}$$

Z kolei wartość z próby wariancji $\tilde{\sigma}_j^2$ wyznacza się z zależności:

$$\tilde{\sigma}_j^2 = \frac{1}{n_i + n_{i+1} - 1}\left(\sum_{s \in S_i} x_{sj}^2 + \sum_{s \in S_{i+1}} x_{sj}^2 - \frac{1}{n_i + n_{i+1}}\left(\sum_{s \in S_i} x_{sj} + \sum_{s \in S_{i+1}} x_{sj}\right)^2\right) \quad \text{(IV.19.3)}$$

Jak wiadomo, przebieg funkcji wiarygodności zależy od wykładnika potęgowego. Wykorzystując zależności (IV.18.3), (IV.19.2), (IV.19.3) otrzymujemy statystykę o postaci (dla każdej kolejnej pary podzbiorów):

$$U(S_i, S_{i+1}) = \frac{\dfrac{n_i + n_{i+1} - 1}{(n_i + n_{i+1}) n_i n_{i+1}} \displaystyle\sum_{j=1}^{k}\left(n_{i+1} \sum_{s \in S_i} x_{sj} - n_i \sum_{s \in S_{i+1}} x_{sj}\right)^2}{\displaystyle\sum_{j=1}^{k}\left(\sum_{s \in S} x_{sj}^2 - \frac{1}{n_i + n_{i+1}}\left(\sum_{s \in S} x_{sj}\right)^2\right)} \quad \text{(IV.20)}$$

gdzie: $S = S_i \cup S_{i+1}$ i=1,..........,M

M – liczba grup

k – liczba rozpatrywanych cech

n_i, n_{i+1} – liczba elementów, odpowiednio, zbiorów S_i, S_{i+1}

x_{sj} - wartości zmiennej losowej ξ_{sj}, będące elementami tablicy obserwacji.

Można udowodnić, że przy wyżej przedstawionych założeniach badanie hipotezy H_0 sprowadza się do badania następującej nierówności:

$$U(S_i, S_{i+1}) \le \chi^2_{\alpha,k} \quad i=1,2, ...,K. \quad \text{(IV.21)}$$

Jeśli nierówność (IV.21) jest spełniona, to można przyjąć, że granica między zbiorami S_i oraz S_{i+1} jest niestabilna. Łączymy wówczas te podzbiory i testujemy hipotezę o stabilności granic między podzbiorami ($S = S_i \cup S_{i+1}$) oraz S_{i+2}, itd.

Jeżeli natomiast $U(S_i, S_{i+1}) > \chi^2_{\alpha,k}$, to granicę między zbiorami S_i oraz S_{i+1} utrzymuje się i przechodzimy do testowania hipotezy o stabilności granic między podzbiorami S_{i+1} oraz S_{i+2}. Badanie przeprowadza się kolejno między wszystkimi sąsiednimi podzbiorami, na jakie został podzielone pierwotny zbiór S.

IV.4. Zastosowanie algorytmu do badania struktur przestrzennych

Algorytm podziału zbioru na jednorodne, rozłączne grupy zastosowano do analizy struktur powiatowych. W pracy przedstawiono następujące analizy: a) podział powiatów województwa mazowieckiego na grupy jednorodne (w sensie ilościowym) ze względu na wybrane cechy opisujące rozwój społeczno-ekonomiczny tych powiatów, b) analiza struktur przestrzennych jednostek administracyjnych na poziomie powiatów oraz weryfikacja hipotezy o (ograniczonej) analogii struktur na poziomie wojewódzkim i powiatowym.

IV.4.1 Analiza jednorodności powiatów województwa mazowieckiego

Omówiony algorytm zastosowano do analizy struktur powiatowych województwa mazowieckiego. Zadanie polegało na podziale 42 powiatów województwa (w tym 5 powiatów grodzkich: Ostrołęka, Płock, Radom, Siedlce i Warszawa) na grupy jednorodne ze względu na wybrane cechy opisujące rozwój społeczno-ekonomiczny tych powiatów.

Rozważono następujące zestawy danych (dla roku 2003):

a) **Kryterium a).** Dane dotyczące podstawowych (ogólnych) informacji o powiatach, zawierające 10 cech: •ludność, • powierzchnia, • zatrudnienie, • bezrobocie (liczba bezrobotnych), •zasoby mieszkaniowe, • dochody ogółem budżetów powiatów, • dochody własne budżetów powiatów, • wydatki budżetów powiatów, • nakłady inwestycyjne w przedsiębiorstwach, • baza noclegowa (miejsca noclegowe). Dochody własne budżetów powiatów składają się między innymi z podatków od nieruchomości, wpływów z podatku dochodowego od osób fizycznych i prawnych, dochodów z najmu i dzierżawy.

b) **Kryterium b).** Dane dotyczące poziomu życia w powiatach, obejmujące 12 cech: • ludność, struktura zatrudnienia - • zatrudnienie w rolnictwie, • zatrudnienie w przemyśle, • zatrudnienie w budownictwie, • zatrudnienie w usługach, • bezrobocie, • zasoby mieszkaniowe, • przeciętne wynagrodzenie miesięczne brutto, • zużycie wody z wodociągów, • zużycie gazu, • zużycie energii elektrycznej, • liczba łóżek w szpitalach.

c) **Kryterium c)**. Dane dotyczące rolniczego i leśnego użytkowania gruntów, obejmujące 4 cechy: • powierzchnia ogólna, • grunty orne, • „użytki zielone", • lasy i grunty leśne. „Użytki zielone" obejmują sady, łąki i pastwiska.

Wartość rozkładu $\chi^2_{\alpha,k}$ dla 4 cech (stopni swobody) i poziomu istotności α=0.05 wynosi 9.488, dla 10 cech (przy tym samym poziomie istotności) wynosi 18.307, a dla 12 cech wynosi 21.026.

Dla przypadków a) i b) jako cechę wiodącą względem, której uszeregowano dane (rosnąco) przyjęto liczbę ludności, natomiast dla przypadku c): wielkość powierzchni ogólnej.

Dla kryterium a) po 13 iteracjach otrzymano 14 grup powiatów. Analiza stabilności granic wykazała, że w dwóch przypadkach granice między sąsiednimi podzbiorami są niestabilne i można je połączyć. W efekcie otrzymano podział na 12 jednorodnych grup powiatów. Dla kryterium b) tj. dla 12 cech, po 9 iteracjach otrzymano 10 jednorodnych grup powiatów o stabilnych granicach między podzbiorami.

Wyniki analizy przedstawiają, odpowiednio, Rys. IV.1 i IV.2. Na obu tych rysunkach cyframi arabskimi oznaczono odpowiednie grupy jednorodne.

Analiza wyników otrzymanych dla wymienionych 10 i 12 cech prowadzi, m.in., do następujących wniosków:

- Warszawa, jako powiat, stanowi samodzielną grupę, zarówno ze względu na cechy charakteryzujące ogólny rozwój powiatu, jak i na cechy opisujące poziom życia w powiecie.
- Miasta o statusie powiatu niekoniecznie są w tej samej grupie, co powiat ziemski, w którym są położone.
- Przy rozpatrywaniu cech charakteryzujących ogólny rozwój powiatu otrzymano większą liczbę jednorodnych grup powiatów aniżeli w przypadku cech opisujących poziom życia w powiecie
- Dla obu przypadków, a) i b), takie powiaty jak:

  miasto Warszawa
  wołomiński, miasto Radom
  piaseczyński, miasto Płock, miński, pruszkowski, radomski

siedlecki, sochaczewski, ostrołęcki, płoński
ciechanowski, legionowski, warszawski, płocki

należą do tych samych grup powiatów; co może oznaczać, że powiaty w tych grupach mają podobny zarówno ogólny poziom rozwoju jak i ogólnie rozumiany poziom życia

Rys. IV.1. Podział powiatów województwa mazowieckiego (kryterium a): cechy opisujące ogólny rozwój.
Źródło: Opracowanie własne.

Rys. IV.2. Podział powiatów województwa mazowieckiego (przykład b): cechy opisujące poziom życia
Źródło: Opracowanie własne.

Dla przypadku c) po 12 iteracjach otrzymano 13 grup powiatów. Analiza stabilności granic wykazała, że w trzech przypadkach granice między sąsiednimi podzbiorami są niestabilne i można je połączyć. W wyniku tego otrzymano podział na 10 jednorodnych, stabilnych grup (Rys. IV.3 na następnej stronie).

W przypadku charakterystyk dotyczących rolniczego i leśnego użytkowania gruntów Warszawa jako powiat należy do wspólnej grupy z powiatami: legionowskim, szydłowieckim i piaseczyńskim, co może stanowić pewnego rodzaju niespodziankę

Przedstawione wyżej podziały powiatów pozwalają na wyodrębnienie jednorodnych grup powiatów zarówno ze względu na ich rozwój i poziom życia, jak i strukturę użytkowania gruntów.

Podsumowanie otrzymanych wyników dla poszczególnych przypadków zamieszczono w Tabeli IV.1 na końcu tego rozdziału. Jest to o tyle istotne, że dalsza analiza odnosi różne województwa właśnie do mazowieckiego.

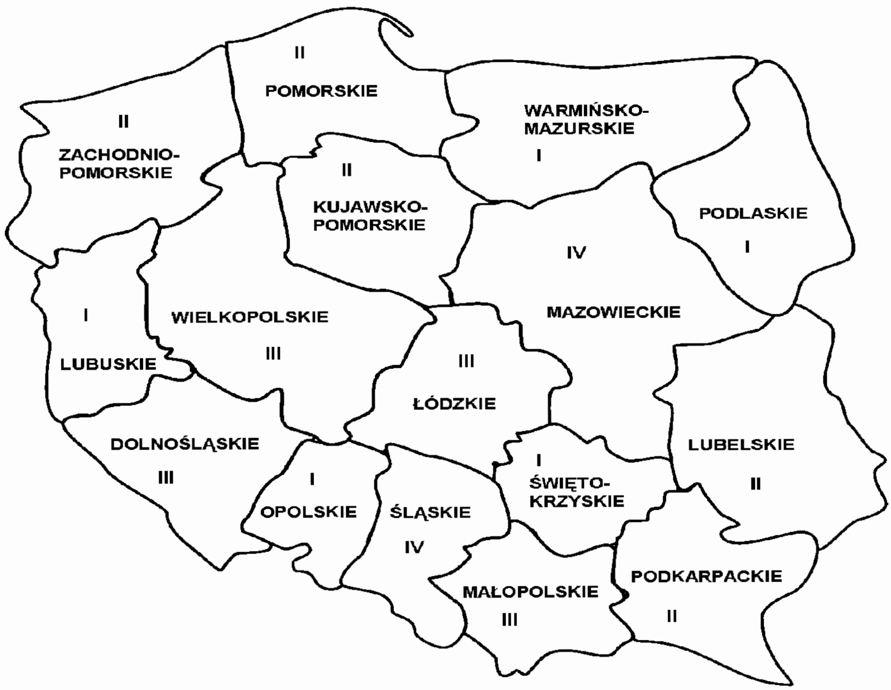
IV.4.2 Analiza porównawcza podziałów powiatów wybranych województw

Przeanalizowano powiaty należące do różnych klas jednorodności województw. Dla wszystkich analizowanych struktur (zarówno województw jak i powiatów) rozpatrzono cechy charakteryzujące ogólny rozwój powiatu, a więc: • ludność, • powierzchnia ogólna, • zatrudnienie, • bezrobocie (liczba bezrobotnych), •zasoby mieszkaniowe, • dochody ogółem budżetów powiatów, • wydatki budżetów powiatów, • nakłady inwestycyjne w przedsiębiorstwach, • baza noclegowa (miejsca noclegowe).

Wybór wymienionych cech związany był z dostępnością jednolitych zestawów danych dla wszystkich powiatów w rozpatrywanych województwach.

Analiza podziału zbioru wszystkich 16 województw na grupy jednorodne ze względu na rozpatrywane cechy pozwoliła na wyodrębnienie 4 jednorodnych grup województw, tak, jak to pokazano na Rys. IV.4. Na rysunku tym liczbami rzymskimi oznaczono odpowiednie grupy jednorodne województw Polski.

Rys.IV.3. Podział powiatów województwa mazowieckiego (przykład c): cechy opisujące rolnicze i leśne użytkowanie gruntów
Źródło: Opracowanie własne.



Rys. IV.4. Podział województw na jednorodne grupy
Źródło: opracowanie własne

W następnym etapie analizy porównano powiaty województwa mazowieckiego z powiatami województw: a) śląskiego, b) łódzkiego, c) świętokrzyskiego, d) pomorskiego.

Jak wynika z analizy podziału województw (Rys. IV.4), województwa mazowieckie i śląskie należą do tej samej grupy jednorodności, natomiast województwa łódzkie, świętokrzyskie oraz pomorskie znajdują się w oddzielnych grupach jednorodności.

Dla wszystkich rozpatrywanych poniżej przypadków podziału powiatów z dwóch porównywanych województw zdefiniowano **zbiory wspólne jako zbiory, do których należą powiaty z obu rozpatrywanych województw.**

Dalej, we wszystkich rozpatrywanych przypadkach porównań (grup) powiatów w parach województw przyjęto następujące oznaczenia:

- cyframi rzymskimi wyróżniono powiaty tworzące zbiory wspólne - są to zbiory, do których należą powiaty z obu rozpatrywanych województw
- cyframi arabskimi w nawiasach oznaczono grupy powiatów jednorodnych (dla danego województwa) nie należące do wspólnych zbiorów - tworzą one osobne zbiory dla każdego z rozpatrywanych województw

Dla przedstawionych zmiennych, użytych w tej analizie, 42 powiaty województwa mazowieckiego zostały podzielone na 11 jednorodnych grup.

Województwo mazowieckie i śląskie:

Województwo śląskie składa się z 36 powiatów, w tym aż 20 miast ma statut powiatu. W wyniku zastosowania algorytmu otrzymano 8 jednorodnych grup powiatów.

Łączny zbiór powiatów województw mazowieckiego i śląskiego składa się z 78 powiatów. Oba województwa należą do tej samej grupy jednorodności ze względu na rozpatrywane cechy. W wyniku podziału zbioru powiatów obu województw wyodrębniono 15 jednorodnych zbiorów. Podział łączny powiatów mazowieckich i śląskich pokazano na Rys. IV.5 i IV.6.

Otrzymane wyniki można podsumować następująco:

- ♦ Wśród 15 jednorodnych grup powiatów istnieje 10 zbiorów wspólnych zawierających powiaty z obu województw. Do zbiorów tych należą 33 powiaty województwa mazowieckiego (79% ogólnej liczby powiatów tego województwa) oraz 29 powiatów województwa śląskiego (81% ogólnej liczby powiatów województwa śląskiego).
- ♦ Województwo mazowieckie ma 2 odrębne jednorodne zbiory, do których należą odpowiednio: do pierwszego: 3 powiaty, zaś do drugiego: 5 powiatów.
- ♦ Województwo śląskie ma jeden osobny zbiór zawierający 6 powiatów.
- ♦ Katowice i Warszawa stanowią dwa osobne zbiory.

Rys. IV.5. Podział powiatów województw mazowieckiego i śląskiego. Województwo mazowieckie (ta sama klasa jednorodności województw)
Źródło: opracowanie własne

1. Gliwice, IX
2. Bytom, IX
3. Chorzów, VII
4. Katowice
5. Piekary Śląskie, II
6. Ruda Śląska, (3)
7. Rybnik, VIII
8. Siemianowice Śl., III
9. Sosnowiec, X
10. Świętochłowice, I
11. Zabrze, IX
12. Żory, II
13. Bielsko-Biała, IX
14. Jastrzębie Zdrój, VI
15. Mysłowice, V
16. Tychy, VIII

Rys. IV.6. Podział powiatów województw mazowieckiego i śląskiego. Województwo śląskie (ta sama klasa jednorodności województw)
Źródło: opracowanie własne

Województwo mazowieckie i łódzkie:

Województwo łódzkie zawiera 24 powiaty, w tym 3 miasta na prawach powiatu. W efekcie podziału wyodrębniono 6 jednorodnych grup powiatów.

Łączny zbiór powiatów województw mazowieckiego i łódzkiego składa się z 66 powiatów. Oba rozpatrywane województwa należą do różnych grup jednorodności ze względu na rozpatrywane cechy.

W efekcie zastosowania algorytmu wyodrębniono 14 jednorodnych grup powiatów. Grupy te przedstawiono na Rys. IV.7 i IV.8.

Rys. IV.7. Podział powiatów województw mazowieckiego i łódzkiego. Województwo łódzkie (różne klasy jednorodności województw)
Źródło: opracowanie własne

Podsumowując otrzymane wyniki można stwierdzić, że:

- Wśród 14 jednorodnych grup powiatów można wyodrębnić tylko 5 grup wspólnych zawierających powiaty z obu województw. Należy do nich 22 powiaty województwa mazowieckiego (52% powiatów mazowieckich) i 16 powiatów województwa łódzkiego (67% powiatów łódzkich).
- Pozostałe powiaty obu województw stanowią odrębne grupy jednorodności: pozostałe powiaty województwa mazowieckiego tworzą cztery odrębne grup jednorodności zawierające, odpowiednio: (1) 3 powiaty, (2) 8 powiatów, (3) 6 powiatów, (4) 2 powiaty, zaś pozostałe powiaty województwa łódzkiego należą do trzech odrębnych grup jednorodności zawierających: (5) 2 powiaty, (6) 2 powiaty, i (7) 3 powiaty.
- Łódź i Warszawa stanowią dwa osobne zbiory.

Rys. IV.8. Podział powiatów województw mazowieckiego i łódzkiego. Województwo mazowieckie (różne klasy jednorodności województw)
Źródło: opracowanie własne

Województwo mazowieckie i świętokrzyskie:

W tym przypadku rozważono województwo świętokrzyskie, które nie należy do żadnej wspólnej grupy jednorodności z omawianymi wcześniej województwami. Województwo świętokrzyskie składa się z 14 powiatów, w tym tylko jedno miasto na prawach powiatu. Analiza jednorodności pozwala podzielić te powiaty na 4 grupy jednorodne.

Łączny zbiór powiatów województw mazowieckiego i świętokrzyskiego składa się z 56 powiatów. W wyniku podziału można wyodrębnić 14 jednorodnych grup powiatów o składzie pokazanym na Rys. IV. 9 i IV.10.

Rys. IV.9. Podział powiatów województw mazowieckiego i świętokrzyskiego. Województwo świętokrzyskie (różne klasy jednorodności)
Źródło: opracowanie własne

Otrzymane wyniki można podsumować w następujący sposób:

- Spośród 14 jednorodnych grup powiatów można wyodrębnić 5 zbiorów wspólnych zawierających powiaty z obu województw. Do tych wspólnych zbiorów należy 38% powiatów województwa mazowieckiego (16 powiatów) oraz 57% powiatów województwa świętokrzyskiego (8 powiatów).
- Warszawa stanowi odrębny zbiór.
- Kielce oraz Radom należą do tej samej grupy jednorodności.
- Pozostała część powiatów z obu województw stanowi odrębne grupy jednorodności, i tak pozostałe powiaty województwa mazowieckiego należą do pięciu odrębnych grup jednorodności zawierających, odpowied-

nio: (1) 3 powiaty, (2) 6 powiatów, (3) 3 powiaty, (4) 7 powiatów, (5) 6 powiatów, natomiast pozostała część powiatów województwa świętokrzyskiego należy do trzech odrębnych grup jednorodności zawierających odpowiednio: (6) 2 powiaty, (7) 2 powiaty, (8) 2 powiaty.

Rys. IV.10. Podział powiatów województw mazowieckiego i świętokrzyskiego. Województwo mazowieckie (różne klasy jednorodności)

Źródło: opracowanie własne

Województwo mazowieckie i pomorskie:

Województwo pomorskie, podobnie jak świętokrzyskie, nie należy do żadnej wspólnej grupy jednorodności z omawianymi wcześniej województwami. Województwo to składa się z 20 powiatów, w tym 4 grodzkie. Analiza jednorodności pozwala podzielić te powiaty na 5 grup jednorodności.

Łączny zbiór powiatów województw mazowieckiego i pomorskiego składa się z 62 powiatów. W efekcie zastosowania algorytmu wyodrębniono 12 jednorodnych grup powiatów, przedstawionych na Rys. IV.11 i IV.12.

Rys. IV.11. Podział powiatów województw mazowieckiego i pomorskiego. Województwo pomorskie (różne klasy jednorodności województw)
Źródło: opracowanie własne

Podsumowanie otrzymanych wyników ma następującą postać:

- Wśród 12 jednorodnych grup powiatów można wyodrębnić cztery grupy wspólne, zawierające powiaty z obu województw. Należy do nich 19 powiatów województwa mazowieckiego (45% powiatów mazowieckich) i 13 powiatów województwa pomorskiego (65% powiatów pomorskich).
- Warszawa stanowi odrębny zbiór.

- Gdynia, Gdańsk stanowią osobny zbiór.
- Pozostałe powiaty obu województw stanowią odrębne grupy jednorodności, i tak: pozostałe powiaty mazowieckie należą do czterech odrębnych grup, zawierających, odpowiednio: (1) 6 powiatów, (2) 6 powiatów, (3) 5 powiatów, (4) 5 powiatów, natomiast pozostałe powiaty województwa pomorskiego należą do trzech odrębnych grup jednorodności zawierających, odpowiednio: (5) 1 powiat, (6) 2 powiaty, (7) 4 powiaty.

Rys. IV.12. Podział powiatów województw Mazowieckiego i Pomorskiego. Województwo Mazowieckie (różne klasy jednorodności województw)
Źródło: opracowanie własne

Z analizy wyników otrzymanych dla wszystkich rozpatrywanych w pracy województw można wnioskować, że

- powiaty należące do województw z tej samej grupy jednorodności mają najwięcej zbiorów wspólnych, co oznacza że procentowy udział powiatów (z tych województw) we wspólnych jednorodnych zbiorach jest duży i mają podobny rozwój społeczno-ekonomiczny.
- powiaty należące do województw z różnych grup jednorodności mają małą liczbę zbiorów wspólnych, co oznacza że procentowy udział powiatów (z tych województw) we wspólnych jednorodnych zbiorach jest mniejszy; powiaty z tych województw różnią się rozwojem społeczno-ekonomicznym.

Podsumowanie otrzymanych wyników przedstawiono w Tabeli IV.2.

IV.5. Uwagi końcowe

Otrzymane wyniki są zaledwie ilustracją możliwości prowadzenia badań przy pomocy zaproponowanego algorytmu w ramach szerszej metodyki, obejmującej (1) wstępną analizę danych, ze szczególnym uwzględnieniem wskazania zmiennych wiodących w strukturze oraz ustalania merytorycznie uzasadnionych zestawów zmiennych, (2) podział na grupy jednorodne według proponowanego algorytmu, (3) badanie wrażliwości wyników ze względu na przyjęte założenia, w tym wybór zmiennej wiodącej, (4) alternatywne podziały otrzymywane przy pomocy innych algorytmów, w tym zwłaszcza zaawansowanej analizy skupień, (5) ocenę jakościową otrzymanych podziałów i grup z punktu widzenia celów rozwoju społeczno-gospodarczego oraz „modeli rozwoju”.

Celem przeprowadzonej przykładowej analizy było zbadanie zachowania się grup obiektów i sprawdzenie czy zachodzi analogia (w sensie przynależności do jednorodnych klas) przy dekompozycji wybranych struktur na mniejsze jednostki.

Wybór cech opisujących rozważane struktury wynikał z dostępności jednolitych zestawów danych dla tych struktur. Oczywiście w przyszłości należałoby zastanowić się nad wyborem odpowiednich cech opisujących różne aspekty rozwoju regionalnego. Ponadto słuszne wydaje się także uwzględnienie jako cechy opisującej powiaty – kategorii porównywalnych (odpowiednie cechy przeliczane na mieszkańca) oraz porównanie wyników otrzymanych dla wartości bezwzględnych cech oraz kategorii względnych.

Z wcześniej przeprowadzonych badań wynika, że zmiana liczby cech (dodanie lub odjęcie) nie wpływa znacząco na liczbę zbiorów jednorodnych; zmianie ulega tylko „zawartość" tych zbiorów.

Przedstawiony algorytm może stanowić część komputerowej procedury służącej wymienionym wyżej zadaniom.

IV.6. Literatura

Byfuglien J., Nordgard A. (1973) Region-building – a comparison of methods. *Norsk Geografisk Tijdsskrift*.

Hołubiec J., Petriczek G. (2004) Homogeneity algorithm for analysis of regional structure. W: *Proceedings of the 15th International Conference on Systems Science*. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, 479-486.

Hołubiec J., Petriczek G. (1997) Homogeneity Algorithm For Modeling Regional Structure. W: *Proceedings of the International Conference on Methods and Models in Automation and Robotics MMAR*, S. Domek, Z. Emirsajłow, R. Kaszyński, red. Technical University of Szczecin, Szczecin, Vol.1, 397-402.

Kildyshev G.S, Abolentzev J.A. (1978) *Mnogomernyie gruppirovki*. Izd. Statistika, Moskva.

Mardia K.V, Kent J.T, Bibby J.M (1979) *Multivariate Analysis*. Academic Press, London.

Owsiński J. (1980) Regionalization revisited: an explicit optimization approach. IIASA, CP-80-26.

Roczniki Statystyczne Województw (2003) Mazowieckiego, Śląskiego, Łódzkiego, Świętokrzyskiego, Główny Urząd Statystyczny, Warszawa.

Taksonomia – Klasyfikacja i analiza danych – teoria i zastosowania (2004, 2005) Prace Naukowe Akademii Ekonomicznej im. Oskara Lanego we Wrocławiu, Wydawnictwo Akademii Ekonomicznej im. Oskara Lanego we Wrocławiu, Wrocław.

Lp	powiat	Kryterium a: 12 zbiorów jednorodnych	Kryterium b: 10 zbiorów jednorod-nych	Kryterium c: 10 zbiorów jedno-rodnych
1	Białobrzeski	**1**	**1**	**4**
2	Ciechanowski	**8**	**7**	**7**
3	Garwoliński	**9**	**7**	**8**
4	Gostyniński	**3**	**3**	**4**
5	Grodziski	**5**	**5**	**2**
6	Grójecki	**9**	**7**	**9**
7	Kozienicki	**4**	**3**	**6**
8	Legionowski	**8**	**7**	**3**
9	Lipski	**2**	**1**	**5**
10	Łosicki	**1**	**1**	**5**
11	Makowski	**3**	**2**	**7**
12	Miński	**10**	**8**	**7**
13	Mławski	**4**	**4**	**8**
14	Nowodworski	**5**	**5**	**4**
15	Ostrołęcki	**7**	**6**	**10**
16	Ostrołęka m.	**3**	**3**	**1**
17	Ostrowski	**5**	**5**	**8**
18	Otwocki	**9**	**8**	**4**
19	Piaseczyński	**10**	**8**	**3**
20	Płock m.	**10**	**8**	**2**
21	Płocki	**8**	**7**	**9**
22	Płoński	**7**	**6**	**9**
23	Pruszkowski	**10**	**8**	**2**
24	Przasnyski	**3**	**3**	**8**
25	Przysuski	**3**	**2**	**5**
26	Pułtuski	**3**	**3**	**5**
27	Radom m.	**11**	**9**	**2**
28	Radomski	**10**	**8**	**9**
29	Siedlce .	**6**	**5**	**1**
30	Siedlecki	**7**	**6**	**9**
31	Sierpecki	**3**	**3**	**5**
32	Sochaczewski	**7**	**6**	**5**
33	Sokołowski	**3**	**3**	**7**
34	Szydłowiecki	**2**	**2**	**3**
35	m.st. Warszawa	**12**	**10**	**3**
36	Warszawski	**8**	**7**	**4**
37	Węgrowski	**4**	**3**	**8**
38	Wołomiński	**11**	**9**	**6**
39	Wyszkowski	**4**	**4**	**6**
40	Zwoleński	**1**	**1**	**4**
41	Żuromiński	**2**	**2**	**5**
42	Żyrardowski	**5**	**5**	**4**

Tabela IV.1. Podsumowanie wyników dotyczących analizy jednorodności powiatów mazowieckich ze względu na trzy różne kryteria (zestawy cech).

Tabela IV.2. Podsumowanie podziału na zbiory jednorodne dla powiatów należących do czterech par województw

Pary województw	Grupa jednorodności województwa	Liczba powiatów	Powiaty należące do zbiorów wspólnych		Powiaty stanowiące oddzielne zbiory		Ogólna liczba zbiorów	Liczba zbiorów wspólnych
			W liczbach	Procentowy udział	Liczba zbiorów	Liczba powiatów		
Mazowieckie	**4**	**42**	**33**	**79%**	**3**	**9**	**15**	**10**
Śląskie	**4**	**36**	**29**	**81%**	**2**	**7**		
Mazowieckie	**4**	**42**	**22**	**52%**	**5**	**20**	**14**	**5**
Łódzkie	**3**	**24**	**16**	**67%**	**4**	**8**		
Mazowieckie	**4**	**42**	**19**	**45%**	**5**	**23**	**12**	**4**
Pomorskie	**2**	**20**	**13**	**65%**	**3**	**7**		
Mazowieckie	**4**	**42**	**16**	**38%**	**6**	**26**	**14**	**5**
Świętokrzyskie	**1**	**14**	**8**	**57%**	**3**	**6**		

Książka poświęcona jest opisowi zastosowań metod sformalizowanych do wybranych zagadnień społeczno-gospodarczych i administracyjnych o charakterze przestrzennym. Rozpatrywane są zagadnienia regionalizacji i typologii przestrzennej, logistyki i organizacji transportu, zrównoważonego rozwoju, czy jakości stron internetowych samorządów w zestawieniu z położeniem odpowiednich jednostek.

ISSN 0208-8029
ISBN 9788389475251

Instytut Badań Systemowych PAN

W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy prosimy o kontakt z Instytutem Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa
tel. (22) 3810 277; e-mail: bibliote@ibspan.waw.pl