

P O L S K A A K A D E M I A N A U K
I N S T Y T U T G E O G R A F I I
I P R Z E S T R Z E N N E G O Z A G O S P O D A R O W A N I A

ZESPÓŁ KOORDYNACYJNY PROBLEMU MIĘDZYRESORTOWEGO
„PODSTAWY PRZESTRZENNEGO ZAGOSPODAROWANIA KRAJU“

Do użytku służbowego

BIULETYN INFORMACYJNY
ZESZYT 23

KRYSTYNA BIELECKA, FRANCISZEK A. SZCZOTKA

**BADANIA NAD OCENĄ STOSOWALNOŚCI
METOD ILOŚCIOWYCH W TYPOLOGII
ROLNICTWA**

WARSZAWA 1978

P O Ł S K A A K A D E M I A N A U K
I N S T Y T U T G E O G R A F I I
I P R Z E S T R Z E N N E G O Z A G O S P O D A R O W A N I A

ZESPÓŁ KOORDYNACYJNY PROBLEMU MIĘDZYRESORTOWEGO
„PODSTAWY PRZESTRZENNEGO ZAGOSPODAROWANIA KRAJU”

Do użytku służbowego

BIULETYN INFORMACYJNY

ZESZYT 23

KRYSTYNA BIELECKA, FRANCISZEK A. SZCZOTKA

BADANIA NAD OCENĄ STOSOWALNOŚCI
METOD ILOŚCIOWYCH W TYPOLOGII
ROLNICTWA

WARSZAWA 1978

Opracowanie redakcyjne: Irena Stańczak

SPIS TREŚCI

strona

Przedmowa - K. Bielecka	
CZĘŚĆ I.	
Analiza i ocena stosowalności metod ilościowych w ba- daniach przestrzennych rolnictwa - Raport z badań - K. Bielecka	
1. Przesłanki badań	
2. Zakres badań: Jakie metody i dlaczego?	
3. Stan wiedzy o metodach	
4. Specyfika metod taksonomicznych	
5. Ukierunkowanie badań	
Aneks	
CZĘŚĆ II.	
Analiza porównawcza wybranych metod taksonomicznych - na materiale empirycznym - F.A. Szczotka, L. Łukaszew- ska	
1. Wstęp	
2. Procedury grupowania	
2.1. Definicje i oznaczenia	
2.2. Metoda najbliższego sąsiada	
2.3. Metoda najdalszego sąsiada	
2.4. Metoda łączonych par grup	
2.5. Huberta metoda podziałowa A	
2.6. Huberta metoda podziałowa B	

3. Metody porządkowania	
3.1. Konstrukcja najkrótszego dendrytu	
3.2. Uporządkowanie liniowe metodą Szczotki	
4. Redukcja wymiarów metodą składowych głównych	
5. Kryteria dobroci podziału	
5.1. Wskaźniki charakteryzujące liczebności grup	
5.2. Wskaźniki charakteryzujące jednorodność grup	
6. Omówienie wyników %	
6.1. Analiza odległości	
6.2. Analiza hierarchii podziałów	
7. Analiza porównawcza jakości metod	
7.1. Wskaźniki charakteryzujące liczebności grup	
7.2. Wskaźniki charakteryzujące jednorodność grup	
8. Wnioski	
Literatura	
Spis tablic	
Spis rycin	

PRZEDMOWA

Celem publikacji jest poinformowanie o przedmiocie badań, programie i sposobie ich wykonywania, stopniu zaawansowania w rozwiązywaniu problemów merytorycznych oraz przekazanie fragmentu wyników badań - dotyczących prac nad oceną możliwości efektywnego wykorzystania aparatu metod ilościowych dla potrzeb analizy przestrzennej rolnictwa. Badania prowadzono w grupie tematycznej O3 "Struktura przestrzenna wyżywienia i rolnictwa" w problemie "Podstawy przestrzennego zagospodarowania kraju".

Ocenianie metod ilościowych pod kątem ich stosowalności w analizie przestrzennej rolnictwa jest przedsięwzięciem pionierskim. Ten rodzaj wiedzy jest nowy zarówno dla matematyków jak i specjalistów branżowych, stojących przed perspektywą konieczności zastosowania metod ilościowych w swoich badaniach. Obiektywna trudność efektywnego stosowania metod ilościowych w nowych dla nich dziedzinach dyscyplin nietechnicznych nakazuje nie rezerwę wobec tych metod lecz przeciwnie - dokładne poznanie ich działania. Jest to niezbędny warunek opracowywania planu badań naukowych, zakładających korzystanie z aparatu metod ilościowych.

Zdając sobie sprawę z obiektywnej trudności samej matematyki jak też z istnienia u wielu nieprofesjonalistów trudnej

do pokonania bariery matematyki, a stąd i małych szans jej upowszechnienia na poziomie pozwalającym na samodzielne stosowanie, autorzy omawianych badań starają się redagować swoje opracowania w ten sposób, by mogły one być zrozumiałe dla czytelników o różnym stopniu znajomości matematyki.

Część pierwsza publikacji ma charakter wprowadzenia metodologicznego. Zapoznanie się z jej treścią ułatwi śledzenie całości toku badań, których wyniki będziemy sukcesywnie udostępniać.

Część druga zawiera omówienie wyników zastosowania siedmiu wybranych metod taksonomicznych, których podstawy teoretyczne zostały przedstawione w pracy F.A. Szczotki "Podstawy taksonomii numerycznej" (Biuletyn ZKPW "Podstawy przestrzennego zagospodarowania kraju", z. 17, 1977, s. 80, IGI PZ PAN).

Opracowania zamieszczone w tym zeszycie Biuletynu należy traktować jedynie jako doniesienie cząstkowe i wyniki prac w toku. Bardziej całościowo jest tu zarysowana problematyka metodologiczna, natomiast sprawa oceny metod - mająca na celu dostarczenie praktycznej wiedzy o efektywności określonych ilościowych metod klasyfikacji - daleka jest od wyczerpania. Wyniki badań nad tymi problemami będą przekazywane w miarę postępu prac.

Część wyników tych badań, obejmująca siedem następnych wybranych ilościowych metod klasyfikacji, będzie opublikowana w Studiach KPZK PAN, t. LXIX, 1979.

Krystyna Bielecka

CZĘŚĆ I

Krystyna Bielecka

ANALIZA I OCENA STOSOWALNOŚCI METOD ILOŚCIOWYCH W BADANIACH PRZESTRZENNYCH ROLNICTWA

- RAPORT Z BADAŃ -

1. Przesłanki badań

Niekwestionowane walory¹ metod ilościowych zachęcają do stosowania tych metod w badaniach dyscyplin nietechnicznych szczególnie wtedy gdy złożoność analizowanego problemu wymaga operowania dużym i trudnym do interpretacji materiałem obserwacyjnym, a uściślenie wyników ma zasadnicze znaczenie dla poznania czy podejmowania decyzji. Jednocześnie obiektywna trud-

¹ Obiektywność działania, jednoznaczność uzyskiwanych wyników i ich precyzja, powtarzalność określonego toku postępowania na różnym materiale empirycznym, tożsamość uzyskiwanych wyników niezależnie od podmiotu stosującego metodę, miejsca i czasu jej stosowania, możliwość stosowania technik wymagających dużej ilości obliczeń oraz możliwość posługiwania się elektroniczną techniką obliczeniową wydatnie zmniejszającą pracochłonność i przyspieszającą uzyskiwanie wyników.

ność efektywnego stosowania metod ilościowych, czego dowodem są dość liczne przykłady zniechęcających wyników zastosowań, powoduje rezerwę wobec tych metod.

Efektywność stosowania metod ilościowych uwarunkowana jest spełnieniem określonych wymogów. Podstawowe znaczenie ma tu interdyscyplinarna metoda pracy, tj. współpraca matematyka ze specjalistą branżowym (zainteresowanym wykorzystaniem aparatu metod ilościowych) przy stawianiu i rozwiązywaniu badanego zadania. Zadaniem specjalisty branżowego jest określenie przedmiotu i celu badań oraz sformułowanie pytań na jakie zamierzone badania mają odpowiedzieć. Zadaniem matematyka jest dobór metody ilościowej, która zapewniałaby prawidłową odpowiedź na postawione pytania. Wymaga to wnikliwego spojrzenia na sprecyzowany przez specjalistę przedmiot i cel badań, a także weryfikację rodzaju informacji z punktu widzenia jej adekwatności do przedmiotu i celu badań.

Praktyka dowodzi, że niekiedy bardzo trudnym zadaniem jest sformułowanie pytań przez specjalistę, wówczas pomijając to ważne - z metodologicznego punktu widzenia - ogniwo procesu poznawczego, podejmuje się badania, kierując się intuicją i domniemaniem.

Skutki prowadzenia badań w sytuacji braku pytań, niedostatecznej precyzji w określeniu przedmiotu i celu badań, niewłaściwego doboru informacji, przejawiają się we wzbudzających wątpliwość wynikach badań. Wówczas to niejednokrotnie pada pod adresem metod ilościowych zarzut "matematyka zawiodła"; na co matematycy zwykle odpowiadają "matematyka nie kłamie i nie zna litości". Powiedzenie to zawiera myśl, że wątpliwe wyniki są pewnego rodzaju testem zarówno właściwości zastosowanej metody ilościowej jak i metodologicznej strony podjętych

badania oraz, że wskazują one badaczom potrzebę dokonania weryfikacji całego programu badań. Przy czym należy tu mieć na uwadze obowiązującą zasadę mówiącą, że - im bardziej precyzyjna metoda ilościowa tym bardziej dokładna znajomość morfologii badanego przedmiotu.

Interdyscyplinarna metoda pracy wymaga, by specjalista branżowy dysponował pewnym zasobem wiedzy o właściwościach potencjalnie użytecznych dla niego metod ilościowych, a matematyk posiadał dostateczną znajomość badanego problemu. Stosunkowo świeża data uświadamiania sobie tej konieczności jest przyczyną szczupłego stanu wiedzy o efektywności metod ilościowych w zastosowaniach praktycznych, co powoduje, że sięgając do aparatu metod ilościowych w celu wykorzystania go w analizie przestrzennej rolnictwa nie mamy możliwości czerpania z istniejącej wiedzy i cudzych doświadczeń w tym względzie, lecz zachodzi potrzeba podjęcia rozpoznania od podstaw.

Prace w tym zakresie podjęłam z inicjatywy prof. Jerzego Kostrowickiego, za którego sprawą problematyka oceny metod ilościowych została wprowadzona do programu badań Komisji Problematyki Przestrzennej Rolnictwa Komitetu Przestrzennego Zagospodarowania Kraju PAN, a następnie także do planu badań problemu węzłowego 11.2.1. w grupie tematycznej O3 tematu 1.A.

Badania prowadził zespół: trzech matematyków - doc. dr F.A. Szczotka, dr Z. Piasecki i mgr M. Paprzycki oraz geograf rolnictwa dr K. Bielecka, kierująca całością prac. Konsultantami prac zespołu są: matematyk prof. dr M. Warmus i geograf rolnictwa prof. dr J. Kostrowicki.

2. Zakres badań: Jakie metody i dlaczego?

Postawionym przez geografów rolnictwa zadaniem było dać odpowiedź na pytanie: która z metod ilościowych nadawałaby się najlepiej dla celów typologii rolnictwa.

x

Według definicji encyklopedycznej - typologia jest zabiegiem systematyzującym, polegającym na wyróżnianiu i grupowaniu obiektów podobnych do siebie pod względem określonych właściwości. Postępowanie to określa się w języku matematycznym (upraszczając rzecz) jako klasyfikację zbioru obiektów wielocechowych (opisanych wieloma zmiennymi).

Zatem rodzaj metod ilościowych, który podlegałby ocenie, wyznaczony jest przez główne problemy metodologiczne typologii rolnictwa, którymi są:

- 1) dobór i właściwe wyrażenie cech diagnostycznych typu rolnictwa,
- 2) metoda grupowania zbioru obiektów (jednostek przestrzennych) mająca zapewnić porównywalność wyników w przestrzeni i w czasie.

Z tego punktu widzenia właściwymi dla typologii są:

- 1) metody analizy czynnikowej służące do:
 - analizy zależności między cechami diagnostycznymi;
 - analizy podobieństw między obiektami;
- 2) metody taksonomii numerycznej zajmujące się:
 - klasyfikacją, tj. podziałem niejednorodnego zbioru obiektów wielocechowych na grupy obiektów podobnych - ze względu na badane cechy;
 - porządkowaniem, tj. sformalizowanym przedstawieniem struktury podobieństw między obiektami analizowanego zbioru.

3. Stan wiedzy o metodach

W momencie podejmowania przez nas badań problem ilościowej metody grupowania zbioru wielocechowego był otwarty i zarazem interesujący wielu badaczy. Pod koniec lat 60-tych pojawiła się w literaturze światowej ogromna różnorodność procedur grupowania, lecz przy braku zadowalającej teorii klasyfikacji nie było i nie ma do dziś jasności co do wyższości - z formalnego punktu widzenia - którejś z nich. Panuje więc pogląd, że najważniejsze z praktycznego punktu widzenia jest eksperymentowanie z wybranymi - pod kątem potrzeb danej dyscypliny - metodami do chwili aż stanie się jasne, która z nich jest najodpowiedniejsza do rozwiązania danego zagadnienia. W tej sytuacji wskazanie metody najlepszej dla potrzeb typologii rolnictwa wymagało przede wszystkim przetestowania wielu z nich na materiale empirycznym. Stawialiśmy w tym samodzielne kroki bez możliwości korzystania z cudzych ustaleń. A o tym jak skromna była nasza wiedza o efektywności metod taksonomicznych świadczy, że nie mogliśmy na początku naszych badań uściślić co należy rozumieć pod określeniem "podział najlepszy". Treść tego określenia formowała się w toku badań. Początkowo jako kryterium oceny dobroci metody przyjęliśmy ostrość podziałów.

Badania nasze wykonywane dla potrzeb typologii rolnictwa służą nie tylko celom tej typologii lecz dostarczają również wskazań metodycznych i metodologicznych wszelkim analizom, których celem jest wyodrębnienie obszarów podobnych - ze względu na zespół rozpatrywanych cech.

4. Specyfika metod taksonomicznych

Przy stosowaniu poszczególnych metod taksonomicznych uzys-

kuje się inny podział tego samego zbioru. Dlaczego tak jest?

x

Odpowiadając krótko na postawione pytanie można powiedzieć, że przyczyna tkwi w specyfice konstrukcji tych metod.

Elementami konstrukcji każdej metody taksonomicznej są:

- 1) miara podobieństwa obiektów,
- 2) kryterium podziału zbioru,
- 3) algorytm (plan czynności) podziału zbioru.

Miar podobieństwa, kryteriów podziału zbioru i algorytmów podziału jest wiele. Doborem ich rodzaju różnią się poszczególne metody (słabą stroną jest tu to że wybór miary podobieństwa i kryterium podziału zbioru są dokonywane arbitralnie, ponieważ nie ma żadnych matematycznych zaleceń w tej kwestii.). Zmieniając tylko jeden z tych elementów możemy uzyskać różne podziały tego samego zbioru. Wpływ na wyniki grupowania ma również sposób normalizacji cech diagnostycznych, dla wyboru którego, nie ma żadnych matematycznych wskazań.

5. Ukierunkowanie badań

W trakcie prowadzenia badań zarysowały się trzy ich wątki. Pierwszy z nich to ocena stosowalności "popularnych", tj. znanych i propagowanych w literaturze rodzimej i światowej, metod analizy wielu zmiennych (multivariate analysis methods). Uzasadnieniem takiego programu badań była chęć uniknięcia tego co wyraża skrót myślowy "wyważania otwartych drzwi", a więc ocena:

- 1) czy i z jakim efektem da się wykorzystać w typologii rolnictwa owe "popularne" metody;
- 2) które z proponowanych w literaturze światowej rozwiązań metodycznych typologii czy regionalizacji rolnictwa można by

przyjąć, bez konieczności modyfikacji, jako efektywne dla celów opracowywanej w Polsce metody typologii rolnictwa.

W badaniach naszych przyjęliśmy dwa sposoby przeprowadzania analizy metod, a mianowicie:

- 1) testowanie wybranych metod ilościowych na naszym materiale empirycznym oraz
- 2) ocenianie metod na podstawie ich zastosowań przez innych autorów.

Wyniki rozpoznawania metod na podstawie własnych prób zastosowań przedstawiamy w następujących opracowaniach syntetyzujących:

1. Podstawy taksonomii numerycznej, F.A. Szczotka, Biuletyn Informacyjny, ZKPW "Podstawy Przestrzennego Zagospodarowania Kraju", z. 17, 1977, s. 80, IGiPZ PAN.
2. Ocena stosowalności wybranych metod ilościowych w typologii rolnictwa, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis, s. 99 + 19 tabel + 17 rycin, IGiPZ PAN archiwum problemu, sygn. A. 711; druk: Studia KPZK PAN, t. LXIX, 1979, PWN.
3. Analiza porównawcza wybranych metod taksonomicznych na materiale empirycznym, F.A. Szczotka, L. Łukaszewska: redakcja naukowa F.A. Szczotka, maszynopis, s. 39 + 31 tabel + 4 ryciny, IGiPZ PAN archiwum problemu, sygn. A. 835/4; druk: w niniejszym Biuletynie, cz. II.

Wyniki rozpoznawania metod przez badaczy radzieckich, brytyjskich i norweskich przedstawiamy w postaci tłumaczenia wyboru prac opublikowanych w Przeglądzie Zagranicznej Literatury Geograficznej, z. 3-4, 1976, s. 151, IGiPZ PAN, "Metody matematyczne w badaniach struktury przestrzennej rolnictwa".

Stosując metodę analizy logicznej wykonano opracowanie: Ocena zastosowanych w pracy Nordgarda pt. "Region-Building a Comparison of Methods" metod z punktu widzenia przydatności ich dla typologii rolnictwa, M. Paprzycki, maszynopis s. 11, Zakład Geografii Rolnictwa IGiPZ PAN.

Pracą w toku jest prowadzona na różnych materiałach obserwacyjnych analiza taksonomii wrocławskiej.

Prace nad oceną stosowalności metod "popularnych" w analizie przestrzennej rolnictwa będziemy kontynuować w miarę pojawiania się nowych, interesujących metod - z tym, że będzie to dyktowane raczej potrzebą "bycia w kursie" niż zamiarem korzystania z gotowych wzorów. Jak mogliśmy bowiem stwierdzić na podstawie prób zastosowań, wybrane metody "popularne" (Aneks) dają stosunkowo mało zadowalające wyniki grupowania, a przy tym żadna z nich nie spełnia wymogu porównywalności wyników grupowania zbiorów w przestrzeni i w czasie.

Wobec braku możliwości efektywnego korzystania z "gotowych" metod ilościowych, punkt ciężkości naszych badań przesuwa się na poszukiwanie własnych, lepszych od istniejących, rozwiązań problemu ilościowej metody typologii rolnictwa.- jest to drugi wątek naszych prac. W pracach tych zarysowują się dwa główne zadania:

- 1) zbudowanie zadowalającej metody klasyfikacji zbioru,
- 2) zbudowanie metody zapewniającej porównywalność wyników grupowania zbiorów w przestrzeni i w czasie.

Dotychczasowymi wynikami naszych prac w zakresie zadania pierwszego są trzy nowe metody taksonomiczne:

- 1) metoda ORLINE - opracowana przez Z. Piaseckiego,
- 2) metoda GRAWITACYJNA - opracowana przez M. Paprzyckiego,
- 3) metoda FARELL-mod - jest to opracowana przez M. Paprzyckiego

modyfikacja metody FARELL.

Opis tych metod oraz wyniki prób ich zastosowań na materiale empirycznym podane są w niżej wymienionych opracowaniach:

- 1) Ocena stosowalności wybranych metod ilościowych w typologii rolnictwa, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis, s. 99 + 19 tabel + 17 rycin, IGiPZ PAN archiwum problemu, sygn. A. 7-11, druk: Studia KPZK, t. LXIX, 1979, PWN.
- 2) An evaluation of applicability of selected mathematical methods for the typology of agriculture, K. Bielecka, M. Paprzycki, Z. Piasecki, /w:/ Agricultural Typology and Lands Utilization, 1975, ss. 55-70, Werona.
- 3) Proposals of new taxonomic methods for agricultural typology, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis, s. 19, druk: Geographia Polonica, t. 41, 1979.
- 4) Z badań nad stosowalnością metod ilościowych w typologii rolnictwa - Propozycja nowych metod taksonomicznych, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis s. 19 + 4; druk: Przegląd Geograficzny 1979 r.

Czy nasze metody są dobre?

Na pytanie to możemy odpowiedzieć następująco: przy zadanych nam zestawie cech opisujących obiekty (jednostki przestrzenne) analizowanych zbiorów są to, jak się wydaje, metody skuteczniejsze od analizowanych metod "popularnych". Jednakże mimo przeprowadzenia serii prób zastosowań naszych nowych metod nie mamy jeszcze jednoznacznego poglądu na to, która z nich jest skuteczniejsza. Żadna z nich jednak nie spełnia warunku porównywalności wyników grupowania zbiorów w przestrzeni i w czasie.

Warunek ten spełnia jedynie nowa, opracowana przez Z. Piaseckiego, identyfikacyjno-weryfikacyjna metoda IDWER. W porówna-

niu z omówionymi metodami cechuje ją odmienność podejścia. W przeciwieństwie do tych metod, które działają na zbiorach zamkniętych (jeden obszar) w ujęciu statycznym (jeden przekrój czasowy) metoda IDWER działa na zbiorze otwartym, umożliwiając klasyfikację nieograniczonej liczby obiektów, dzięki czemu uzyskuje się porównywalność w przestrzeni. Porównywalność w czasie jest tu równoważna do porównywalności w przestrzeni, ponieważ każdy obiekt opisywany w dwóch różnych przekrojach czasowych można traktować formalnie jako dwa obiekty. Możliwe to jest przy przyjęciu określonych założeń.

Istniejącą postacią metody IDWER można uznać jako propozycję formalnego zapisu idei rozwiązania zagadnienia porównywalności wyników podziału w przestrzeni i w czasie. Uzyskane próby zastosowań nie podważają idei metody IDWER ani technicznej strony jej działania. Kierują natomiast uwagę na potrzebę rozwiązania wielu problemów, nad czym obecnie pracujemy. Opis metody IDWER i jej zastosowań zawierają następujące opracowania:

- 1) Propozycja nowych metod taksonomicznych dla celów typologii rolnictwa, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis, s. 16, Zakład Geografii Rolnictwa IGiPZ PAN.
- 2) Proposal of a new identification-verification method to be used in agricultural typology, Z. Piasecki, [w:] Agricultural Typology and Land Utilisation, 1975, ss. 269-276, Werona.
- 3) A typology of Poland's Agriculture - A Study of Application of the IDWER Method, Z. Piasecki, maszynopis powielony, s. 8 + 1, 1975, Zakład Geografii Rolnictwa IGiPZ PAN.
- 4) Proposal of new taxonomic methods for agricultural typology, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis, s. 19, Zakład Geografii Rolnictwa IGiPZ PAN, druk: Geographia Polonica, t. 41, 1979.

W trzeciej wątek naszych badań układają się - ujawnione w trakcie prac - problemy metodologii diagnostyki metod klasyfikacji zbiorów.

Aby powiedzieć, który z dwóch podziałów tego samego zbioru jest lepszy, należy posłużyć się odpowiednim kryterium. Wyłania się tu ważny i zarazem otwarty problem formalnej oceny dobroci podziału zbioru.

Jak dotychczas, matematyka nie dysponuje jeszcze odpowiednim formalnym kryterium. "Zastępczo" ocenę formalną opiera się na analizie logicznej instrukcji metody klasyfikacji, materiału obserwacyjnego oraz wyników grupowania, oraz na intuicji - zdając sobie sprawę z ograniczonego zakresu stawianej na tej podstawie diagnozy. Ten mało zadowalający sposób oceny dobroci podziału zbioru należy zastąpić odpowiednią metodą formalną. Jest to zadanie trudne do rozwiązania.

Podjęte przez nas prace nad stworzeniem ilościowej metody formalnej oceny dobroci podziału zbioru znajdują się na wstępnym etapie. Prowadzimy je metodą kolejnych przybliżeń. Wyniki tego etapu prac przedstawione są w opracowaniach:

- 1) Zagadnienie względnej skuteczności metod taksonomii numerycznej, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis s. 10, IGIPIZ PAN.
- 2) Formalna ocena względnej skuteczności metod taksonomii numerycznej - Podjęcie zagadnienia metody oceny, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis, s. 12 + Aneks s. 12, IGIPIZ PAN, archiwum problemu, sygn. A. 1108/2.
- 3) O metodach oceny efektywności metod taksonomii numerycznej, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis, s. 4, 1978, Zakład Geografii Rolnictwa IGIPIZ PAN.
- 4) On methods applied in the evaluation of the effectiveness

of numeric taxonomy methods - Summary, K. Bielecka, M. Paprzycki, Z. Piasecki, maszynopis, s. 5, 1978, Zakład Geografii Rolnictwa IGiPZ PAN.

- 5) Badania nad stosowalnością metod ilościowych w typologii rolnictwa - Refleksje metodologiczne, K. Bielecka, M. Paprzycki, Z. Piasecki, Przegląd Geograficzny, t. XLIX, z. 4, 1977, ss. 807-817.
- 6) Analiza porównawcza wybranych metod taksonomicznych na materiale empirycznym, F.A. Szczotka, L. Łukaszewska, redakcja naukowa F.A. Szczotka, maszynopis, s. 39 + 31 tabel + 4 rycin, IGiPZ PAN archiwum problemu, sygn. A. 835/4; druk w niniejszym Biuletynie, cz. II.

A N E K S

Poddane ocenie "popularne" metody klasyfikacji zbioru obiektów wielocechowych:

1. Technika R analizy czynnikowej
2. Technika Q analizy czynnikowej
3. Metoda Berry'ego - grupowanie oparte na wyjściowych wartościach cech opisujących
4. Metoda Berry'ego - grupowanie oparte na składowych Hotellinga
5. Metoda FARELL
6. Metoda najbliższego sąsiada
7. Metoda najdalszego sąsiada
8. Metoda par grup
9. Huberta metoda podziałowa A
10. Huberta metoda podziałowa B
11. McQuitty'ego metoda analizy powiązań jednostkowych
12. McQuitty'ego metoda powiązań ograniczonych
13. McQuitty'ego modyfikacja metody Berry'ego - metoda zastępowania odległością maksymalną
14. Metoda Warda
15. Metoda graficzna grupowania w układzie dwóch najistotniejszych czynników wspólnych
16. Taksonomia wrocławska

CZĘŚĆ II

Franciszek A. Szczotka, Lidia Łukaszewska

Redakcja Naukowa - Franciszek A. Szczotka

ANALIZA PORÓWNAWCZA WYBRANYCH METOD TAKSONOMICZNYCH

- NA MATERIALE EMPIRYCZNYM -

1. Wstęp

W ostatnich latach nastąpił bardzo intensywny rozwój tzw. taksonomii numerycznej. Ujawnia się on przede wszystkim w powstaniu dużej ilości metod grupowania, tj. metod mających za zadanie podział zbioru niejednorodnego na grupy jednorodne.

Zadanie to przy sprecyzowanym kryterium optymalności podziału już dla umiarkowanej liczby obiektów porządkowanych przedstawia zadanie numeryczne na ogół nie do wykonania nawet przy użyciu maszyn szybko liczących. Istnieją jednak algorytmy podziału oparte na przesłankach heurystycznych, które dają podziały zadowalające.

Znane są niektóre właściwości formalne tych metod, natomiast wiele właściwości charakteryzujących ich przydatność

praktyczną nie wynika z opisu algorytmu.

Celem niniejszego studium jest porównanie wybranych metod taksonomicznych na empirycznym materiale obserwacyjnym.

Porównanymi metodami są:

- metoda najbliższego sąsiada (Johnson, 1967, single linkage method) identyczna z tzw. taksonomią wrocławską (Florek i inni, 1951), zwana dalej w skrócie SIM,
- metoda najdalszego sąsiada (Johnson, 1967, complete linkage method), zwana w skrócie CIM,
- metoda łączonych par grup (Sokal, 1963) oznaczona dalej przez MPG,
- Huberta metoda podziałowa A (Hubert, 1973) oznaczona przez MPA, oraz
- Huberta metoda podziałowa B (Hubert, 1973) oznaczona przez MPB.

Metody te opisane są w p. 2 opracowania (por. także Szczotka, 1976).

Porównanie wymienionych metod przeprowadzono na dwóch zbiorach danych liczących po 40 powiatów Polski. W układzie przestrzennym zbiór pierwotny stanowi pas biegnący z północy na południowy wschód Polski (tabl. 1), zaś zbiór drugi - pas biegnący z północnego wschodu na południowy zachód Polski (tabl. 2).

Powiaty obu zbiorów opisano 21 cechami typologicznymi rolnictwa (tabl. 3).

Podobieństwo między powiatami określano za pomocą odległości euklidesowej $d(A_k, A_2)$ obliczonej na podstawie wartości unormowanych cech według wzoru:

$$d^2(A_k, A_l) = d_{kl}^2 = \sum_{i=1}^p \frac{(x_{ik} - x_{il})^2}{s_i^2}$$

gdzie

p jest liczbą cech,

x_{ik} , x_{il} wartością i -tej cechy dla k -tego i l -tego powiatu,

s_i - średnie kwadratowe odchylenia i -tej cechy.

Dla każdego podziału obliczono następujące wskaźniki ilościowe:

- liczba grup jednoelementowych,
- liczba obiektów w grupie najliczniejszej,
- wariancja liczby obiektów,
- suma kwadratów odległości wewnątrz grup,
- suma przeciętnych kwadratów odległości wewnątrz grup,
- suma średnich kwadratów odległości wewnątrz grup,
- suma kwadratów odległości wewnątrz grup od środków ciężkości (centroidów) grup,
- ważona suma kwadratów odległości między środkami ciężkości grup,
- suma kwadratów odległości między środkami ciężkości grup, podzielona przez sumę kwadratów odległości od centroidów wewnątrz grup.

Wskaźniki te stanowią charakterystyki metod podziału i są podstawą ich porównywania.

Dla przybliżonego zilustrowania wzajemnego położenia rozważanych powiatów, traktowanych jako punkty przestrzeni 21-wymiarowej o współrzędnych równych wartościom unormowanych cech, przeprowadzono redukcję wymiarów do dwóch zmiennych za

pomocą Hotellinga składowych głównych. Opis tego postępowania podano w p. 4.

Poza tym, że zbiory 1 i 2 pogrupowane zostały wymienionymi poprzednio metodami, uporządkowano oba te zbiory w dwojaki sposób:

- w sposób liniowy metodą Szczotki (Szczotka, 1972) oraz
- w sposób nieliniowy za pomocą najkrótszego dendrytu (dendrytu wrocławskiego), (Florek i inni, 1951). Opis tych metod podano w p. 3.

2. Procedury grupowania

2.1. Definicje i oznaczenia

W dalszym ciągu używane będą następujące oznaczenia i pojęcia:

Badane obiekty (powiaty): A_1, A_2, \dots, A_n .

Zbiór badanych obiektów: $G_1 = \{A_1, A_2, \dots, A_n\}$.

Zbiór liczb naturalnych: N

Zbiór indeksów: $I_m = \{i_1, i_2, \dots, i_m\}$, $i_k \in N$.

O zbiorze indeksów zakłada się, że $i_k < i_l$ dla $k < l$.

$$I_m - \{i_l\} = \{i_1, i_2, \dots, i_{l-1}, i_{l+1}, \dots, i_m\}.$$

Podział zbioru G na m rozłącznych grup oznaczono przez

$$P^{(m)} = \{G_1^{(m)}, G_2^{(m)}, \dots, G_m^{(m)}\}.$$

Podział spełnia następujące warunki:

$$G_k^{(m)} \cap G_l^{(m)} = \emptyset \quad \text{dla } k \neq l$$

$$G_1^{(m)} \cup G_2^{(m)} \cup \dots \cup G_m^{(m)} = G,$$

gdzie $G_k \cap G_l$ oznacza część wspólną, a $G_k \cup G_l$ połączenie (sumę) zbiorów G_k i G_l . Symbol \emptyset oznacza zbiór pusty.

Ciąg podziałów $H = \{P^{(1)}, P^{(2)}, \dots, P^{(n)}\}$ tworzy hierarchię, jeżeli spełnione są warunki:

1. $P^{(1)} = \{G_1^{(1)}\}$ gdzie $G_1^{(1)} = G$

2. Dla każdego $u < v$

$$G_k^{(u)} \cup G_l^{(v)} = G_k^{(v)} \quad \text{lub} \quad G_k^{(u)} \cap G_l^{(v)} = \emptyset.$$

Warunek drugi oznacza, że grupa w podziale drobniejszym albo zawiera się w grupie podziału grubszego albo jest z nią rozłączna.

Hierarchie podziałów otrzymuje się bądź przez łączenie grup bądź przez podział pojedynczych grup.

Połączenie (suma) grup $G_i^{(m)}$ i $G_j^{(m)}$, gdzie $i < j$, oznaczone jest przez $G_i^{(m-1)}$:

$$G_i^{(m)} \cup G_j^{(m)} = G_i^{(m-1)} \quad \text{dla } i < j.$$

Przy podziale grupy na dwie stosuje się następujące oznaczenia:

niech $G_i^{(m-1)} = G' \cup G''$ gdzie

$$G' = \{A_{1_1}, A_{1_2}, \dots, A_{1_{n_1}}\}$$

$$G'' = \{A_{j_1}, A_{j_2}, \dots, A_{j_{n_j}}\}$$

$$i = i_1 < j_1 = j,$$

wtedy $G_i^{(m)} = G'$

$$G_j^{(m)} = G''$$

Odległość między obiektami A_k i A_l oznaczona jest przez

$$d(A_k, A_l) = d_{kl}.$$

2.2. Metoda najbliższego sąsiada (single linkage method) - SIM

Metoda ta jest metodą łączeniową. Łączy ona tę parę grup, dla których odległość najbliższych obiektów jest najmniejsza.

Daje ona więc hierarchię podziałów $H = \{P^{(1)}, P^{(2)}, \dots, P^{(m)}\}$ według następujących zasad:

$$1. P^{(n)} = \{(A_1), (A_2), \dots, (A_n)\}.$$

2. Konstrukcja podziału $P^{(m-1)}$ z podziału $P^{(m)}$.

$$\text{Niech } P^{(m)} = \{G_{i_1}^{(m)}, G_{i_2}^{(m)}, \dots, G_{i_m}^{(m)}\}$$

$$I_m = \{i_1, i_2, \dots, i_m\}$$

$$d(G_k^{(m)}, G_l^{(m)}) = \min_{u, v \in I_m} d(G_u^{(m)}, G_v^{(m)})$$

$$k < l$$

wtedy $G_{i_k}^{(m-1)} = G_k^{(m)}$ $G_l^{(m)}$

$$G_i^{(m-1)} = G_i^{(m)} \text{ dla } i \in I_m \text{ oraz } i \neq k, l$$

$$I_{m-1} = I_m - \{l\}.$$

Można wykazać, że metoda najbliższego sąsiada dla każdego m daje podział, dla którego długość najkrótszego dendrytu łączącego grupy jest możliwie duża.

Metoda najbliższego sąsiada jest identyczna z metodą podziału najkrótszego dendrytu w taksonomii wrocławskiej (Florek i inni, 1951).

2.3. Metoda najdalszego sąsiada (complete linkage method) -CIM

Metoda ta jest także metodą łączeniową. Inaczej jednak niż w metodzie najbliższego sąsiada SIM, łączy ona tę parę grup, dla której odległość najdalszych obiektów jest najmniejsza. Metoda ta daje więc hierarchię podziałów

$$H = \{P^{(1)}, P^{(2)}, \dots, P^{(n)}\} \quad \text{według następujących}$$

zasad:

$$1. P^{(n)} = \{(A_1), (A_2), \dots, (A_n)\}.$$

2. Konstrukcja podziału $P^{(m-1)}$ z podziału $P^{(m)}$.

$$\text{Niech } P^{(m)} = \{G_{i_1}^{(m)}, G_{i_2}^{(m)}, \dots, G_{i_m}^{(m)}\}$$

$$I_m = \{i_1, i_2, \dots, i_m\}$$

$$D(G_k^{(m)}, G_l^{(m)}) = \min_{u, v \in I_m} d(G_u^{(m)}, G_v^{(m)})$$

$$k < l$$

$$\text{wtedy } G_k^{(m-1)} = G_k^{(m)} \cup G_l^{(m)}$$

$$G_i^{(m-1)} = G_i^{(m)} \quad \text{dla } i \neq k, l$$

$$I_{m-1} = I_m - \{l\}.$$

2.4. Metoda łączonych par grup Sokala i Michenera (pair group method) MPG

W obu poprzednich metodach decyzję o połączeniu grup podejmuje się na podstawie odległości najbliższych lub najdalszych obiektów obu grup. W metodzie par grup decyzję o połączeniu grup podejmuje się na podstawie średniej odległości między obiektami jednej i drugiej grupy.

W niniejszym opracowaniu posługiwano się wprost średnią arytmetyczną $\bar{d}(G_u, G_v)$ odległości między obiektami należącymi do różnych grup. Odległości określone są wzorem (1). Metoda ta daje hierarchię podziałów $H = \{P^{(1)}, P^{(2)}, \dots, P^{(n)}\}$ według następujących zasad:

$$1. P^{(n)} = \{(A_1), A_2, \dots, (A_n)\}.$$

2. Konstrukcja podziału $P^{(m-1)}$ z podziału $P^{(m)}$.

$$\text{Niech } P^{(m)} = \{G_{i_1}^{(m)}, G_{i_2}^{(m)}, \dots, G_{i_m}^{(m)}\}$$

$$I_m = \{i_1, i_2, \dots, i_m\}$$

$$\bar{d}(G_k^{(m)} + G_l^{(m)}) = \min_{u, v \in I_m} \bar{d}(G_u^{(m)}, G_v^{(m)})$$

$$k < l$$

$$\text{wtedy } G_k^{(m-1)} = G_k^{(m)} \cup G_l^{(m)}$$

$$G_i^{(m-1)} = G_i^{(m)} \quad \text{dla } i \neq k, l$$

$$I_{m-1} = I_m - \{l\}.$$

2.5. Huberta metoda podziałowa A, skrót MPA

Metoda ta konstruuje hierarchię podziałów drogą kolejnego dzielenia grup. Zaczyna ona więc od podziału jednogrupowego obejmującego wszystkie obiekty. Następnie zwiększa się krok po kroku liczbę grup, dzieląc wg określonego przepisu grupę najbardziej niejednorodną na dwie części. Sformalizowany opis tej metody jest następujący:

1. $P^{(1)} = \{G_1^{(1)}\} = \{(A_1, A_2, \dots, A_n)\}$.

2. Konstrukcja podziału $P^{(m+1)}$ z podziału $P^{(m)}$.

Niech $P^{(m)} = \{G_{i_1}^{(m)}, G_{i_2}^{(m)}, \dots, G_{i_m}^{(m)}\}$.

Najpierw znajduje się grupę, która ma być dzielona.

Niech $(G_k^{(m)}) = \max_{i \in I_m} \delta(G_i^{(m)})$

gdzie $\delta(G_i^{(m)}) = \max d(A_s, A_t)$, $A_s, A_t \in G_i^{(m)}$.

Grupę $G_k^{(m)}$ dzieli się na grupy G' i G'' , czyli

$$G_k^{(m)} = G' \cup G''.$$

Definicja grup G' i G'' jest następująca:

Niech A_{r_1} i A_{r_t} będą najbardziej oddalonymi obiektami w grupie $G_k^{(m)}$, tzn.

$$\delta(G_k^{(m)}) = d(A_{r_1}, A_{r_t})$$

wtedy A_{r_1} jest pierwszym obiektem w grupie G' a A_{r_t} pierwszym obiektem w grupie G'' , czyli na tym etapie

$$G' = (A_{r_1})$$

$$G'' = (A_{t_1}) .$$

Dalszy opis postępowania jest indukcyjny.

Niech \bar{G} oznacza zbiór obiektów grupy $G_k^{(r)}$ niezaliczonych jeszcze ani do G' ani do G'' ,

$$\bar{G} = G_k^{(m)} \cap \bar{G}' \cap \bar{G}''$$

gdzie \bar{G}' i \bar{G}'' oznaczają odpowiednio obiekty nie należące do G' lub do G'' , i niech A_s i A_w będą parą najbardziej odległych obiektów z grupy \bar{G} i $G' \cup G''$, tzn.

$$d(A_s, A_w) = \max_{A_u \in \bar{G}} d(A_u, A_v)$$

$$A_v \in G' \cup G''$$

wtedy, jeżeli $A_w \in G'$, to A_s włącza się do G'' ,

jeżeli $A_w \in G''$, to A_s włącza się do G' .

Pó rozdzieleniu całej grupy $G_k^{(m)}$ otrzymuje się podział

$$G_i^{(m+1)} = G_i^{(m)} \quad i < k$$

$$G_k^{(m+1)} = G'$$

$$G_{k+1}^{(m+1)} = G''$$

$$G_{i+1}^{(m+1)} = G_i^{(m)} \quad i > k$$

2:6. Huberta metoda podziałowa B, skrót MHP

Metoda ta różni się od metody podziałowej A zasadą dzielenia grupy $G_k^{(m)}$ na dwie: G' i G'' .

Modyfikacja jest następująca: niech A_S i A_W będą parą najbliższych obiektów z grup \bar{G} i $G' G''$, tzn.

$$d(A_S, A_W) = \min_{A_U \in \bar{G}} d(A_U, A_V)$$

$$A_V \in G' \cup G'',$$

Jeżeli $A_W \in G'$, to A_S włącza się do G' ,
jeżeli $A_W \in G''$, to A_S włącza się do G'' .

3. Metody porządkowania

Materiały obserwacyjne zostały uporządkowane w dwojaki sposób. Pierwsze uporządkowanie stanowi najkrótszy dendryt łączący wszystkie obiekty badanego zbioru; jest on częścią tzw. taksonomii wrocławskiej (Florek i inni, 1951). Drugie uporządkowanie jest liniowe, przeprowadzono je metodą Szczotki (Szczotka, 1972).

3.1. Konstrukcja najkrótszego dendrytu

Przez graf rozumie się w matematyce parę $\{G, R\}$, gdzie G jest skończonym zbiorem elementów a R jest zbiorem ich par. W naszym zagadnieniu G jest zbiorem porządkowanych obiektów, czyli powiatów, natomiast R jest zbiorem łączonych par. Przez dendryt rozumie się taki graf spójny, w którym połączenia nie tworzą krzywej zamkniętej. Z definicji tej wynika, że każda para obiektów połączona jest dokładnie jednym ciągiem połączeń.

Niech d_{ij} oznacza długość odcinka łączącego obiekty A_i i A_j , jeżeli para ta jest bezpośrednio z sobą połączona. Sumę długości wszystkich połączeń nazywa się długością dendrytu

Najkrótszy dendryt otrzymuje się następująco:

$$\text{niech } \min_{k \neq 1} d(A_k, A_1) = d(A_{i_1}, A_{i_2}).$$

Wtedy pierwszą parę połączoną stanowią obiekty A_{i_1} i A_{i_2} .

Przypuśćmy, że połączono już obiekty $A_{i_1}, A_{i_2}, \dots, A_{i_m}$.

Podobnie jak poprzednio oznaczmy przez $I_m = \{i_1, i_2, \dots, i_m\}$

zbiór indeksów. Niech

$$\min_{k \in I_m} d(A_k, A_1) = d(A_{i_j}, A_{i_{m+1}}), \quad j < m+1$$

$$i \in I_m$$

wtedy obiekt $A_{i_{m+1}}$ zostaje połączony z obiektem A_{i_j} .

Dendryt ma tę właściwość, że można go narysować z zachowaniem długości połączeń.

3.2. Uporządkowanie liniowe metodą Szczotki

Nieformalnie mówiąc, uporządkowanie liniowe polega na ustawieniu zbioru obiektów w kolejności. Uporządkowanie wyznaczone jest przez permutacje indeksów $(i_1, i_2, i_3, \dots, i_n)$. Dążeniem jest uzyskanie takiego uporządkowania, by odległości między kolejno po sobie następującymi obiektami były możliwie małe, a obiekty o dużych odległościach były oddalone także w uporządkowaniu.

Intuicje te można sformalizować następująco:

niech v oznacza oddalenie obiektów w uporządkowaniu. Obiekty

sąsiednie są oddalone o jeden, obiekty co drugie oddalone o dwa itd. Skrajne obiekty oddalone są o $v = n - 1$. Następnie dla dowolnego uporządkowania tworzy się sumę iloczynów faktycznych odległości obiektów i ich oddalenia w uporządkowaniu. Tę wielkość można napisać w postaci:

$$K(i_1, i_2, \dots, i_n) = \sum_{v=1}^{n-1} v \sum_{j=1}^{n-v} d(A_j, A_{j+v})$$

Dążeniem jest, by małym oddaleniom odpowiadały małe odległości, a dużym oddaleniom duże odległości. Za lepsze uporządkowanie z dwóch uznać więc należy to, dla którego wartość K jest większa.

Istnieje algorytm (Szczotka, 1972) na znalezienie takiego uporządkowania, którego nie da się poprawić przez przestawienie dwóch obiektów.

4. Redukcja wymiarów metodą składowych głównych

W interpretacji geometrycznej zespół rozważanych obiektów scharakteryzowany p cechami uważać można za zbiór punktów w przestrzeni p -wymiarowej.

Dla przybliżonego i uproszczonego graficznego przedstawienia zrzutowano zbiór badanych powiatów na najlepiej dopasowaną płaszczyznę dwuwymiarową. Płaszczyzna ta wyznaczona jest przez dwie pierwsze składowe główne Y_1 i Y_2 . Składowe te są pewnymi funkcjami liniowymi cech, określonymi wzorami

$$Y_1 = \sum_{k=1}^p a_{1k} x_k^* \quad (2)$$

$$Y_2 = \sum_{k=1}^p a_{2k} x_k^*$$

lub w zapisie wektorowym

$$Y_1 = \underline{a}_1' \underline{x}^*$$

$$Y_2 = \underline{a}_2' \underline{x}^*$$

gdzie $\underline{x}^* = (x_1^*, x_2^*, \dots, x_p^*)'$ jest wektorem wartości cech unormowanych na średnią zero i dyspersję jeden.

Wektor $\underline{a}_1 = (a_{11}, a_{12}, \dots, a_{1p})$ jest wektorem własnym macierzy korelacji cech R, odpowiadającym największej wartości własnej λ_1 .

Wektor $\underline{a}_2 = (a_{21}, a_{22}, \dots, a_{2p})$ jest wektorem własnym macierzy korelacji R, odpowiadającym drugiej wartości własnej λ_2 .

Wzory (2) i (3) wyznaczają w przestrzeni p-wymiarowej pewną płaszczyznę $\Pi_{Y_1 Y_2}$ przechodzącą przez początek układu.

Wzory te określają równocześnie współrzędne rzutu A_i punktu A_i na płaszczyznę $\Pi_{Y_1 Y_2}$.

Suma kwadratów odległości punktu A_i od tej płaszczyzny jest równa.

$$\sum_{i=1}^n d^2(A_i, \Pi) = n(p - \lambda_1 - \lambda_2), \quad (4)$$

gdzie λ_1 i λ_2 są dwiema największymi wartościami własnymi macierzy R.

Płaszczyzna zdefiniowana przez obie pierwsze składowe główne ma tę właściwość, że suma kwadratów odległości obiektów od niej jest mniejsza niż od jakiegokolwiek innej płaszczyzny.

5. Kryteria dobroci podziału

W celu porównania analizowanych pięciu metod grupowania wprowadzono kilka liczbowych kryteriów dobroci. Można je podzielić na trzy kategorie. Wskaźniki pierwszej kategorii charakteryzują liczebność grup w poszczególnych podziałach. Wskaźniki kategorii drugiej charakteryzują jednorodność we wnętrzu grup. Do trzeciej kategorii należą wskaźniki charakteryzujące różnice międzygrupowe.

5.1. Wskaźniki charakteryzujące liczebność grup

Niektóre metody podziału mogą być czułe na obiekty odbiegające od pozostałych i mogą wykazywać większą tendencję wyodrębniania takich obiektów w postaci grup jednoobektowych. Inną właściwością metody podziału jest tendencja do zachowania grupy lub grup bardzo licznych. Trzeci aspekt, jaki można rozważać, to tendencja do tworzenia grup mało zróżnicowanych lub bardzo zróżnicowanych pod względem liczebności.

W niniejszym opracowaniu używa się trzech parametrów wiążących się z liczebnością grup:

- liczba grup jednoobektowych,
- liczba obiektów w grupie najliczniejszej,
- wariancja liczby obiektów w grupach.

Parametry te obliczono dla każdego zbioru 1. Ich wartości podano w w tablicach 8, 9 i 10.

5.2. Wskaźniki charakteryzujące jednorodność grup

Dla określenia jednorodności grup uzyskanych w poszczególnych podziałach obliczono następujące wskaźniki:

1. Suma kwadratów odległości wewnątrz grup

Niech liczba grup m będzie ustalona i niech $G_i^{(m)} = A_{i1}, A_{i2}, \dots, A_{in}$ oznacza dowolną grupę w podziale na m grup. Suma kwadratów odległości wewnątrz tej grupy (SKOWG) D_i^2 jest równa sumie kwadratów odległości między obiektami należącymi do tej grupy

$$D_i^2 = \sum_k \sum_l d^2(A_{ik}, A_{il}) \quad (5)$$

Dla grup jednoobektowych $D_i^2 = 0$.

Charakterystyką jednorodności wszystkich grup podziału jest suma sum kwadratów odległości wewnątrz grup

$$D^2 = \sum_{i=1}^m D_i^2, \quad (6)$$

W myśl tego kryterium z dwóch podziałów na tę samą liczbę grup za lepszy należy uznać ten, dla którego D^2 jest mniejsze. W hierarchiach podziałów D^2 maleje wraz z liczbą grup.

Następne dwa wskaźniki są pochodnymi sumy kwadratów odległości wewnątrz grup.

2. Średni kwadrat odległości wewnątrz grup

Niech jak poprzednio $G_i^{(m)} = \{A_{i1}, A_{i2}, \dots, A_{in}\}$ oznacza dowolną grupę w podziale na m grup.

Średni kwadrat odległości wewnątrz tej grupy \bar{d}_i^2 określony jest jako iloraz sumy kwadratów odległości (SKOWG) D_i^2 i liczby odległości (LODL) w tej grupie

$$\bar{d}_i^2 = D_i^2 / \left(\frac{n_i (n_i - 1)}{2} \right). \quad (7)$$

Dla całego podziału określa się średni kwadrat odległości wewnątrz grup za pomocą wzoru:

$$\bar{d}^2 = D^2 \left(\sum_{i=1}^m \frac{n_i (n_i - 1)}{2} \right) = \sum_{i=1}^m \frac{n_i (n_i - 1)}{2} \bar{d}_i^2, \quad (8)$$

W myśl tego wskaźnika z dwóch podziałów na tę samą liczbę grup za lepszy uznać należy ten, dla którego \bar{d}^2 jest mniejsze.

Średnie kwadraty odległości \bar{d}^2 podane są w tablicach 11 (zbiór 1) i 12 (zbiór 2).

3. Względna suma kwadratów odległości wewnątrz grup

Dla określonej grupy $G_i^{(m)}$, względna suma kwadratów odległości \bar{d}_i^2 określona jest jako suma kwadratów odległości wewnątrz grupy (SKOWG) podzielona przez liczbę obiektów (LOB) w tej grupie, czyli

$$\bar{d}_i^2 = D_i^2 / n_i \quad (9)$$

Dla całego podziału określa się względną sumę kwadratów za pomocą wzoru:

$$\bar{d}^2 = \frac{\sum \bar{d}_i^2}{\sum n_i} \quad (10)$$

W myśl tego wskaźnika z dwóch podziałów na tę samą liczbę grup za lepszy uznać należy ten, dla którego \bar{d}^2 jest mniejsze.

Względne sumy kwadratów odległości \bar{d}^2 podane są w tablicach 13 (zbiór 1) i 14 (zbiór 2).

4. Suma kwadratów odległości od środków ciężkości (centroidów) wewnątrz grup

Niech $C_1 = (\bar{x}_{11}, \bar{x}_{21}, \dots, \bar{x}_{p1})$,

gdzie $\bar{x}_{ki} = \frac{1}{n_i} \sum_{j=1}^p x_{kij}$

będzie środkiem ciężkości (centroidem) i -tej grupy.

Suma kwadratów odległości od środków ciężkości wewnątrz grup S_w^2 określona jest wzorem

$$S_w^2 = \sum_{i=1}^m \sum_{j=1}^{n_i} d^2 (A_{ij}, C_i) \quad (11)$$

gdzie A_{ij} jest j -tym obiektem i -tej grupy.

Przy podziale na tę samą liczbę grup za lepszą uznać należy tę metodę, dla której suma kwadratów odległości wewnątrz grup jest mniejsza.

Sumy kwadratów odległości wewnątrz grup podane są w tablicach 15 (zbiór 1) i 16 (zbiór 2).

5. Suma kwadratów odległości między grupami

Niech C oznacza środek ciężkości (centroid) wszystkich obiektów:

$$C = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$$

Jako miary zróżnicowania grup używa się sumy kwadratów odległości między grupami S_m^2 , określonej wzorem

$$k_3 = S_m^2 = \sum_i n_i d^2 (C_i, C). \quad (12)$$

Przy podziale na tę samą liczbę grup za lepszą uznać należy tę metodę, dla której suma kwadratów odległości między grupami jest większa.

Sumy kwadratów odległości między grupami podano w tablicach 17 (zbiór 1) i 18 (zbiór 2).

Niezależnie od ilości grup suma S_w^2 i S_m^2 jest stała i równa sumie kwadratów odległości wszystkich obiektów od centroidu C , czyli

$$S_w^2 + S_m^2 = \text{const.}$$

6. Iloraz S_m^2 / S_w^2

Jako wskaźnika charakteryzującego jednorodność wewnątrz grup i zróżnicowanie między grupami użyto stosunku sumy kwadratów odległości między grupami S_m^2 do sumy kwadratów odległości od centroidów wewnątrz grup S_w^2 , czyli ilorazu

$$I = S_m^2 / S_w^2 \quad (13)$$

Przy podziale na tę samą liczbę grup za lepszą należy uznać tę metodę, dla której iloraz ten jest większy.

Wartości ilorazów S_m^2 / S_w^2 podano w tablicach 19 i 20.

6. Omówienie wyników

6.1. Analiza odległości

W części tej omówione zostanie wzajemne podobieństwo powiatów na podstawie odległości obliczonych wg wzoru (1).

Zbiór 1.

Obliczone odległości wahają się od 1.28 dla powiatów Słupsk i Lębork do 14.06 dla powiatów Kazimierza Wielka i Ustrzyki.

Poniżej podano pary powiatów, których odległość wynosi co najmniej 11.00.

39 Lesko	7 Malbork	d = 11.81
39 "	13 Chełmno	d = 11.12
39 "	23 Grójec	d = 11.81
39 "	15 Inowrocław	d = 11.48
39 "	30 Kazimierza Wielka	d = 12.33
40 Ustrzyki	7 Malbork	d = 11.56
40 "	13 Chełmno	d = 11.50
40 "	15 Inowrocław	d = 12.02
40 "	16 Aleksandrów	d = 11.75
40 "	17 Radziejów	d = 12.11
40 "	18 Włocławek	d = 11.17
40 "	23 Grójec	d = 12.21
40 "	28 Pińczów	d = 11.84
40 "	29 Busko	d = 11.42
40 "	30 Kazimierza Wielka	d = 14.06

Największe odległości między sąsiadującymi geograficznie ze sobą powiatami okazały się następujące:

23 Grójec	22 Rawa Mazowiecka	d = 7.70
40. Ustrzyki	38 Sanok	d = 7.28
40 Ustrzyki	39 Lesko	d = 6.12
6 Pruszcz Gd.	4 Kartuzy	d = 6.07
31 Dąbrowa T.	30 Kazimierza Wielka	d = 5.64

Jak to omówiono w p. 4 wzajemne położenie powiatów w sensie obliczonej odległości można w przybliżeniu przedstawić za pomocą składowych głównych. Wykres 1 przedstawia rzut zbioru tych powiatów na płaszczyznę wyznaczoną przez obie pierwsze składowe główne.

Średni kwadrat odległości powiatów od tej płaszczyzny równa

się $p - \lambda_1 - \lambda_2 = 21 - 7.38 - 5.24 = 8.38$, co stanowi około 39% zmienności całkowitej. W sytuacji takiej odchylenia punktów (powiatów) od płaszczyzny mogą być znaczne, a ich położenie na płaszczyźnie nie zawsze odpowiada ich położeniu w wyjściowej przestrzeni cech. Mimo to, rysunek dość dobrze oddaje strukturę podobieństwa między powiatami. Jako najbardziej odbiegające od pozostałych okazują się powiaty o numerach 39 i 40, 30 i 7 oraz 1 i 2.

Na rysunku 2 przedstawiono najkrótszy dendryt łączący wszystkie powiaty. Jak wynika z opisu jego konstrukcji przedstawia on strukturę podobieństwa między powiatami na podstawie najkrótszych odległości.

W tabelicy 6 podano optymalną według metody Szczotki kolejność powiatów.

Zbiór 2.

W zbiorze 2 obliczone odległości wahają się od 1.46 dla powiatów: Wrocław i Dzierżoniów do 14.91 dla powiatów Gołdap i Pruszków.

Oba ostatnio wymienione powiaty najbardziej odbiegają od pozostałych. Odległości większych niż 11.0 jest w zbiorze 2. dwadzieścia dwie. W 20 przypadkach spośród nich jednym z powiatów jest powiat pruszkowski.

Powiat pruszkowski odbiega także znacznie od powiatów sąsiadujących z nim. I tak, odległość między powiatem Pruszków i powiatem Grodzisk^{M. BZ.} wynosi 10.63, powiatem Nowy Dwór Maz. 7.81 oraz powiatem Sochaczew 8.65.

Dalsze sąsiadujące ze sobą powiaty w największych odległościach są następujące:

37 Wałbrzych	34 Świdnica	d = 8.29.
3 Suwałki	1 Gołdap	d = 7.89
37 Wałbrzych	35 Dzierżoniów	d = 7.53
4 Augustów	2 Olecko	d = 6.35
31 Milicz	30 Rawicz	d = 6.08

Ryc. 23 przedstawia rzut powiatów zbioru 2 na płaszczyźnie wyznaczonej przez obie pierwsze składowe główne. Średni kwadrat odległości powiatów od tej płaszczyzny wynosi $p - \lambda_1 - \lambda_2 = 21 - 7.64 - 4.58 = 8.78$, co stanowi 40% całej zmienności. Podobnie jak dla zbioru 1. odchylenie to jest znaczne, jednak dla celów praktycznych rysunek może być bardzo przydatną ilustracją wzajemnego położenia powiatów i struktury podobieństwa między nimi. Przede wszystkim rzucają się w oczy jako skrajne powiaty gołdapski (1) i olecki (2) oraz pruszkowski (18).

Na rysunku 4 przedstawiono najkrótszy dendryt dla zbioru 2.

Kolejność powiatów, uporządkowanych metodą Szczołki podana jest w tablicy 7. Kolejność ta w dużym stopniu zgodna jest z kolejnością w położeniu geograficznym.

6.2. Analiza hierarchii podziałów

Szczegółowa analiza merytoryczna poszczególnych podziałów z punktu widzenia geografii rolnictwa powinna być przeprowadzona przez specjalistów z tej dziedziny. Może ona prowadzić do interesujących wniosków fachowych. Podstawą do takiej analizy są szczegółowe wykazy grup podane w tabulogramach 9 - 10 dla zbioru 1. i 19 - 20 dla zbioru 2. W niniejszym opracowaniu omówione zostaną niektóre tylko aspekty i to raczej w ujęciu formalnym.

Zbiór 1.

Metoda par grup. Metoda ta już przy podziale na dwie grupy rozbiła cały zbiór na grupę liczącą 25 powiatów i grupę liczącą 15 powiatów. Do grupy drugiej należą powiaty południowe od powiatu opoczyńskiego poczynając. Wyjątkami są powiaty jędrzejowski (27) i kazimierzowski (30), które postępowanie zaliczyło do grupy pierwszej powiatów północnych.

Zwiększanie liczby grup powoduje odszczepianie pojedynczych powiatów: ustrzyckiego (40), grójeckiego (23) i kazimierskiego (30) oraz pary powiatów: sanockiego (38) i leskiego (39).

Dalsze rozdrobnienie podziału rozbiło grupę powiatów północnych, łącząc w odrębną grupę powiaty: pruszczański (6), malborski (7), tczewski (9), chełmiński (13), toruński (14) i inowrocławski (15). W następnym kroku następuje rozdzielanie powiatów południowych przez oddzielenie grupy powiatów: nowosądeckiego (34), nowotarskiego (35), jasielskiego (36) i krośnieńskiego (37). Dalej następuje utworzenie nowej grupy, obejmującej cztery powiaty: aleksandrowski (16), radziejowski (17), włocławski (18) i kutnowski (19). Przy dalszym podziale wyodrębnia się powiat malborski (7) jako nowa grupa; potem rozdziela się duża grupa powiatów północnych na dwie. Jedna z nich obejmuje powiaty: słupski (1), lęborski (2), bytowski (3), starogardzki (8), tucholski (10), świeciewski (11) i bydgoski (12); druga powiaty: kartuski (4), kościerzyński (5), łowicki (20), skierniewicki (21), rawsko-mazowiecki (22) i jędrzejowski (27). Dalej z grupy powiatów południowych wydziela się grupa powiatów: opoczyńskiego (24), koneckiego (25) i kieleckiego (26).

Na tym etapie liczba grup wynosi 12, przy czym 4 grupy są jednopowiatowe (powiaty 7, 23, 30 i 40), a jedna grupa obejmuje dwa powiaty (38 i 39). Skład pozostałych grup jest następujący:

(1, 2, 3, 8, 10, 11, 12)

(4, 5, 20, 21, 22, 27)

(6, 9, 13, 14, 15)

(16, 17, 18, 19)

(24, 25, 26)

(28, 29, 31, 32, 33)

(34, 35, 36, 37).

Grupy 1 i 2 są dość stabilne, bo dopiero przy podziale na 17 grup oddzielają się powiaty: 4 i 5, a przy podziale na 18 grup powiaty: 1, 2 i 3.

Metoda podziałowa A. Metoda ta także dzieli cały zbiór na dwie duże grupy. Podział ten jest jednak inny niż przy metodzie par grup. Do jednej grupy zaliczone zostały trzy powiaty północne (słupski, lęborski i bytowski), powiaty południowe od powiatu nowosądeckiego (powiaty 34, 35, 36, 37, 38, 39 i 40) oraz powiaty opoczyński (24), kielecki (26) i tarnowski (32). W następnym kroku z grupy tej oddzielają się powiaty leski (38) i ustrzycki (40) jako nowa grupa, a dalej kolejnemu rozdzieleniu ulega duża grupa pozostałych powiatów.

Przy 12 grupach podział przedstawia się następująco. Są grupy po jednym powiecie (pow. grójecki i pow. kazimierski) oraz trzy grupy po dwa powiaty: opoczyński i sanocki, nowosądecki i nowotarski oraz leski i ustrzycki. Następne grupy składają się z powiatów o następujących numerach:

(6, 7, 15)

(9, 10, 11, 12)

(4, 5, 8, 21, 22, 25)

(20, 27, 28, 29, 31, 33)

(13, 14, 16, 17, 18, 19)

(1, 2, 3)

(26, 32, 36, 37)

Podział ten odbiega znacznie od podziału na tę samą liczbę grup, jaki dała metoda par grup. Wypowiedź, który z podziałów jest sensowniejszy z punktu widzenia merytorycznego, należy do specjalisty z zakresu geografii rolnictwa. Z punktu widzenia formalnego powyższy podział daje grupy bardziej jednorodne niż metoda par grup.

Metoda najdalszego sąsiada. Metoda ta przy podziale na 2 grupy oddziela grupę powiatów "bieszczadzkich": sanocki, leski i ustrzycki.

Przy zwiększającej się liczbie grup wydzielają się z pozostałych powiatów nowe grupy i to w kolejności:

- grupa obejmująca powiaty: pruszczański (6), malborski (7), trzewski (9), chełmiński (13), toruński (14), inowrocławski (15), grójecki (23), kazimierski (30). Z grupy tej od razu oddzielają się powiaty: grójecki (23) i kazimierski (30) jako oddzielne grupy;
- grupa obejmująca następujące powiaty południowe: opoczyński (24), konecki (25), kielecki (26), pińczowski (28), buski (29), tarnowsko-dąbrowski (31), tarnowski (32), dębicki (33), nowosądecki (34), nowotarski (35), jasielski (36) i krośnieński (38). Ostatnie cztery powiaty (34, 35, 36 i 38) oddzielają się zaraz w następnym kroku tworząc

nowe grupy,

- grupa obejmująca powiaty: słupski (1), lęborski (2), bytowski (3), starogardzki (8), tucholski (10), świeciewski (11) i bydgoski (12),
- grupa obejmująca powiaty: kartuski (4), kościerzynski (5), łowicki (20), skierniewicki (21), mazowiecko-rawski (22) i jędrzejowski (27).

Dalsze zwiększenie grup odbywa się przez wyodrębnienie się powiatów ustrzyckiego (40) i malborskiego (7), jako pojedynczych grup. Podział na 12 grup okazuje się taki sam jak w metodzie par grup.

Metoda podziałowa B. Metoda podziałowa B oddzieliła początkowo 5 grup jednopowiatowych i grupę złożoną z dwóch powiatów (38, 39). Dopiero potem z dużej grupy pozostałych powiatów wyodrębniła się grupa złożona z powiatów: tarnowsko-dąbrowskiego (31), tarnowskiego (32), dębickiego (33), nowosądeckiego (34), jasielskiego (36) i krośnieńskiego (37), a następnie grupa złożona z dwóch powiatów: koneckiego (25) i kieleckiego (26).

Przy podziale na 12 grup mamy 6 grup złożonych z pojedynczych powiatów, trzy grupy złożone z par powiatów (powiaty konecki i kielecki, pińczowski i buski oraz sanocki i leski) oraz trzy grupy duże obejmujące powiaty:

(6, 9, 13, 14, 15, 16, 17, 18, 19)

(1, 2, 3, 4, 5, 8, 10, 11, 12, 20, 21, 22, 27)

(31, 32, 33, 34, 36, 37).

Tendencja tworzenia grup jedno lub dwuobiektywnych utrzymuje się przy zwiększaniu liczby grup.

Metoda najbliższego sąsiada. Metoda ta w jeszcze większym stopniu niż metoda podziałowa B wykazuje tendencję do tworzenia grup jednoobektowych. Jeszcze przy podziale na 10 grup daje ona 8 grup jednoobektowych utworzonych z powiatów pruszkowskiego (6), malborskiego (7), grójeckiego (23), kazimierskiego (30), nowotarskiego (35), sanockiego (38), leskiego (39) i ustrzyckiego (40) i jedną grupę złożoną z powiatów pińczowskiego (28) i buskiego (29). Na wykresie składowych głównych (rys. 1) można sprawdzić, że są to powiaty o położeniu skrajnym. Pozostała grupa obejmuje 30 powiatów. Dopiero w następnym etapie grupa ta rozdziela się na dwie, z których jedna zawiera powiaty dąbrowski (31), tarnowski (32), dębicki (33), nowosądecki (34), jasielski (36) i krośnieński (37), druga grupa pozostałe powiaty.

Zbiór 2.

Metoda par grup. Podobnie jak poprzednio, zacznę od omówienia podziałów uzyskanych metodą par grup. Przy podziale na dwie i trzy grupy oddziela się najpierw jako osobna grupa powiat pruszkowski, a potem grupa złożona z powiatów gołdapskiego (1) i oleckiego (2). Takie same podziały dają także metoda najbliższego sąsiada i metoda podziałowa B. Dopiero w następnych krokach grupa 37 pozostałych powiatów zostaje kolejno rozdzielona. I tak przy podziale na 4 grupy wyodrębnia się z niej grupa obejmująca powiaty krotoszyński (28), gostyński (29), rawicki (30), wrocławski (33), świdnicki (34), dzierzoniowski (35) i strzeliński (36). Przy podziale na 5 grup - grupa obejmująca powiaty północno-wschodnie: Suwałki (3), Augustów (4), Dąbrowa-Białostocka (5), Sokółka (6), Mońki (7), Białystok (8), Łapy (9), Wysokie Mazowieckie (10).

Przy podziale na 6 grup - grupa obejmująca powiaty Ostrów Wlkp. (27), Milicz (31), Wałbrzych (37), Kamienna Góra (38), Jelenia Góra (39) i Lwówek Śląski (40). Przy podziale na 6 grup mamy więc grupę liczącą 16 powiatów; trzy grupy liczące 8, 7 i 6 powiatów; jedną grupę liczącą dwa powiaty i jedną grupę jednopowiatową.

Przy dalszym zwiększaniu liczby grup, grupa obejmująca powiaty o numerach: 28, 29, 30, 33, 34, 35 i 36 rozdziela się na dwie. Jedna z nich zawiera powiaty: krotoszyński (28), gostyński (29) i rawicki (30); druga powiaty wrocławski (33), świdnicki (34), dzierzoniowski (35) i strzeliński (36).

W następnym kroku (8 grup) duża grupa licząca 16 powiatów rozdziela się na dwie: grupę obejmującą powiaty: Ostrów Maz. (11), Sokołów Podlaski (12), Węgrów (13), Wołomin (14), Wyszaków (15), Pułtusk (16), Grodzisk Maz. (19), Łowicz (21), Podębnice (23) i Turek (24) oraz grupę obejmującą powiaty: Nowy Dwór Maz. (17), Sochaczew (20), Łęczyca (22), Kalisz (25), Pleszew (26) i Trzebnica (32). Pierwsza z tych grup jest bardzo stabilna i rozdziela się dopiero na dalszym etapie przy podziale na 15 grup.

Metoda najdalszego sąsiada. Podział na grupy uzyskany tą metodą jest podobny do podziału uzyskanego metodą par grup. Porównamy podziały uzyskane obiema metodami przy liczbie 8 grup. I w jednej i drugiej metodzie mamy powiat pruszkowski (8) jako odrębną grupę oraz grupę złożoną z powiatów gołdapskiego (1) i oleckiego (2). Następne dwie grupy różnią się parą powiatów: Łapy (9) i Wysokie Maz. (10). Metoda par grup zaliczyła je do jednej grupy z powiatami o numerach 3, 4, 5,

6, 7 i 8, natomiast metoda najdalszego sąsiada zaliczyła je do jednej grupy z powiatami 11, 12, 13, 14, 15, 16, 19, 21, 23, 24. Występujące w metodzie najdalszego sąsiada w jednej grupie powiaty: Krotoszyn (28), Gostyń (29), Rawicz (30), Wrocław (33), Świdnica (34), Dzierżoniów (35) i Strzelin (36) w metodzie par grup są rozdzielone na grupę powiatów 28, 29 i 30 oraz grupę powiatów 33, 34, 35 i 36. Zarówno w jednej jak i w drugiej metodzie powiaty 17, 20 i 32 oraz 31, 37, 38, 39 i 40 występują w tych samych grupach. Występująca jednak w metodzie najdalszego sąsiada grupa obejmująca powiaty: Łęczycą (22), Kalisz (25), Pleszew (26) i Ostrów Wlkp. (27) jest w metodzie par grup rozdzielona. Powiaty 22, 25 i 26 zaliczone są do jednej grupy z powiatami 17, 20 i 32, natomiast powiat 27 do grupy powiatów 31, 37, 38, 39 i 40. Zaznaczyć trzeba, że podział metodą par grup jest bardziej jednorodny niż podział metodą najdalszego sąsiada.

Metoda podziałowa B. Metoda ta daje także podziały zbliżone do metody par grup. Tak jak w tamtej metodzie najpierw oddziela się powiat pruszkowski (18) jako odrębna grupa, potem grupa złożona z powiatów: gołdapskiego i oleckiego. Podział na cztery grupy jest nieco inny, bo powstaje nowa grupa licząca tylko cztery powiaty: wrocławski (33), świdnicki (34), dzierżoniowski (35) i strzeliński (36). W podziale na pięć grup dochodzi grupa złożona z powiatów: Krotoszyn (28), Gostyń (29) i Rawicz (30), a następnie grupa obejmująca powiaty: Suwałki (3), Augustów (4), Dąbrowa Białostocka (5), Sokółka (6), Mońki (7), Białystok (8), Łapy (9) i Wysokie Maz.(10).

Duża jest zgodność podziałów na 7 grup uzyskanych obiema metodami. Pięć grup jest takich samych, a dwie grupy róż-

nią się występowaniem powiatu ostrowskiego (27).

Metoda podziałowa A. Metoda ta działa inaczej niż omówiona poprzednio. Już przy podziale na 2 i 3 grupy dzieli ona całość na dwie grupy po 28 i 17 powiatów oraz na 3 grupy po 8, 20 i 12 powiatów.

Przy podziale na 4 grupy wyodrębnia się grupa złożona z dwóch powiatów: nowodworskiego (17) i pruszkowskiego (18). Jest to dziwne, gdyż oba te powiaty są dość odległe od siebie ($d = 7.81$).

Podział na 8 grup jest następujący: Występują dwie grupy jednoobiektywne (pow. nowodworski i pow. pruszkowski) oraz grupa łącząca powiaty gołdapski (1) i wałbrzyski (37). Dalsze grupy obejmują powiaty o numerach: -

(2, 3, 6, 7, 8, 38)

(4, 31, 32, 40)

(5, 9, 10, 11, 12, 13, 15, 16, 23, 24)

(14, 19, 21, 27, 30, 39)

(20, 22, 25, 26, 28, 29, 33, 34, 35, 36).

Metoda łączy powiaty geograficznie bardziej odległe niż metody poprzednie. Suma kwadratów odległości wewnątrz grup wynosi 2613 i jest prawie dwa razy większa niż dla metody par grup ($D^2 = 1411$).

Metoda najbliższego sąsiada. Podobnie jak dla zbioru 1. Metoda ta charakteryzuje się tendencją odszczepiania małych grup złożonych z obiektów skrajnych. Tak więc najpierw wyodrębnia jako grupę powiat pruszkowski (18), potem grupę powiatów gołdapski (1) i olecki (2), następnie powiat augustowski

(4). Przy dalszym zwiększaniu grup powstają, podobnie jak w metodzie najdalszego sąsiada, grupy obejmujące powiaty 28, 29, 30 oraz 33, 34, 35 i 36. Dalej pozostaje grupa z powiatu Dąbrowa Białostocka (5), następnie rozdziela się grupa złożona z powiatów 1 i 2 na dwie grupy jednoobiektywne. Duża grupa licząca 27 powiatów zostaje rozdzielona dopiero przy podziale na 10 grup; na grupę liczącą 6 i grupę liczącą 21 powiatów. Z grupy tej wyodrębnia się w następnym kroku grupa licząca 5 powiatów. Pozostała grupa 16 powiatów jest dość stabilna.

7. Analiza porównawcza jakości metod .

W niniejszej części omówione zostaną formalne kryteria dobroci, opisane w p. 5, dla wszystkich pięciu metod.

7.1. Wskaźniki charakteryzujące liczebność grup

Liczba grup jednoobiektywych. Liczba grup jednoobiektywych dla rozpatrywanych pięciu metod (dla zbioru 1) podana jest w tablicy 8. W tablicy 21 podano w ilu podziałach metoda wymieniona w wierszu dała mniej grup jednoelementowych niż metoda wymieniona w nagłówku kolumny.

Najsilniejszą tendencję do tworzenia grup jednoelementowych wykazuje metoda najbliższego sąsiada (SIM), która zawsze dała, co najmniej, tyle grup jednoobiektywych co inne metody. Najmniej grup jednoelementowych tworzą metoda podziałowa A (MPA) i metoda najdalszego sąsiada (CIM).

Liczba obiektów w grupie najliczniejszej. Jest to druga charakterystyka liczebności w grupach. Zachowanie grup bardzo licznych uznać należy za niepożądane.

Liczbę obiektów w grupach najliczniejszych (dla zbioru 1) podano w tabl. 9. Tablica 22 zawiera liczbę podziałów, w których metoda wymieniona w tytule wiersza miała najliczniejszą grupę mniejszą, niż metoda wymieniona w tytule kolumny.

Okazało się, że metoda najbliższego sąsiada we wszystkich przypadkach dała największe grupy. Najlepsza pod tym względem okazała się metoda podziałowa A, dla której w większości podziałów liczebność maksymalnej grupy była mniejsza niż w innych metodach.

Wariancja liczby obiektów w grupach. Wariancja liczby obiektów w grupach podana jest w tabl. 10. Zbiorcze zestawienie podano w tabl. 23. Największe zróżnicowanie liczby obiektów w grupach wykazuje metoda najbliższego sąsiada (SLM). Metoda ta dała tylko jeden podział, w którym wariancja liczby obiektów w grupach była mniejsza niż przy podziale na tę samą liczbę grup uzyskanych inną metodą (metodą podziałową B).

Najmniejszym zróżnicowaniem liczby obiektów charakteryzuje się metoda podziałowa A (MPA) i metoda najdalszego sąsiada (CIM).

Średni kwadrat odległości wewnątrz grup d^2 . Wielkość ta obliczona została wg wzoru (8) z p. 5. Średnie kwadraty odległości wewnątrz grup podano w tablicy 11 dla zbioru 1 i tablicy 12 dla zbioru 2. Tablice 24 i 25 zawierają zestawienia zbiorcze liczby podziałów, w których metoda wymieniona w tytule wiersza dała grupy bardziej jednorodnie niż metoda wymieniona w tytule kolumny.

Metodami dającymi podziały o największym średnim kwadracie

odległości, a więc dające podziały na grupy najmniej jednorodne w tym sensie, okazały się metoda najbliższego sąsiada (SIM) i metoda podziałowa A (MPA). W zbiorze 1. gorsza okazała się metoda najbliższego sąsiada, w zbiorze 2. metoda podziałowa A. Następną co do dobroci była metoda podziałowa B (MPB), potem metoda par grup (MPG). Najlepszą okazała się metoda najdalszego sąsiada (CIM).

Analizując dane z tablic 11 i 12 stwierdza się, że w porównaniu z metodą najbliższego sąsiada metoda podziałowa A daje bardziej jednorodne podziały przy małej liczbie grup, a mniej jednorodne przy dużej. Podobną tendencję wykazuje metoda podziałowa B w porównaniu z metodą podziałową A.

Względna suma kwadratów odległości \bar{d}^2 . Wielkość ta obliczona została wg wzoru (10) z p. 5. Zestawienie względnych sum kwadratów dla hierarchii podziałów uzyskanych różnymi metodami podano w tabl. 13 dla zbioru 1. i tablicy 14 dla zbioru 2. W tablicach 26 i 27 podano ilość podziałów, dla których metoda wymieniona w kolumnie tytułowej dała mniejszą względną sumę kwadratów niż metoda wymieniona w wierszu tytułowym.

Podobnie jak poprzednio, najgorszą w omawianym sensie okazała się metoda najbliższego sąsiada (SIM). Lepsze od niej, ale gorsze od metody par grup i metody najdalszego sąsiada okazały się metody podziałowe A (MPA) i B (MPB), przy czym w zbiorze 1. lepszą okazała się metoda A, w zbiorze 2. metoda B. Najmniejsze sumy kwadratów odległości wewnątrz grup dały metoda par grup (MPG) i metoda najdalszego sąsiada (CIM).

Suma kwadratów odległości od centroidów wewnątrz grup S_w^2 .

Sumy kwadratów odległości od środków ciężkości wewnątrz grup obliczone zostały wg wzoru (11) z p. 5. Są one podane dla poszczególnych hierarchii podziałów w tablicach 15 i 16. Ponieważ suma kwadratów odległości od centroidów wewnątrz grup S_w^2 i suma kwadratów odległości między grupami S_m^2 (12) są stałe, analiza porównawcza nastąpi na podstawie tej drugiej.

Suma kwadratów odległości między grupami S_m^2 . Suma kwadratów odległości między grupami rozumiana jest jako ważona suma kwadratów odległości centroidów grup od środka ciężkości (centroidu) całego materiału i rachowana według wzoru (12). Wartości S_m^2 dla zbioru 1 podane są w tablicy 17, a dla zbioru 2 w tablicy 18. W tablicach 28 i 29 podano zestawienia zbiorcze ilości podziałów, w których metoda wymieniona w tytule wiersza (M_w) dała większe zróżnicowanie grup niż metoda wymieniona w tytule kolumny (M_k). Z zestawienia tego wynika, że podziały o największych odległościach między grupami uzyskuje się metodą najdalszego sąsiada (CJM). Drugą po niej jest metoda par grup (MPG). Różnice we wskaźniku S_m^2 między obiema tymi metodami były niewielkie. Na trzecim miejscu jest metoda podziałowa B (MPB). Różnica w S_m^2 między tą metodą a pierwszymi dwiema jest duża dla podziałów z małą liczbą grup. Najmniejsze zróżnicowanie między grupami daje metoda podziałowa A (MPA) i metoda najbliższego sąsiada (SIM), które dają podziały o zbliżonych wartościach S_m^2 .

Iloraz I = S_m^2 / S_w^2 . Wartości ilorazu sumy kwadratów odległości między grupami i sumy kwadratów odległości od cen-

troidów wewnątrz grup podano w tablicach 19 i 20. Dla wszystkich pięciu metod iloraz ten jest funkcją rosnącą liczby grup. Zbiorcze zestawienia ilości podziałów, w których jedna metoda (M_w) okazała się lepsza niż druga (M_k) podane są w tablicach 30 i 31.

Podobnie jak poprzednio, tak i ze względu na ten wskaźnik najlepszymi okazały się metoda najdalejzego sąsiada (CIM) i metoda par grup (MPG), z tym, że różnice między nimi są niewielkie. Środkowe miejsce zajmuje metoda podziałowa B. Najwięcej podziałów o najmniejszej wartości ilorazu I dały metody najbliższego sąsiada (SIM) i metoda podziałowa A (MPA).

8. Wnioski

1. Analizowane w niniejszym opracowaniu metody dają w praktyce odbiegające od siebie hierarchie podziałów. Z punktu widzenia rozpatrywanych kryteriów dobroci: liczby obiektów w grupach, jednorodności wewnątrz grupy i zróżnicowania między grupami, metody te ułożyły się prawie w tej samej kolejności. Najlepszymi okazały się metoda najdalejzego sąsiada i metoda grupowanych par. Najgorszymi natomiast okazały się metoda najbliższego sąsiada i metoda podziałowa A. Metoda podziałowa B zajmuje miejsce pośrednie. Ponieważ czasy egzekucji obliczeń różnią się nieznacznie, przeto jako najbardziej przydatne spośród badanych metod uznać należy metodę par grup i metodę najdalejzego sąsiada.

2. Jako postępowanie pomocnicze, pozwalające zilustrować wzajemne położenie obiektu w przestrzeni cech, zalecić należy graficzne przedstawienie danych w układzie dwóch pierwszych składowych głównych.

3. Bardzo przydatne w nieformalnej analizie skupień jest korzystanie z tablicy odległości odpowiadającej optymalnemu uporządkowaniu liniowemu obiektów (odpowiednik diagramu Czekanowskiego).

4. Sądzić należy, że analizowane metody, jak i inne podobne metody są zawodne w przypadku jednorodnych zbiorów obiektów.

5. Niepraktyczne i pracochłonne mogą okazać się metody grupowania podobne do omawianych w przypadku zbiorów bardzo licznych. Dlatego też należy szierować uwagę na procedury pozwalające analizować bardzo duże zbiory oraz na metody pozwalające przenieść w sposób sensowny podziały losowo wybranych podzbiorów na całość.

Literatura

- Florek K. Łukaszewicz J., Perkal J., Steinhaus, H. et
Zubrzycki S., (1951) Sur la liaison et la division des
points d'un ensemble fini, Coll. Math., 2, p. 282-285.
- Hubert L., (1973), Monotone invariant clustering procedures,
Psychometrika, 38, p. 47-62.
- Johnson S. C., (1967), Hierarchical clustering schemes,
Psychometrika, 32, p. 241-254.
- Sokal R.R., Sneath P.H., (1963), Principles of Numerical
Taxonomy, Freeman and Co., San Francisco.
- Szczotka F.A., (1972), On a method of ordering and cluster-
ing of objects, Zast. Mat., XIII, str. 23-35.
- Szczotka F.A., (1976), Podstawy taksonomii numerycznej,
Biuletyn Informacyjny ZKPW "Podstawy przestrzennego zagospo-
odarowania kraju", z. 19, IGI PZ PAN.

SPIS TABLIC

	strona
Tab. 1. Zbiór 1.40 powiatów pasa biegnącego z północy na południowy wschód	63
Tab. 2. Zbiór 2.40 powiatów pasa biegnącego z północnego wschodu na południowy zachód	64
Tab. 3. Zespół cech dla zbiorów 1 i 2	65
Tab. 4. Średnie arytmetyczne i standardowe odchylenia cech dla zbioru 1	66
Tab. 5. Średnie arytmetyczne i standardowe odchylenia cech dla zbioru 2	67
Tab. 6. Kolejność powiatów zbioru 1 uporządkowanych metodą Szczotki	68
Tab. 7. Kolejność powiatów zbioru 2 uporządkowanych metodą Szczotki	69
Tab. 8. Liczba grup jednoobektowych. Zbiór 1	70
Tab. 9. Liczba obiektów w grupie najliczniejszej. Zbiór 1	71
Tab. 10. Wariancja liczby obiektów w grupach. Zbiór 1	72
Tab. 11. Średni kwadrat odległości wewnątrz grup \bar{d}^2 . Zbiór 1...	73
Tab. 12. Średni kwadrat odległości wewnątrz grup \bar{d}^2 . Zbiór 2 ..	74
Tab. 13. Względna suma kwadratów odległości wewnątrz grup \bar{d}^2 . Zbiór 1	75
Tab. 14. Względna suma kwadratów odległości wewnątrz grup \bar{d}^2 . Zbiór 2	76

Tab. 15. Suma kwadratów odległości od centroidów wewnątrz grup. Zbiór 1	77
Tab. 16. Suma kwadratów odległości od centroidów wewnątrz grup. Zbiór 2	78
Tab. 17. Suma kwadratów odległości między grupami. Zbiór 1	79
Tab. 18. Suma kwadratów odległości między grupami. Zbiór 2	80
Tab. 19. Iloraz: suma kwadratów odległości między grupami podzielona przez sumę kwadratów odległości wewnątrz grup. Zbiór 1	81
Tab. 20. Iloraz: suma kwadratów odległości między grupami podzielona przez sumę kwadratów odległości wewnątrz grup. Zbiór 2	82
Tab. 21. Liczba podziałów, w których metoda M_w dała mniej grup jednoobektowych niż metoda M_k	83
Tab. 22. Liczba podziałów, w których najliczniejsza grupa w metodzie M_w była mniej liczna niż w metodzie M_k	83
Tab. 23. Liczba podziałów, w których wariancja liczebności grup w metodzie M_w była mniejsza niż w metodzie M_k ..	84
Tab. 24. Liczba podziałów, w których metoda M_k dała mniejszy średni kwadrat odległości \bar{d}^2 niż metoda M_w . Zbiór 1..	84
Tab. 25. Liczba podziałów, w których metoda M_k dała mniejszy średni kwadrat odległości \bar{d}^2 niż metody M_w . Zbiór 2..	85
Tab. 26. Liczba podziałów, w których metoda M_k dała mniejszą sumę kwadratów odległości \tilde{d}^2 niż metoda M_w . Zbiór 1..	85
Tab. 27. Liczba podziałów, w których metoda M_k dała mniejszą sumę kwadratów odległości \bar{d}^2 niż metoda M_w . Zbiór 2..	86
Tab. 28. Liczba podziałów, w których metoda M_w dała większą sumę kwadratów odległości między grupami S_m^2 niż metoda M_k . Zbiór 1	86

- Tab. 29. Liczba podziałów, w których metoda M_w dała większą sumę kwadratów odległości między grupami S_m^2 niż metoda M_k . Zbiór 2 87
- Tab. 30. Liczba podziałów, w których metoda M_w dała większą wartość ilorazu $I = S_m^2 / S_w^2$ niż metoda M_k .
Zbiór 1 87
- Tab. 31. Liczba podziałów, w których metoda M_w dała większą wartość ilorazu $I = S_m^2 / S_w^2$ niż metoda M_k .
Zbiór 2 88

Tab.1. Zbiór 1 : 40 powiatów pasa biegnącego z północy na południowy wschód

- | | |
|---------------------|-----------------------|
| 1. Słupsk | 26. Kielce |
| 2. Łęborg | 27. Jędrzejów |
| 3. Bytów | 28. Pinczów |
| 4. Kartuzy | 29. Busko |
| 5. Kościerzyna | 30. Kazimierza Wielka |
| 6. Pruszcz Gdański | 31. Dąbrowa Tarnowska |
| 7. Malbork | 32. Tarnów |
| 8. Starogard | 33. Dębica |
| 9. Tczew | 34. Nowy Sącz |
| 10. Tuchola | 35. Nowy Targ |
| 11. Świecie | 36. Jasło |
| 12. Bydgoszcz | 37. Krosno |
| 13. Chełmno | 38. Sanok |
| 14. Toruń | 39. Lesko |
| 15. Inowrocław | 40. Ustrzyki |
| 16. Aleksandrów | |
| 17. Radziejów | |
| 18. Włocławek | |
| 19. Kutno | |
| 20. Łowicz | |
| 21. Skierniewice | |
| 22. Rawa Mazowiecka | |
| 23. Grójec | |
| 24. Opoczno | |
| 25. Końskie | |

Tab.2. Zbiór 2: 40 powiatów pasa biegnącego z północnego wschodu na południowy zachód

- | | |
|------------------------|-------------------|
| 1. Gołdap | 26. Pleszew |
| 2. Olecko | 27. Ostrów Wlkp. |
| 3. Suwałki | 28. Krotoszyn |
| 4. Augustów | 29. Gostyń |
| 5. Dąbrowa Białostocka | 30. Rawicz |
| 6. Sokółka | 31. Milicz |
| 7. Monki | 32. Trzebnica |
| 8. Białystok | 33. Wrocław |
| 9. Łapy | 34. Świdnica |
| 10. Wysokie Maz. | 35. Dzierżoniów |
| 11. Ostrów Maz. | 36. Strzelin |
| 12. Sokółów Podlaski | 37. Wałbrzych |
| 13. Węgrów | 38. Kamienna Góra |
| 14. Wołomin | 39. Jelenia Góra |
| 15. Wyszaków | 40. Lwówek Śląski |
| 16. Pułtusk | |
| 17. Nowy Dwór Maz. | |
| 18. Pruszków | |
| 19. Grodzisk Maz. | |
| 20. Sochaczew | |
| 21. Łowicz | |
| 22. Łęczyca | |
| 23. Poddębice | |
| 24. Turek | |
| 25. Kalisz | |

Tab. 3. Zestaw cech diagnostycznych typu rolnictwa

1. Produktywność pracy
2. Nawożenie mineralne
3. Nawożenie organiczne
4. Produktywność ziemi
5. Poziom towarowości
6. Stopień towarowości
7. Struktura własności
8. Średnia wielkość gospodarstw indywidualnych w ha
9. Ludność rolnicza na 100 ha UR
10. Konie na 100 ha UR
11. Traktory na 100 ha UR
12. Procentowy udział produkcji globalnej zwierzęcej w produkcji globalnej ogółem
13. Procentowy udział produkcji towarowej zwierzęcej w produkcji towarowej ogółem
14. Procentowy udział roślin przemysłowych w produkcji globalnej ogółem
15. Procentowy udział w produkcji warzyw i owoców w produkcji globalnej ogółem
16. Procentowy udział produkcji mleka w produkcji globalnej ogółem
17. Procentowy udział produkcji trzody w produkcji globalnej ogółem
18. Procentowy udział produkcji żywca bydłęcego w produkcji globalnej ogółem
19. Procentowy udział roślin ekstraktywnych w powierzchni globalnej zasiewów
20. Procentowy udział roślin intensyfikujących w powierzchni globalnej zasiewów
21. Procentowy udział roślin strukturotwórczych w powierzchni globalnej zasiewów

Tab.4. Średnie arytmetyczne \bar{x} i standardowe odchylenia s /
Zbiór 1

nr cechy	\bar{x}	s
1	46.22	15.72
2	54.98	25.39
3	67.99	14.74
4	27.15	5.87
5	3.87	1.41
6	45.28	10.30
7	20.52	15.60
8	5.44	1.72
9	64.18	23.91
10	13.54	4.08
11	0.72	0.55
12	48.61	6.81
13	67.55	13.42
14	4.97	5.35
15	5.23	4.32
16	18.84	5.42
17	15.33	2.54
18	6.48	1.91
19	54.71	3.21
20	27.20	4.05
21	18.11	3.55

Tab.5. Średnie arytmetyczne \bar{x} i standardowe odchylenia s /
Zbiór 2

nr cechy	\bar{x}	s
1	50.56	10.81
2	62.88	29.47
3	66.85	11.79
4	27.52	5.43
5	3.64	1.81
6	43.96	9.28
7	21.07	12.21
8	5.73	1.56
9	54.98	10.57
10	13.44	2.11
11	0.66	0.39
12	45.64	6.57
13	65.96	13.22
14	5.15	4.16
15	5.04	7.47
16	16.09	3.34
17	16.49	3.60
18	6.23	1.94
19	56.46	5.42
20	27.24	4.67
21	16.30	3.22

Tab. 6. Kolejność powiatów zbioru 1 uporządkowanych metodą
Szczotki

Grójec	/23/	Busko	/29/
Malbork	/ 7/	Opoczno	/24/
Inowrocław	/15/	Dębica	/33/
Chełmno	/13/	Dąbrowa Tarnowska	/31/
Tczew	/ 9/	Tarnów	/32/
Pruszcz Gdański	/ 6/	Końskie	/25/
Toruń	/14/	Kielce	/26/
Włocławek	/18/	Krosno	/37/
Radziejów	/17/	Jasło	/36/
Aleksandrów	/16/	Nowy Sącz	/34/
Kutno	/19/	Sanok	/38/
Bydgoszcz	/12/	Kazimierza Wielka	/30/
Świecie	/11/	Nowy Targ	/35/
Słupsk	/ 1/	Lesko	/39/
Lębork	/ 2/	Ustrzyki	/40/
Tuchola	/10/		
Bytów	/ 3/		
Starogard	/ 8/		
Kościerzyna	/ 5/		
Kartuzy	/ 4/		
Łowicz	/20/		
Rawa Mazowiecka	/22/		
Skierzwice	/21/		
Jędrzejów	/27/		
Pinczów	/28/		

Tab.7 . Kolejność powiatów zbioru 2 uporządkowanych metodą Szczotki

Gołdap	/1/	Sochaczew	/20/
Olecko	/2/	Lwówek Śląski	/40/
Suwałki	/3/	Milicz	/31/
Sokółka	/6/	Kamienna Góra	/38/
Augustów	/4/	Wałbrzych	/37/
Dąbrowa Białostocka	/5/	Nowy Dwór Mazowiecki	/17/
Mońki	/7/	Trzebnica	/32/
Białystok	/8/	Kalisz	/25/
Łapy	/9/	Pleszew	/26/
Wysokie Mazowieckie	/10/	Ostrów Wielkopolski	/27/
Ostrów Mazowiecka	/11/	Łęczyca	/22/
Wyszków	/15/	Jelenia Góra	/39/
Węgrów	/13/	Krotoszyn	/28/
Wołomin	/14/	Wrocław	/33/
Pułtusk	/16/	Rawicz	/30/
Sokołów Podlaski	/12/	Dzierżoniów	/35/
Turek	/24/	Gostynin	/25/
Grodzisk Mazowiecki	/19/	Strzelin	/36/
Podębnice	/23/	Świdnica	/34/
Łowicz	/21/	Pruszków	/18/

Tab. 8. Liczba grup jednoobiektowych. Zbiór 1

Łączna liczba grup	SIM	CLM	MPG	MPA	MPB
2	1	0	0	0	1
3	2	0	1	0	1
4	2	1	2	0	2
5	3	2	3	1	3
6	5	2	3	1	4
7	6	2	3	1	5
8	8	3	3	1	5
9	8	3	3	1	5
10	8	3	4	2	6
11	8	3	4	2	6
12	8	4	4	2	6
13	9	5	5	4	6
14	9	5	7	6	7
15	10	5	7	7	8
16	11	7	8	7	9
17	11	7	8	7	9
18	12	8	8	8	10
19	13	9	9	9	11
20	14	10	10	10	11
21	14	10	11	12	12
22	16	11	12	12	14
23	16	12	13	14	15
24	17	13	13	14	15
25	17	14	14	16	16
26	18	15	16	17	17
27	19	17	17	18	18
28	20	18	19	19	19
29	21	20	20	20	20
30	22	22	22	21	22
31	24	23	24	22	23
32	26	25	25	24	25
33	27	26	27	26	26
34	29	28	28	28	28
35	30	30	30	30	30
36	32	32	32	32	32
37	34	34	34	34	34
38	36	36	36	36	36
39	38	38	38	38	38

Tab. 9. Liczba obiektów w grupie najliczniejszej. Zbiór 1.
Łączna

liczba grup	SLM	CLM	MPG	MPA	MPB
2	39	37	25	27	39
3	38	29	24	27	37
4	36	29	24	19	36
5	35	29	23	19	35
6	35	17	23	12	34
7	34	17	17	12	33
8	33	17	17	12	27
9	32	10	13	6	25
10	30	8	13	6	24
11	24	7	7	6	15
12	22	7	7	6	13
13	21	7	7	6	10
14	18	7	7	6	9
15	18	7	7	6	9
16	18	7	7	6	9
17	10	4	7	5	8
18	9	4	4	5	7
19	9	4	4	5	6
20	9	4	4	5	5
21	7	4	4	5	5
22	7	4	4	4	5
23	6	4	4	3	5
24	6	4	4	3	4
25	4	3	3	3	4
26	4	3	3	3	3
27	4	3	3	3	3
28	4	3	3	3	3
29	4	3	3	3	3
30	3	3	3	3	3
31	3	3	3	2	3
32	3	3	3	2	3
33	3	2	3	2	2
34	3	2	2	2	2
35	2	2	2	2	2
36	2	2	2	2	2
37	2	2	2	2	2
38	2	2	2	2	2
39	2	2	2	2	2

Tab. 10. Wariacja liczby obiektów w grupach. Zbiór 1

Liczba grup	SLM	CLM	MFG	MPA	MFB
2	361.0	289.000	25.000	49.000	361.000
3	281.3	126.889	89.556	106.889	280.222
4	250.4	125.000	93.500	37.500	225.500
5	147.1	113.600	81.600	43.200	182.400
6	160.5	35.556	68.889	16.889	149.556
7	133.35	26.776	35.347	11.347	124.204
8	112.00	26.500	26.500	10.500	71.750
9	94.91	10.469	14.469	3.802	55.136
10	75.20	6.400	13.800	4.000	46.600
11	43.50	4.413	6.231	3.868	19.140
12	33.55	4.222	4.222	3.222	14.222
13	28.53	4.225	4.225	3.609	10.076
14	19.41	3.694	4.408	3.837	7.980
15	17.95	2.889	3.822	3.422	7.289
16	17.00	3.000	3.500	2.500	6.250
17	6.93	1.758	2.699	1.640	3.875
18	5.84	1.506	1.506	1.617	3.173
19	5.04	1.357	1.463	1.568	2.620
20	4.40	1.300	1.300	1.400	1.900
21	3.03	1.039	1.229	1.420	1.610
22	2.96	0.876	1.058	0.967	1.603
23	1.032	0.802	0.975	0.802	1.474
24	1.639	0.722	0.722	0.722	0.972
25	1.040	0.560	0.560	0.720	0.800
26	0.864	0.479	0.556	0.633	0.633
27	0.694	0.472	0.472	0.546	0.546
28	0.602	0.388	0.459	0.459	0.459
29	0.511	0.373	0.373	0.373	0.373
30	0.356	0.356	0.355	0.289	0.356
31	0.335	0.271	0.335	0.206	0.271
32	0.312	0.250	0.250	0.188	0.250
33	0.228	0.167	0.227	0.167	0.167
34	0.204	0.145	0.145	0.145	0.145
35	0.122	0.122	0.122	0.122	0.122
36	0.099	0.099	0.099	0.099	0.099
37	0.075	0.075	0.075	0.075	0.075
38	0.050	0.050	0.050	0.050	0.050
39	0.025	0.025	0.025	0.025	0.025

Tab.11. Średni kwadrat odległości wewnątrz grup d^2 /SKOWG/LODL/.Zbiór

Liczba grup	SLM	GLM	MPG	MPA	MPB
2	39.81	37.42	32.68	35.09	39.81
3	37.96	29.41	30.68	33.32	37.40
4	35.44	28.66	28.14	28.19	36.19
5	34.06	28.21	25.86	27.09	34.11
6	34.09	19.69	24.24	23.57	31.97
7	32.79	18.54	19.46	19.71	30.43
8	30.45	18.25	18.25	18.92	26.25
9	30.11	15.38	15.46	16.71	23.96
10	30.35	13.10	15.21	15.18	23.18
11	25.21	11.55	12.59	13.96	17.34
12	22.33	10.83	10.83	13.11	14.85
13	21.26	10.43	10.43	12.74	13.02
14	19.19	9.87	10.31	12.49	12.35
15	19.23	9.26	9.74	11.31	12.03
16	19.34	9.10	9.42	10.25	11.34
17	12.20	7.94	8.68	10.13	10.18
18	11.29	7.27	7.27	9.75	9.50
19	10.76	7.01	7.04	9.51	8.74
20	10.46	6.74	6.74	9.15	7.97
21	10.06	6.49	6.56	8.98	7.48
22	10.12	6.14	6.24	8.30	7.16
23	8.44	5.85	6.01	7.83	6.85
24	7.98	5.55	5.55	7.37	6.32
25	6.52	5.12	5.12	7.08	6.04
26	5.99	4.84	4.99	6.82	5.58
27	5.68	4.68	4.68	6.58	5.11
28	5.43	4.39	4.53	6.33	4.69
29	4.98	4.19	4.19	5.90	4.22
30	3.99	3.99	3.99	5.62	3.99
31	3.87	3.76	3.87	5.19	3.76
32	3.78	3.60	3.60	4.91	3.60
33	3.54	3.27	3.44	4.57	3.27
34	3.40	3.01	3.01	4.17	3.01
35	2.72	2.72	3.78	3.78	2.72
36	2.41	2.41	2.41	3.39	2.41
37	2.17	2.17	2.17	2.91	2.17
38	1.93	1.93	1.93	2.38	1.93
39	1.63	1.63	1.63	1.63	1.63

Tab.12. Średni kwadrat odległości wewnątrz grup \bar{d}^2 /SKOWG/LOD

liczba grup	Zbiór 2					
2	38.89	38.89	38.89	34.340	38.89	
3	35.26	34.30	35.26	29.075	35.26	
4	34.82	23.89	26.40	26.887	31.10	
5	31.83	23.26	22.02	24.097	26.35	
6	25.48	17.73	16.59	22.968	21.75	
7	25.48	14.34	15.79	22.740	16.41	
8	24.99	13.83	12.49	20.578	15.62	
9	25.02	12.27	11.88	19.610	14.74	
10	21.69	10.89	11.30	16.875	13.12	
11	15.304	10.12	10.46	16.505	11.72	
12	15.387	8.94	10.32	15.862	10.89	
13	15.397	8.63	9.64	15.290	10.23	
14	14.317	8.40	9.52	14.659	9.76	
15	14.133	8.23	7.76	14.379	9.21	
16	11.671	7.54	7.54	12.904	8.85	
17	11.642	7.01	7.30	10.457	8.53	
18	11.694	6.69	7.21	9.499	7.94	
19	8.989	6.54	6.30	9.173	6.88	
20	8.519	6.40	6.00	8.957	6.38	
21	7.633	5.87	5.77	8.595	6.19	
22	7.609	5.58	5.67	7.572	5.83	
23	7.497	5.44	5.46	7.306	5.63	
24	5.800	5.31	5.24	7.034	5.41	
25	5.329	5.02	5.12	6.397	5.29	
26	5.051	4.67	4.95	6.044	5.13	
27	4.991	4.46	4.86	5.809	4.77	
28	4.935	4.25	4.41	5.352	4.34	
29	4.883	4.08	4.08	5.160	4.08	
30	4.410	3.91	3.91	4.547	3.91	
31	4.020	3.72	3.72	4.280	3.73	
32	3.937	3.52	3.52	3.920	3.52	
33	3.374	3.37	3.37	3.642	3.37	
34	3.264	3.26	3.26	3.390	3.26	
35	3.187	2.97	3.19	3.187	2.97	
36	3.135	2.74	2.74	2.741	2.74	
37	2.474	2.47	2.47	2.474	2.47	
38	2.178	2.18	2.18	2.178	2.18	
39	2.118	2.12	2.12	2.118	2.12	

Tab.13. Względna suma kwadratów odległości wewnątrz grup d^2 /SKOWG/LOB/
Zbiór 1

Liczba grup	SLM	CLM	MPG	MPA	MPB
2	737.45	625.90	330.93	376.35	737.45
3	667.17	321.27	292.26	339.08	623.64
4	559.02	308.05	258.21	179.68	570.85
5	507.56	299.05	222.41	167.93	508.18
6	507.11	108.31	193.90	96.63	449.24
7	459.83	87.12	106.05	66.03	402.40
8	401.95	84.86	84.86	57.70	240.85
9	373.39	44.61	51.80	35.93	189.90
10	330.77	30.13	49.06	30.37	169.77
11	184.06	22.23	27.39	25.83	68.49
12	138.44	19.49	19.49	21.63	49.00
13	120.65	17.99	17.99	20.70	33.19
14	83.01	15.55	17.53	19.98	28.71
15	80.76	12.74	15.09	16.68	26.46
16	80.28	12.28	13.66	12.81	22.67
17	26.22	8.34	10.85	10.39	15.27
18	21.74	6.91	6.91	9.55	12.58
19	18.83	6.13	6.33	8.80	10.27
20	16.74	5.56	5.56	7.77	7.77
21	12.57	4.71	5.08	7.41	6.54
22	12.40	3.99	4.37	5.60	6.09
23	7.81	3.51	3.91	4.70	5.31
24	6.58	3.05	3.05	4.05	3.95
25	4.08	2.43	2.43	3.72	3.32
26	3.30	2.06	2.25	3.24	2.65
27	2.70	1.87	1.87	2.80	2.17
28	2.31	1.54	1.70	2.37	1.76
29	1.87	1.36	1.36	1.92	1.37
30	1.20	1.20	1.20	1.55	1.20
31	1.06	0.94	1.07	1.17	0.94
32	0.94	0.81	0.81	0.98	0.81
33	0.71	0.57	0.69	0.80	0.57
34	0.60	0.45	0.45	0.63	0.45
35	0.34	0.34	0.34	0.47	0.34
36	0.24	0.24	0.24	0.34	0.24
37	0.16	0.16	0.16	0.22	0.16
38	0.10	0.10	0.10	0.22	0.10
39	0.04	0.04	0.04	0.05	0.04

Tab 14. Względna suma kwadratów odległości wewnątrz grup \bar{d}^2 /SKOWG/L
Zbiór 2

Liczba grup	SLM	CLM	MPG	MPA	MPB
2	720.50	720.50	720.50	381.06	720.50
3	587.90	443.35	587.90	206.43	587.90
4	549.34	204.21	301.65	177.45	415.91
5	423.39	193.06	154.72	108.44	293.19
6	265.02	96.65	76.73	96.47	146.26
7	247.19	48.04	68.27	94.94	75.49
8	225.50	42.19	35.28	65.34	65.60
9	225.22	33.73	31.48	57.85	56.37
10	126.88	24.51	27.39	39.66	35.74
11	58.92	20.74	22.22	38.37	26.66
12	56.93	12.96	21.40	33.71	22.86
13	55.04	11.65	17.84	31.73	19.95
14	45.81	10.71	16.90	29.32	17.80
15	40.28	10.29	9.12	28.04	14.96
16	21.59	8.30	8.48	22.58	12.83
17	20.67	6.83	7.85	12.81	11.51
18	20.47	6.19	7.57	10.92	9.52
19	12.136	5.72	5.36	10.09	6.19
20	10.010	5.44	4.65	9.63	5.26
21	6.679	4.11	4.18	8.81	4.80
22	6.278	3.63	3.97	5.49	4.08
23	5.810	3.40	3.55	5.11	3.80
24	3.915	3.05	3.15	4.75	3.38
25	3.108	2.39	2.95	3.84	3.04
26	2.778	1.98	2.60	3.32	2.69
27	2.620	1.78	2.43	2.90	2.03
28	2.467	1.59	1.77	2.41	1.63
29	2.319	1.43	1.43	2.19	1.43
30	1.654	1.27	1.27	1.59	1.27
31	1.306	1.12	1.12	1.39	1.12
32	1.181	0.97	0.97	1.08	0.97
33	0.843	0.84	0.84	0.91	0.84
34	0.734	0.73	0.73	0.76	0.73
35	0.637	0.37	0.64	0.64	0.37
36	0.549	0.27	0.27	0.27	0.27
37	0.186	0.19	0.19	0.18	0.19
38	0.109	0.11	0.11	0.11	0.11
39	0.053	0.05	0.05	0.05	0.05

Tab. 15. Suma kwadratów odległości od centroidów wewnątrz grup. Zbiór 1

liczba grup	SIM	CLM	MPG	LPA	MPB
2	756.38	709.93	633.79	666.34	756.36
3	702.29	575.44	584.89	611.29	682.81
4	629.72	519.63	506.24	525.71	642.86
5	588.64	473.29	458.07	474.75	589.35
6	579.55	341.89	400.97	398.32	537.07
7	540.98	304.57	314.75	330.85	496.30
8	487.21	277.44	277.44	298.25	388.40
9	466.73	233.48	231.49	263.54	337.26
10	443.12	193.79	216.71	235.15	316.03
11	331.15	168.78	179.00	210.23	233.17
12	277.79	154.00	154.00	192.37	195.93
13	255.49	141.06	141.06	173.64	166.32
14	214.43	124.60	131.97	159.24	153.01
15	202.38	111.99	115.51	140.53	140.96
16	196.26	102.90	106.45	126.28	129.06
17	127.16	88.22	93.84	117.86	108.96
18	113.84	79.16	79.16	108.89	99.94
19	103.59	72.75	72.41	100.13	90.43
20	95.32	66.04	66.04	91.74	79.51
21	87.17	60.37	59.92	84.51	71.71
22	83.67	54.87	54.42	74.15	62.62
23	64.53	48.75	49.42	66.73	56.34
24	56.72	43.74	43.74	59.16	49.15
25	45.80	38.52	38.52	52.48	44.59
26	39.53	34.32	34.79	46.36	38.94
27	35.34	30.60	30.60	41.00	32.82
28	30.68	26.66	27.09	36.27	27.98
29	25.47	23.16	23.16	31.40	23.33
30	19.83	19.83	19.83	27.60	19.83
31	17.20	16.86	17.20	23.36	16.86
32	14.79	14.23	14.23	19.63	14.23
33	12.00	11.44	11.82	16.01	11.44
34	9.76	9.03	9.03	12.51	9.03
35	6.79	6.79	6.79	9.44	6.79
36	4.81	4.81	4.81	6.78	4.81
37	3.25	3.25	3.25	4.36	3.25
38	1.93	1.93	1.93	2.38	1.93
39	0.81	0.81	0.81	0.81	0.81

Tab. 16. Suma kwadratów odległości od centroidów wewnątrz grup. Zbiór

Liczba grup	SIM	GIM	MPG	MPA	MPB
2	738.98	738.98	738.98	651.65	738.98
3	640.87	617.55	640.87	550.52	640.87
4	615.67	424.16	448.99	496.57	513.55
5	525.17	370.69	359.41	453.92	406.89
6	381.45	284.36	278.39	425.98	317.32
7	368.97	238.66	236.30	395.46	237.81
8	349.84	218.52	192.71	353.41	221.91
9	344.24	176.42	179.08	330.18	205.66
10	275.34	156.62	161.15	282.66	179.91
11	192.56	141.75	144.99	257.00	157.23
12	182.06	125.50	135.19	240.29	143.61
13	173.54	116.18	120.52	216.40	131.41
14	157.30	108.26	112.60	195.76	120.39
15	149.29	99.73	96.35	181.15	111.51
16	124.00	89.24	88.86	149.20	104.20
17	116.63	82.24	81.78	127.81	95.91
18	112.62	74.75	76.18	110.89	85.42
19	88.60	69.16	68.03	101.12	73.40
20	81.29	63.56	62.31	91.88	66.02
21	69.27	58.10	56.73	82.08	60.44
22	64.94	52.22	52.47	72.07	54.72
23	59.35	47.65	47.86	64.55	49.12
24	46.50	43.61	43.53	57.23	44.51
25	40.78	39.30	39.52	50.02	40.52
26	36.17	34.96	35.48	43.68	36.48
27	33.01	30.96	32.13	39.07	32.17
28	29.96	27.18	27.82	33.05	27.83
29	27.00	23.83	23.83	28.74	23.83
30	22.69	20.67	20.67	24.19	20.67
31	18.65	17.61	17.61	20.18	17.61
32	16.14	14.66	14.66	16.84	14.66
33	12.15	12.15	12.15	13.49	12.15
34	9.97	9.97	9.97	10.53	9.97
35	8.03	7.42	8.03	8.03	7.42
36	6.25	5.48	5.48	5.48	5.48
37	3.71	3.71	3.71	3.71	3.71
38	2.18	2.18	2.18	2.18	2.18
39	1.06	1.06	1.06	1.06	1.06

Tab. 17. Suma kwadratów odległości między grupami. Zbiór 1

Liczba grup	SLM	GLM	MPG	MPA	MPB
2	83.64	130.06	206.21	153.67	83.64
3	137.71	264.57	255.11	228.71	157.19
4	210.28	320.37	333.76	314.29	197.14
4	251.36	366.72	381.93	365.26	250.65
6	260.45	498.12	439.03	441.68	302.93
7	299.02	535.43	525.25	509.15	343.70
8	352.79	562.56	562.56	541.75	451.60
9	373.27	606.53	608.51	576.46	502.74
10	396.88	646.21	623.29	604.85	523.97
11	508.85	671.22	661.00	629.77	606.83
12	562.21	686.00	686.00	647.63	644.07
13	584.51	698.94	699.00	666.36	673.68
14	625.57	715.40	708.03	680.76	686.99
15	637.62	728.01	724.49	699.47	699.04
16	643.74	737.10	733.55	713.72	710.94
17	712.84	751.78	746.16	722.14	731.04
18	726.16	760.84	760.84	731.11	740.06
19	736.42	767.21	767.59	739.87	749.57
20	744.69	773.96	773.96	748.26	760.49
21	752.83	779.63	780.08	755.49	768.29
22	756.34	785.13	785.58	765.85	777.38
23	775.47	791.25	790.58	773.27	783.66
24	783.28	796.26	796.26	780.84	790.85
25	794.20	801.48	801.48	787.52	795.42
26	800.47	805.68	805.21	793.64	801.06
27	804.66	809.40	809.40	799.00	807.18
28	809.32	813.34	812.91	803.74	812.02
29	814.53	816.84	816.84	808.60	816.67
30	820.17	820.17	820.17	812.40	820.17
31	822.80	823.14	822.80	816.64	823.14
32	825.21	825.77	825.77	820.37	825.77
33	828.00	828.56	828.18	823.99	828.56
34	830.24	830.97	830.97	827.50	830.97
35	833.21	833.21	833.21	830.56	833.21
36	835.19	835.19	835.19	833.22	835.19
37	836.75	836.75	836.75	835.64	836.75
38	838.07	838.07	838.07	837.62	838.07
39	839.19	839.19	839.19	839.19	839.19

Tab. 18. Suma kwadratów odległości między grupami.

Zbiór 2

Liczba grup	SLM	CLM	MPG	MPA	MPB
2	101.02	101.02	101.02	188.35	101.02
3	199.13	222.45	199.13	189.48	199.13
4	224.34	415.84	391.01	343.43	326.45
5	314.83	469.30	480.59	386.08	433.11
6	458.55	555.64	561.61	414.02	522.68
7	471.03	601.34	603.70	444.54	602.19
8	490.16	621.49	647.29	486.59	618.09
9	495.76	663.58	660.92	509.82	634.34
10	564.66	683.38	678.85	557.34	660.09
11	647.44	698.25	695.01	583.00	682.77
12	657.94	714.50	704.81	599.72	696.40
13	666.46	723.82	719.48	623.60	708.60
14	682.70	731.74	727.40	644.24	719.62
15	690.71	740.27	743.65	658.85	728.49
16	716.00	750.76	751.14	690.80	735.80
17	723.37	757.78	758.22	712.19	744.09
18	727.38	765.25	763.82	729.11	754.58
19	751.40	770.84	771.97	738.88	766.60
20	758.71	776.44	777.69	748.12	773.98
21	770.73	781.90	783.27	757.92	779.56
22	775.06	787.78	787.53	767.93	785.28
23	780.65	792.35	792.14	775.45	790.88
24	793.50	796.39	796.47	782.77	795.49
25	799.22	800.71	800.48	789.98	799.48
26	803.83	805.04	804.52	796.32	803.52
27	806.99	809.05	807.87	800.93	807.83
28	810.04	812.82	812.18	806.95	812.17
29	813.00	816.17	816.17	811.26	816.17
30	817.31	819.33	819.33	815.81	819.33
31	821.35	822.39	822.39	819.82	822.39
32	823.86	825.34	825.34	823.16	825.34
33	827.85	827.85	827.85	826.51	827.85
34	830.03	830.03	830.03	829.42	830.03
35	831.97	832.58	831.97	831.97	832.58
36	833.75	834.52	834.52	834.52	834.52
37	836.29	836.29	836.29	836.29	836.29
38	<u>837.82</u>	<u>837.82</u>	<u>837.82</u>	<u>837.82</u>	<u>837.82</u>
39	838.94	838.94	838.94	838.94	838.94

Tab. 19. Iloraz: Suma kwadratów odległości między grupami podzielona przez sumę kwadratów odległości wewnątrz grup. Zbiór 1

Liczba grup	SIM	CLM	MPG	MPA	MPB
2	0.11	0.18	0.33	0.22	0.11
3	0.20	0.46	0.44	0.37	0.23
4	0.33	0.62	0.66	0.60	0.31
5	0.43	0.77	0.83	0.77	0.43
6	0.45	1.46	1.09	1.11	0.56
7	0.55	1.76	1.67	1.54	0.69
8	0.72	2.03	2.03	1.82	1.16
9	0.80	2.60	2.63	2.19	1.49
10	0.90	3.33	2.88	2.57	1.66
11	1.54	3.98	3.69	3.00	2.60
12	2.02	4.45	4.45	3.37	3.29
13	2.29	4.95	4.95	3.84	4.05
14	2.92	5.74	5.37	4.27	4.49
15	3.15	6.50	6.27	4.98	4.96
16	3.28	7.16	6.89	5.65	5.51
17	5.61	8.52	7.95	6.13	6.71
18	6.38	9.61	9.61	6.71	7.40
19	7.11	10.54	10.60	7.39	8.29
20	7.81	11.72	11.72	8.16	9.56
21	8.64	12.91	13.02	8.94	10.71
22	9.04	14.31	14.44	10.33	12.42
23	12.02	16.23	16.00	11.59	13.91
24	13.81	18.20	18.20	13.20	16.09
25	17.34	20.81	20.81	15.01	17.84
26	20.25	23.47	23.14	17.12	20.57
27	22.77	26.46	26.46	19.49	24.59
28	26.38	30.51	30.00	22.16	29.02
29	31.98	35.27	35.27	25.76	35.01
30	41.37	41.37	41.37	29.44	41.37
31	47.84	48.83	47.84	34.96	48.83
32	55.79	58.04	58.04	41.79	58.04
33	68.98	72.43	70.06	51.48	72.43
34	85.02	92.00	92.00	66.17	92.00
35	122.63	122.63	122.63	87.96	122.63
36	173.62	173.62	173.62	122.96	173.62
37	257.60	257.60	257.60	191.65	257.60
38	434.61	434.61	434.61	352.52	434.61
39	1030.95	1030.95	1030.95	1030.95	1030.95

Tab. 20. Iloraz: Suma kwadratów odległości między grupami podzielona przez sumę kwadratów odległości wewnątrz grup. Zbiór 2

Liczba grup	SIM	CLM	MPG	MPA	MPB
2	0.147	0.14	0.14	0.29	0.14
3	0.31	0.36	0.31	0.53	0.31
4	0.36	0.98	0.87	0.69	0.64
5	0.60	1.27	1.34	0.85	1.06
6	1.20	1.95	2.02	0.97	1.65
7	1.28	2.52	2.55	1.12	2.53
8	1.40	2.84	3.36	1.38	2.79
9	1.44	3.76	3.69	1.54	3.08
10	2.05	4.36	4.21	1.97	3.67
11	3.36	4.93	4.79	2.27	4.34
12	3.61	5.69	5.21	2.50	4.85
13	3.84	6.23	5.91	2.88	5.39
14	4.34	6.76	6.46	3.29	5.98
15	4.63	7.42	7.72	3.64	6.53
16	5.77	8.41	8.45	4.63	7.06
17	6.20	9.21	9.27	5.57	7.76
18	6.46	10.23	10.03	6.58	8.83
19	8.48	11.15	11.35	7.31	10.44
20	9.33	12.22	12.48	8.14	11.72
21	11.13	13.46	13.81	9.23	12.90
22	11.94	15.09	15.01	10.66	14.35
23	13.15	16.63	16.55	12.01	16.10
24	17.07	18.26	18.30	13.68	17.87
25	19.60	20.38	20.26	15.79	19.73
26	22.22	23.03	22.08	18.23	22.03
27	24.44	26.14	25.14	20.50	25.11
28	27.04	29.91	29.20	24.42	29.18
29	30.11	34.25	34.25	28.23	34.25
30	36.03	39.63	39.63	33.73	39.63
31	44.05	46.69	46.69	40.63	46.69
32	51.05	56.31	56.31	48.88	56.31
33	68.14	68.14	68.14	61.26	68.14
34	83.29	83.29	83.29	78.75	83.29
35	103.67	112.16	103.67	103.67	112.16
36	133.33	152.21	152.21	152.21	152.21
37	225.38	225.38	225.38	225.38	225.38
38	384.73	384.73	384.73	384.73	384.73
39	792.18	792.18	792.18	792.18	792.18

Tab.21. Liczba podziałów w których metoda M_w dała mniej grup jednoobektowych niż metoda M_k

$M_w \backslash M_k$	SLM	CLM	MPG	MPA	MPB
SLM		0	0	0	0
CLM	32		18	10	27
MPG	28	0		5	22
MPA	33	15	18		25
MPB	29	0	2	0	

Tab.22. Liczba podziałów w których najliczniejsza grupa w metodzie M_w była mniej liczna niż w metodzie M_k

$M_w \backslash M_k$	SLM	CLM	MPG	MPA	MPB
SLM	.	0	0	0	0
CLM	30		4	5	23
MPG	29	5		6	22
MPA	33	19	18		24
MPB	26	0	2	0	

Tab.23. Liczba podziałów w których wariancja liczebności grup w metodzie M_w była mniejsza niż w metodzie M_k

$M_w \backslash M_k$	SLM	CLM	MPG	MPA	MPB
SLM		0	0	0	1
CLM	32		16	11	27
MPG	30	5		10	27
MPA	33	17	19		27
MPB	30	0	2	0	

Tab.24. Liczba podziałów w których metoda M_k dała mniejszy średni kwadrat odległości \bar{d}^2 niż metoda M_w . Zbiór 1

$M_w \backslash M_k$	SLM	CLM	MPG	MPA	MPB
SLM		0	0	14	2
CLM	31		18	34	28
MPG	31	3		35	39
MPA	23	3	3		16
MPB	29	0	1	21	

Tab.25. Liczba podziałów w których metoda M_k dała mniejszy średni kwadrat odległości \bar{d}^2 niż metoda M_w . Zbiór 2

$M_w \backslash M_k$	SIM	CLM	MPG	MPA	MPB
SIM		0	0	18	1
CLM	32		17	32	24
MPG	30	9		31	25
MPA	16	2	2		4
MPB	31	2	1	30	

Tab.26. Liczba podziałów w których metoda M_k dała mniejszą sumę kwadratów odległości \tilde{d}^2 niż metoda M_w . Zbiór 1

$M_w \backslash M_k$	SIM	CLM	MPG	MPA	MPB
SIM		0	0	13	2
CLM	32		17	31	28
MPG	31	4		27	28
MPA	25	7	11		20
MPB	28	0	2	17	

Tab. 27. Liczba podziałów w których metoda M_k dała mniejszą sumę kwadratów odległości \tilde{d}^2 niż metoda M_w . Zbiór 2

$M_w \backslash M_k$	SLM	CLM	MPG	MPA	MPB
SLM		0	0	9	0
CLM	30		20	29	25
MPG	33	6		27	26
MPA	25	5	4		6
MPB	31	1	1	28	

Tab. 28. Liczba podziałów w których metoda M_w dała większą sumę kwadratów odległości między grupami \tilde{s}_m^2 niż metoda M_k . Zbiór 1

$M_w \backslash M_k$	SLM	CLM	MPG	MPA	MPB
SLM		0	0	16	1
CLM	32		14	36	28
MPG	31	8		37	28
MPA	21	1	0		13
MPB	29	0	2	24	

Tab. 29. Liczba podziałów w których metoda M_w dała większą sumę kwadratów odległości między grupami S_m^2 niż metoda M_k . Zbiór 2

$M_w \backslash M_k$	SIM	CLM	MPG	MPA	MPB
SIM		0	0	25	1
CLM	32		16	32	27
MPG	30	11		31	23
MPA	9	2	2		3
MPB	30	1	1	31	

Tab. 30. Liczba podziałów w których metoda M_w dała większą wartość ilorazu $I = S_m^2 / S_w^2$ niż metoda M_k . Zbiór 1

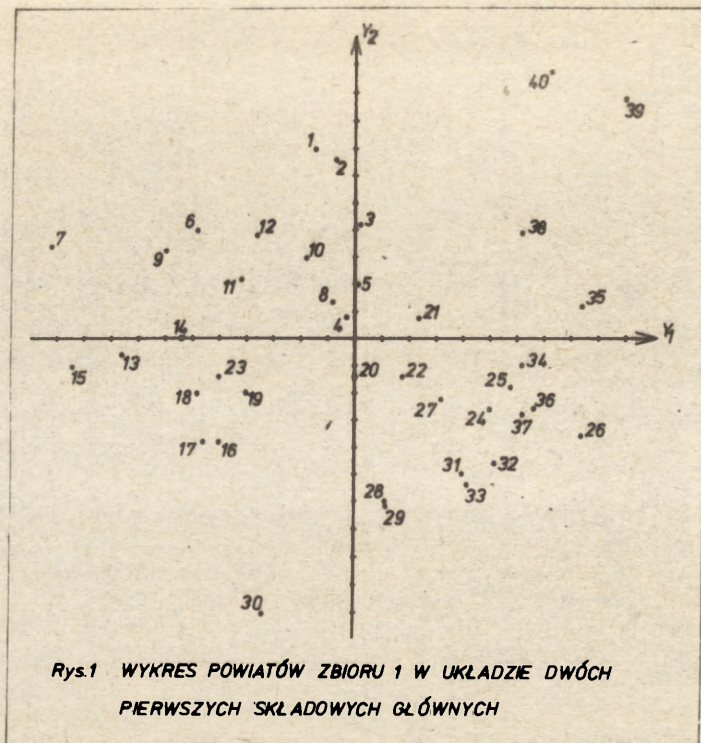
$M_w \backslash M_k$	SIM	CLM	MPG	MPA	MPB
SIM		0	0	16	1
CLM	32		14	35	28
MPG	31	7		36	28
MPA	21	1	1		13
MPB	29	0	2	24	

Tab.31. Liczba podziałów, w których metoda M_w dała większą wartość ilorazu $I = S_m^2 / S_w^2$ niż metoda M_k . Zbiór 2

$M_w \backslash M_k$	SLM	CLM	MPG	MPA	MPB
SLM		0	0	27	1
CLM	34		16	32	24
MPG	30	11		31	25
MPA	7	2	3		3
MPB	30	1	1	31	

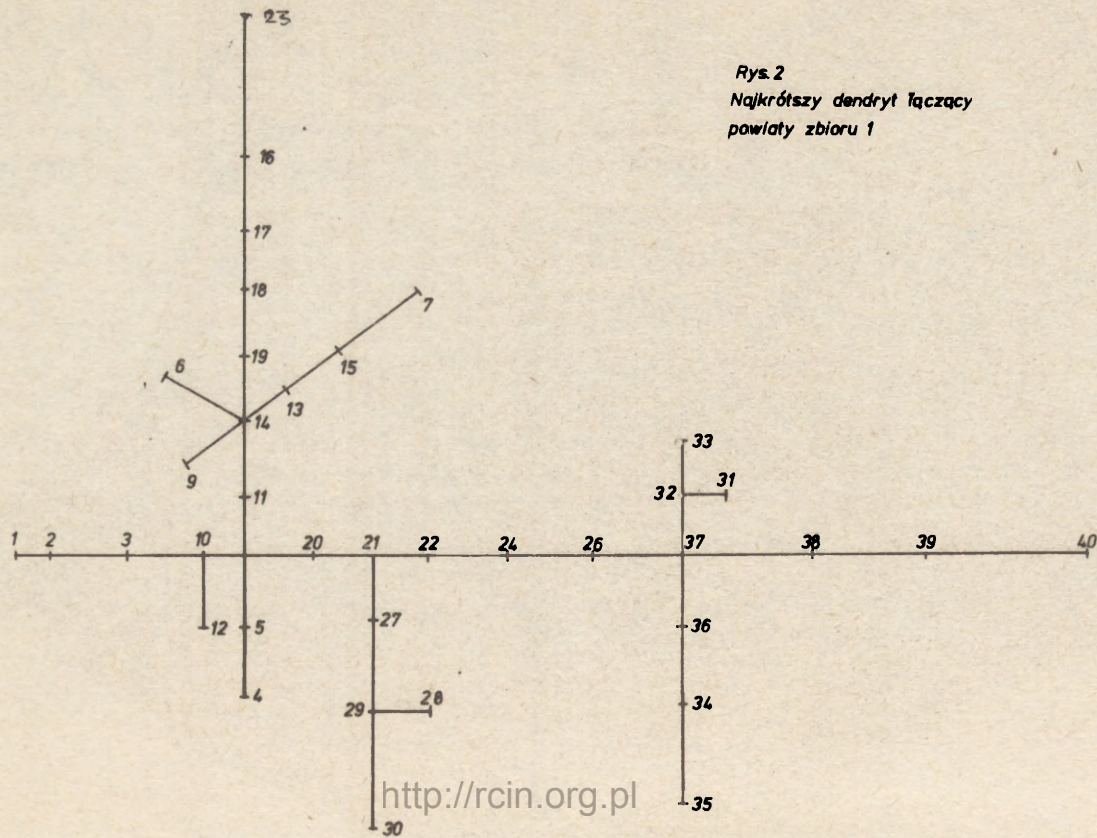
SPIS RYCIN

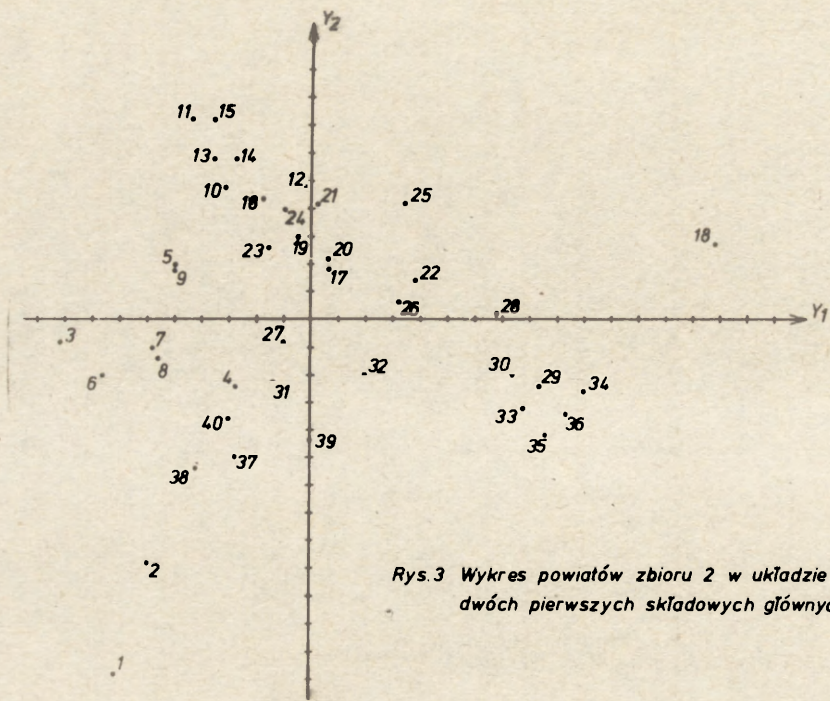
	strona
Ryc. 1. Wykres powiatów zbioru 1 w układzie dwóch pierwszych składowych głównych	91
Ryc. 2. Najkrótszy dendryt łączący powiaty zbioru 1 ...	92
Ryc. 3. Wykres powiatów zbioru 2 w układzie dwóch pierwszych składowych głównych	93
Ryc. 4. Najkrótszy dendryt łączący powiaty zbioru 2 ...	94



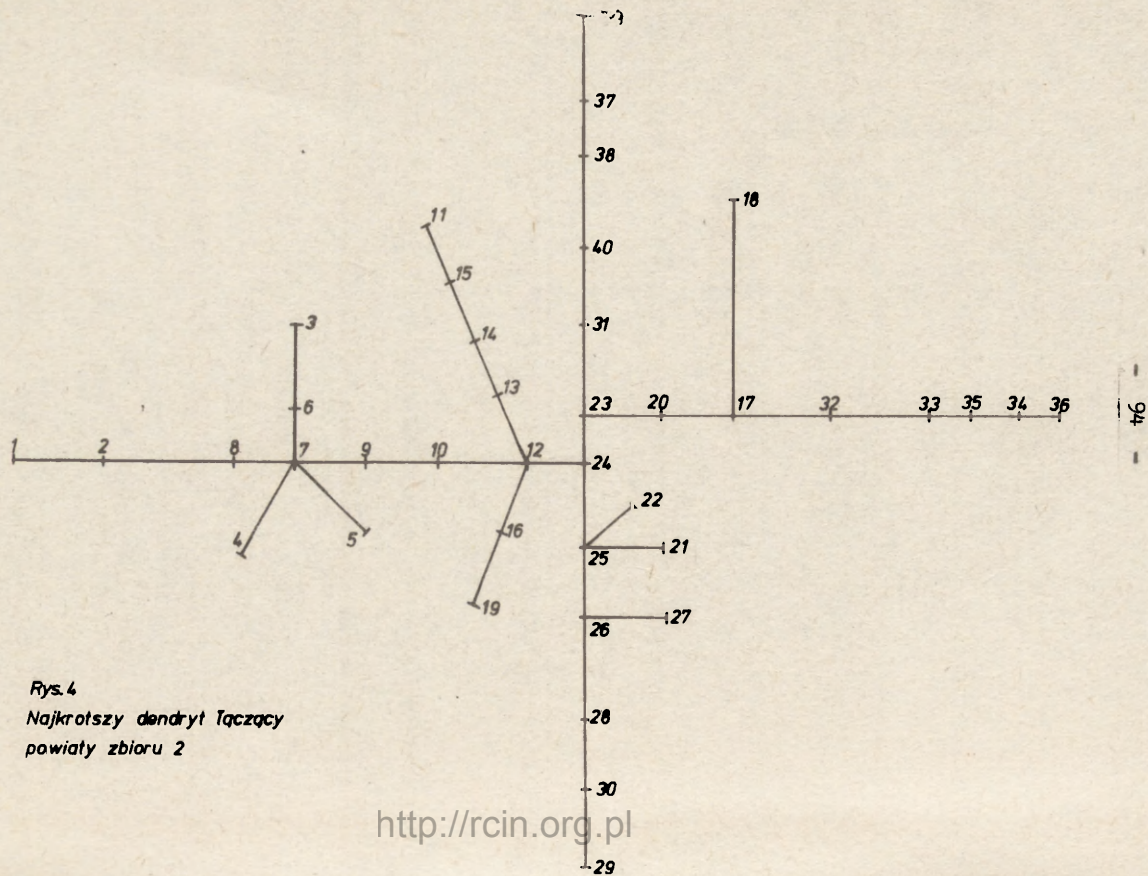
Rys.1 WYKRES POWIATÓW ZBIORU 1 W UKŁADZIE DWÓCH
PIERWSZYCH SKŁADOWYCH GŁÓWNYCH

Rys. 2
 Najkrótszy dendryt łączący
 powiaty zbioru 1





Rys.3 Wykres powiatów zbioru 2 w układzie dwóch pierwszych składowych głównych



Rys. 4
 Najkrótszy dendryt łączący
 powiaty zbioru 2

WYKAZ ZESZYTÓW BIULETYNU INFORMACYJNEGO

1. Bibliografia opracowań wykonanych w latach 1970—1973. 1974.
2. Studia nad infrastrukturą społeczną „sensu stricto”. 1974.
3. Struktura przestrzenna rolnictwa. Wybór prac. 1975.
4. Wariantowa prognoza rozwoju transportu w Polsce do roku 2000. 1975.
5. Gospodarka wodna. 1975.
6. System regionalnych centrów i ośrodków wzrostu. Materiały z seminarium — Warszawa 22 V 1973. 1975.
7. Wybrane problemy przestrzennego kształtowania infrastruktury społecznej. 1975.
8. Surowik Cz. Wybrane problemy przestrzennego kształtowania infrastruktury ochrony zdrowia i pomocy społecznej. 1975.
9. Prognoza rozwoju infrastruktury ekonomiczno-technicznej Polski do roku 2000. 1975.
10. Uprzemysłowienie i struktura przestrzenna przemysłu. 1976.
11. Nowe metody i techniki w planowaniu przestrzennym. 1976.
12. Problem węzłowy 11.2.1. Podstawy przestrzennego zagospodarowania kraju. Synteza badań I. 1976.
13. Problem węzłowy 11.2.1. Podstawy przestrzennego zagospodarowania kraju. Synteza badań II. 1976.
14. Kształtowanie wewnętrznej struktury jednostek osadniczych. Raport końcowy części B problemu węzłowego 11.2.1. „Zagospodarowanie przestrzenne miast i osiedli wiejskich”. 1976.
15. Dziembowski Z. Ekonomiczna efektywność infrastruktury gospodarczej w aglomeracjach miejskich. Raport końcowy tematu. 1976.
16. Bibliografia opracowań wykonanych w latach 1971—1975. 1977.
17. Szczotka F. A. Podstawy taksonomii numerycznej. 1976.
18. Ekonomiczna efektywność przemysłu zlokalizowanego w aglomeracjach miejskich. 1977.
19. System gospodarki przestrzennej, mechanizmy i organizacja. 1977.
20. Truskowska R. Charakterystyka warunków przyrodniczych rolnictwa. Metoda i mapa. Załącznik — rejestr. 1977.
21. Zarys organizacji i planowania gospodarki przestrzennej. 1977.
22. Problemy kształtowania przestrzennej struktury przemysłu. 1978.