Akademia Ekonomiczna im. Oskara Langego
we Wrocławiu
Wydział Zarządzania i Informatyki

Krzysztof Jajuga

METODY ANALIZY WIELOWYMIAROWEJ W ILOŚCIOWYCH BADANIACH PRZESTRZENNYCH

Praca doktorska przygotowana
pod kierunkiem
doc. dr Stanisławy Bartosiewicz

Wrocław 1981

SPIS TREŚCI.

WSTĘP.

Udoskonalanie narzędzi i sposobów poznania jest jedną z najważniejszych prawidłowości rozwoju cywilizacyjnego człowieka. W ostatnich kilkudziesięciu latach obserwowany jest szczególnie burzliwy rozwój procesu poznania, przy czym dotyczy to zarówno poznania przyrody, jak i poznania praw rządzących działalnością ludzką. W tej ostatniej dziedzinie poczesne miejsce zajmuje odkrywanie praw funkcjonujących w sferze działalności ekonomicznej.

Coraz bardziej skomplikowane procesy zachodzące w sferze gospodarki powodowały poszukiwanie nowych narzędzi opisu tych procesów. Do tego celu zapożyczano sposoby i metody stosowane w innych dziedzinach /np. w technice/. Sporą popularność zdobyły wszelkiego rodzaju metody matematyczno-statystyczne, a wśród nich metody analizy wielowymiarowej. Zaleta metod analizy wielowymiarowej tkwi w możliwości stosunkowo prostej kwantyfikacji skomplikowanych zależności, które występują w zbiorze jednostek badania.

Większość zjawisk i problemów badawczych w dziedzinie ekonomii ma charakter złożony, tzn. opisywane są one za pomocą wielu charakterystyk. Prosty opis tych złożonych, wielowymiarowych zjawisk to zadanie metod analizy wielowymiarowej. Poczesne miejsce wśród zadań analizy wielowymiarowej zajmuje klasyfikacja, pozwalająca zidentyfikować pewne wspólne cechy różnych jednostek badania.

Celem pracy jest prezentacja teoretycznych podstaw analizy wielowymiarowej, ze szczególnym uwzględnieniem metod klasyfikacji, regionalizacji i klasyfikacji rozmytej, na tle możliwości ich zastosowania w ilościowych badaniach przestrzennych, tzn. takich badaniach ilościowych, w których jednostkami /obiektami/ badania są obszary kraju.

Problematyka ta ma niebagatelne znaczenie praktyczne, zwłaszcza dla potrzeb właściwej polityki społeczno-ekonomicznej kraju. Najważniejszym elementem tej polityki jest takie planowanie rozwoju społeczno-gospodarczego lub rozwoju wybranej dziedziny, które likwidowałoby istniejące dysproporcje występujące w różnych regionach kraju. Do identyfikacji tych dysproporcji przydatne mogą być wyżej wymienione metody, przy czym każde z nich niosą nieco odrębne informacje.

Podstawowym narzędziem identyfikacji dysproporcji w przekroju przestrzennym są metody klasyfikacji. Klasyfikacja ma na celu wyodrębnienie klas obszarów kraju, charakteryzujących się zbliżonym poziomem badanych zjawisk , czyli zbliżonymi wartościami zmiennych opisujących te zjawiska. Przynosi to określone walory poznawcze dla potrzeb polityki społeczno-ekonomicznej, umożliwia zróżnicowanie tej polityki wg układu obszarów regionalnych.

Dodatkowe informacje niesie stosowanie metod regionalizacji. Regionalizacja ma na celu wyodrębnienie klas obszarów, w których istnieje zbliżony poziom zjawisk, a które ponadto posiadają własność ciągłości przestrzennej /na mapie tworzą zwarty kompleks/. Takie obszary /regiony/ dla praktycznych potrzeb polityki społeczno-ekonomicznej są bardziej przydatne niż obszary typu regionalnego /tzn. takie, które nie posiadają własności ciągłości przestrzennej/, gdyż mogą być łatwiej planowane i zarządzane.

Jeszcze inne, bogatsze w stosunku do klasyfikacji, informacje niesie klasyfikacja rozmyta. Ma ona na celu wyodrębnienie klas obszarów o zbliżonym poziomie zjawisk, przy czym każdy obiekt /obszar/ należy do różnych klas w różnym stopniu. Stopień ten określany jest liczbą z przedziału $\langle 0,1 \rangle$. Im liczba ta jest większa, w tym większym stopniu obiekt należy do klasy. Pozwala to wobec tego na większe niż klasyfikacja różnicowanie informacji o danym obiekcie, nie zatracając przy tym stopnia ogólności, jaki niesie klasyfikacja. Dzięki temu decyzje podejmowane w sferze polityki społeczno-ekonomicznej mogą być bardziej precyzyjne.

Empiryczna część pracy dotyczy przestrzennych badań usług. W miarę rozwoju cywilizacyjnego potrzeby człowieka rosną, a przy tym zmienia się ich struktura. Charakterystyczny jest wzrost udziału w tej strukturze potrzeb różnego rodzaju usług.

Usługi, podobnie jak dobra, są elementem warunków ludzkiego bytu. Z punktu widzenia szeroko rozumianego dobra człowieka, zdaniem autora nadrzędnego celu rozwoju ludzkości, należy dążyć do tego, aby dostępność usług była w różnych obszarach zbliżona. Elementem polityki społeczno-ekonomicznej w sferze usług powinna być, jak już wspomniano, likwidacja istniejących dysproporcji. Stąd wynika duże znaczenie opisanych powyżej narzędzi.

Praca, która jest poświęcona zasygnalizowanym powyżej zagadnieniom, składa się z pięciu rozdziałów.

Rozdział pierwszy zawiera wprowadzenie do badań dotyczących regionu. W 1.1. przedstawiona została krótka historia tych badań. Na jej tle uwaga została poświęcona ewolucji pojęcia "region". Paragraf 1.2. zawiera analizę podstawowych koncepcji regionu ekonomicznego. Może on być traktowany jako pojęcie czy narzędzie analityczne stosowane w badaniach geograficzno-ekonomicznych, jako ukształtowana w wyniku rozwoju sił wytwórczych naturalna jednostka obszaru kraju lub jako obszar kraju wyodrębniony dla potrzeb planowania i zarządzania. Przedstawiony został pogląd autora dotyczący wzajemnych zależności między tymi koncepcjami. W 1.3. zostały sformułowane i zapisane w postaci sformalizowanej podstawowe cechy regionu ekonomicznego.

W rozdziale drugim zawarto teoretyczne podstawy analizy wielowymiarowej. Poprzedzone są one przedstawionymi w 2.1. syntetycznymi uwagami na temat terminu "analiza wielowymiarowa", a także na temat stosowanego przez autora podejścia. Można by je nazwać deterministycznym, przy czym determinizm wynika nie z braku mechanizmu losowego rządzącego zjawiskami społeczno-ekonomicznymi, lecz z faktu, iż badania za pomocą metod analizy wielowymiarowej są z reguły badaniami całej populacji, nie ma zatem prób losowych. W 2.2. przedstawiono podstawowe pojęcia analizy wielowymiarowej. Są to pojęcia "obiektu" i "zmiennej". Wskazano na dualizm tych pojęć. Uwagę poświęcono problemowi normowania zmiennych. Zdefiniowano również inne ważne pojęcia, "podobieństwo" i "odległość". Pozwalają one mierzyć różnego rodzaju zależności występujące w zbiorze obiektów badania. Przedstawiono różne sposoby wyznaczania funkcji podobieństwa i funkcji odległości. Paragraf 2.3. poświęcony jest systematyzacji i syntetycznej prezentacji metod i zadań analizy wielowymiarowej. Wyodrębniono dwie podstawowe grupy metod: metody badania związków między zmiennymi i metody badania zbioru obiektów. Omówiono metody wchodzące w skład obu grup. Nieco więcej uwagi poświęcono metodom klasyfikacji i analizie dyskryminacyjnej. Wskazano na związki tych metod z modną ostatnio teorią rozpoznawania obrazów.

Rozdział trzeci poświęcony jest obszernemu omówieniu metod klasyfikacji i regionalizacji. W 3.1. dokonano zagadnienia klasyfikacji i regionalizacji, przedstawiono również wzór /wraz z dowodem/ na ilość możliwych klasyfikacji. Paragraf 3.2. zawiera syntetyczne omówienie i analizę metod klasyfikacji wg zaproponowanego kryterium zestawienia z podziałem na trzy grupy metod: metody grupowania, metody podziału i metody iteracyjne. Odnośnie każdej grupy metod zaprezentowano najciekawsze propozycje, porównanie różnych metod, a także pewne modyfikacje. Szczególną uwagę poświęcono sposobowi wyboru z ciągu klasyfikacji otrzymanych w wyniku hierarchicznych metod grupowania /jest to najpowszechniej stosowana grupa metod/ klasyfikacji ostatecznej, w pewnym sensie najlepszej. W 3.3. przedstawiono metody regionalizacji. Warunek ciągłości przestrzennej obszarów poddawanych klasyfikacji jest realizowany przez koncepcję tzw. warunku przyległości. Podano sposoby adaptacji metod klasyfikacji do zagadnienia regionalizacji, a także inne, między innymi własne, propozycje dotyczące metod regionalizacji.

Rozdział czwarty zawiera propozycję zastosowania do niektórych metod analizy wielowymiarowej koncepcji tzw. zbiorów rozmytych. W 4.1. przedstawione zostały podstawowe pojęcia teorii zbiorów rozmytych. Zbiór rozmyty jest uogólnieniem zbioru w zwykłym sensie. Dany element może należeć do zbioru w różnym stopniu, przy czym stopień ten określany jest przez tzw. funkcję przynależności, przyjmującą wartości z przedziału $\langle 0,1 \rangle$. Paragraf 4.2. zawiera propozycję zastosowania pojęcia zbioru rozmytego w zagadnieniu klasyfikacji. Wprowadzono pojęcie klasyfikacji rozmytej bezwzględnej i względnej, podano również zaczerpnięte z literatury i własne propozycje konstruowania klasyfikacji rozmytej. Koncepcja klasyfikacji rozmytej może być szczególnie przydatna, zwłaszcza ze względów praktycznych, tam, gdzie nie istnieje klasyfikacja naturalna /tzn. gdy granice między klasami nie są "ostre", gdy pewne obiekty mogą być przydzielone do kilku różnych klas/. W 4.3. przedstawiono własną propozycję budowy modeli dla zbioru obiektów, z których każdy należy do badanego zbioru w pewnym stopniu. Korzystać można tutaj z uprzednio skonstruowanej klasyfikacji rozmytej. Wprowadzone zostały pojęcia zmiennej rozmytej oraz wartości średniej zmiennej rozmytej. Wyznaczono wzory na wariancję, kowariancję oraz współczynnik korelacji zmiennych rozmytych. Następnie pokazano sposób budowy liniowego modelu regresji dla zmiennych rozmytych. Jest to, podobnie jak poprzednie pojęcia, uogólnienie znanych ze statystyki opisowej pojęć. Zaproponowano również postać mierników dobroci modelu: współczynnika zbieżności i współczynnika zmienności losowej. Na końcu podany został wzór na prognozę wartości zmiennej objaśnianej na podstawie modelu rozmytego.

Rozdział piąty zawiera zastosowanie niektórych z opisanych w poprzednich rozdziałach metod do ilościowych badań przestrzennych usług konsumpcyjnych dla ludności dla województw Polski. W 5.1. dokonano analizy pojęcia usług. W 5.2. wyodrębniono 5 grup usług, które zostały zbadane. Są to: usługi kulturalno-rozrywkowe, usługi w zakresie ochrony zdrowia i opieki społecznej, usługi komunalno-mieszkaniowe, usługi w zakresie oświaty i wychowania, usługi bytowe. W paragrafie tym zawarto również kilka uwag na temat doboru zmiennych ,a także przedstawiono własną, stosowaną w pracy, heurystykę doboru zmiennych. W 5.3. w oparciu o wybrany zbiór zmiennych dokonano klasyfikacji w poszczególnych grupach usług. Podobnie, paragrafy 5.4. i 5.5. zawierają regionalizację i klasyfikację rozmytą. Wyniki badań uzupełnione są ich analizą oraz graficzną prezentacją. W 5.6. przedstawiono otrzymane na bazie klasyfikacji rozmytej modele zależności poziomu usług od poziomu urbanizacji, a także od poziomu inwestycji. Jest to ilustracja propozycji podanej w 4.3.

Stosowanie metod ilościowych, w szczególności metod analizy wielowymiarowej, upraszcza opis. Stwarza również pozory obiektywności. Nie należy jednak zapominać, iż stosowanie metod analizy wielowymiarowej wiąże się z szeregiem subiektywnych decyzji, których nie może podejmować laik, zaopatrzony jedynie w algorytm metody. Subiektywne decyzje podejmuje się przede wszystkim przy wyborze zbioru zmiennych, subiektywne, a czasami wręcz arbitralne decyzje występują przy samym stosowaniu metod. Dotyczy to zwłaszcza wyboru różnych wielkości krytycznych czy progowych. Ten subiektywizm decyzji poparty oczywiście wiedzą merytoryczną, a także znajomość obiektywnych narzędzi metodycznych powoduje, iż ilościowa i jakościowa analiza na równi uczestniczą w procesie badania.

1. WPROWADZENIE DO BADAŃ REGIONALNYCH.

1.1. Krótka historia badań regionalnych.

Rozwój badań regionalnych /tzn. dotyczących problematyki regionu/ wiąże się ściśle z rozwojem geografii. W geografii nowożytnej pierwsze opisy regionalne pojawiają się we Francji w wieku XVIII. W owym czasie badania regionalne podjął m.in. J.E. Guettard. Regiony były wtedy utożsamiane z jednostkami administracyjnymi, tworzonymi przez władze kraju w celu usprawnienia systemu administrowania krajem. Dopiero od połowy XVIII w. w geografii zaczynają dominować inne poglądy, głoszące, iż głównym kryterium wyodrębniania regionów winny być elementy środowiska naturalnego. Poglądy te opierają się na tezie Monteskiusza, który głosił, iż losy historyczne ludzkości wiążą się przede wszystkim ze środowiskiem geograficznym. Tego samego zdania był m.in. francuski geograf Filip Buache /1700-1773/, który uważał, że regiony należy tworzyć w oparciu o dorzecza. Jak widać z powyższych faktów, pojęcie regionu w owym czasie miało charakter fizyczno-geograficzny, region był regionem "naturalnym", tworzonym na podstawie elementów środowiska naturalnego /przy czym uwagę szczególną zwracano na środowisko hydrograficzne/.

W wieku XIX, a ściślej mówiąc, w jego pierwszej połowie, badania regionalne, oprócz Francji, prowadzone są również w Niemczech /nadmienić wypada, iż Niemcy były wtedy krajem, w którym geografia rozwinęła się najbardziej/.

Twórcą niemieckiej geografii regionalnej był Karol Ritter /1779-1859/. Dla Rittera geografia była nauką o charakterze regionalnym. Twierdził, iż wszelkie zjawiska geograficzne powinny być rozważane w ujęciu regionalnym.

W miarę upływu czasu wśród naukowców zajmujących się problematyką regionu pojawia się zainteresowanie działalnością człowieka w przekształcaniu środowiska naturalnego. Jednym z pionierów w tej dziedzinie był Fryderyk Ratzel, autor regionalnej geografii Niemiec. On, a także inni geografowie niemieccy stworzyli pojęcie tzw. "krajobrazu kulturalnego", polegające na tym, iż obok elementów środowiska naturalnego rozważane były również elementy działalności człowieka. Była to pierwsza próba stworzenia regionów uniwersalnych.

Jeszcze mocniej w badaniach regionalnych ujęli zjawiska ekonomiczno-społeczne geografowie rosyjscy.
Podziału Rosji na regiony dokonali m.in. K.I. Arseniew /1848/ i P. Siemionow Tian-Szański /1871 i 1880/. Ich regiony są to regiony uniwersalne, a więc takie, w których wyodrębnieniu czynnik ekonomiczny odgrywa rolę na równi z czynnikiem fizyczno-geograficznym.

Wiek XX przynosi rozwój badań regionalnych i nowe koncepcje. Badaniami regionalnymi zaczynają parać się przedstawiciele innych dyscyplin, między innymi ekonomiści. Powoduje to, że funkcjonować zaczynają dwa pojęcia regionu: region fizyczno-geograficzny i region ekonomiczny. Fakt, iż zjawiska ekonomiczno-społeczne są w coraz mniejszym stopniu zależne od środowiska naturalnego, uzasadnia słuszność tego rozgraniczenia.

Obecnie skoncentrujemy się jedynie na regionie ekonomicznym /tzn. wyodrębnianym tylko na podstawie zjawisk ekonomicznych/.

Rozwój badań nad regionem ekonomicznym obserwuje się po I wojnie światowej. W 1919 r. Anglik Fleure proponuje koncepcję tzw. "regionów ludzkich". Również u innego Anglika - Dickinsona występuje tzw. human region /region człowieka/, wyodrębniany na podstawie cech ekonomicznych, kulturowych, politycznych. Jeszcze bardziej konsekwentnie wyróżnia region ekonomiczny C. Vallaux w wydanej w 1929 r. pracy "Les sciences géographiques".

Wszyscy wyżej wspomniani badacze traktują region ekonomiczny jako obszar jednorodny w sensie pewnych cech opisujących zjawiska ekonomiczne /dalej zajmiemy się bardziej szczegółowo tym pojęciem/.

W okresie międzywojennym funkcjonuje również inny pogląd. Wg niego region ekonomiczny jest rozumiany jako "strefa wpływów" dużych miast. Wynika to z faktu, że występowanie szeregu zjawisk ekonomicznych ogniskuje się wokół aglomeracji miejskich. Autorem najpełniejszej teorii regionalizacji bazującej na kryterium ciążenia do aglomeracji miejskich jest Niemiec Walter Christaller. Koncepcję tę przedstawił w latach trzydziestych. Była ona bardzo teoretyczna i przedstawiała rozwiązania wynikające bardziej z geometrii niż z ekonomii. Odbiegała ona od rzeczywistości i dlatego spotkała się z dużą krytyką. Zbliżoną koncepcję przedstawił inny badacz niemiecki A. Losch.

W Polsce badania regionalne rozwinęły się w okresie międzywojennym. Do pionierów zaliczyć należy Wiktora Ormickiego, który w wydanej w 1929 r. pracy habilitacyjnej dokonuje podziału Polski na regiony.

Natomiast pierwszym Polakiem, który dokonał podziału kraju na podstawie kryterium ciążeń usługowych, był Włodzimierz Wakar, autor wydanej w 1928 r. pracy "Podział Polski na regiony gospodarcze".

Również Eugeniusz Romer zajmował się regionami. Podkreślał on regionalny charakter zjawisk gospodarczych w Polsce. Regionalizm ów wywodził przede wszystkim z różnych możliwości rozwoju poszczególnych części Polski, które pozostawały pod różnymi zaborami.

Po II wojnie światowej obserwuje się dalszy rozwój badań regionalnych. Rodzą się nowe koncepcje, dokonuje się nowych podziałów krajów na regiony. Dla przykładu, w Związku Radzieckim rozwijana jest koncepcja tzw. kompleksu produkcyjnego /różnymi koncepcjami regionu ekonomicznego zajmiemy się szerzej w 1.2/. Różnorodność koncepcji, a także fakt, iż problematyką regionu zajmowali się przedstawiciele różnych dyscyplin spowodował konieczność uporządkowania wiedzy o regionie. Podkreślić należy tutaj niebagatelny wkład Polaków w tej dziedzinie /a zwłaszcza K. Dziewońskiego, S. Leszczyckiego i A. Wróbla/.

W 1960 r. podczas XIX Międzynarodowego Kongresu Geograficznego w Sztokholmie utworzona zostaje Komisja Metod Regionalizacji Ekonomicznej, której celem była "analiza i porównanie celów i środków badań geograficznych nad problemem regionalizacji ekonomicznej, podejmowanych w różnych krajach, zarówno z punktu widzenia ich wartości dla rozwoju teorii naukowych, jak i praktycznych zastosowań" /por. [43], s. 9 /. Efektem kilkuletnich prac /w tym kilku konferencji/ tej komisji są prace [15], [42], [45], [92], [117].
W latach sześćdziesiątych i siedemdziesiątych zainteresowania

badaczy koncentrują się bardziej wokół problematyki metod regionalizacji ekonomicznej /m.in. [9],[21],[24],[2[illegible]],[26], [31],[34],[48],[49],[69],[74],[99]/, jak również ich praktycznych zastosowań /m.in.[27],[32],[34],[48],[58], [98],[99],[101],[116]/, choć nie brakuje również i pozycji dotyczących regionu ekonomicznego /por. np. [3],[44], [67],[117],[142],[150]/.

1.2. Podstawowe koncepcje regionu ekonomicznego.

Zajmiemy się obecnie krótką analizą najważniejszych koncepcji regionu ekonomicznego. W przeszłości między zwolennikami tych koncepcji istniały spory, jednak koncepcje te mogą funkcjonować równolegle, gdyż są efektem różnych podejść do problemu regionu ekonomicznego.

Region ekonomiczny może być rozważany jako:

- <u>pojęcie</u> stosowane w badaniach geograficznych i ekonomicznych;
- <u>obiekt</u> funkcjonujący <u>rzeczywiście</u>, "obiektywnie",
- <u>obszar kraju</u> wyodrębniony dla potrzeb planowania i zarządzania gospodarką narodową.

1. Region ekonomiczny jako pojęcie, narzędzie analityczne stosowane w badaniach geograficzno-ekonomicznych.

   Koncepcja ta narodziła się w Stanach Zjednoczonych. Wg jej zwolenników region ekonomiczny w rzeczywistości nie istnieje. Jest jedynie narzędziem analitycznym badania, jest pojęciem, służy jedynie do badania przestrzennego zróżnicowania zjawisk ekonomicznych. Koncepcja ta została szczegó-

łowo przedstawiona w [145]. Jej autor, David Whittlesey, twierdzi, iż "każdy wycinek, każda część powierzchni Ziemi /a zatem i kraju - przyp. K.J./ jest regionem, jeśli jest jednorodna" /por.[145], s.46/. Przy czym zaznaczyć należy, że jednorodność ta określana jest na podstawie kryteriów uzależnionych od celu badania. Widać stąd, że koncepcja ta dopuszcza istnienie wielu ugrupowań regionalnych, ściślej mówiąc tylu, ile można postawić celów badania.
Inne ugrupowanie regionalne otrzymamy badając rolnictwo kraju, inne przy badaniu przemysłu itd.
Różne ugrupowania regionalne można otrzymać również rozpatrując różne cechy opisujące to samo zjawisko.
Whittlesey uważa, że "tak zdefiniowany region nie jest przedmiotem ani samookreślonym, ani danym przez naturę. Jest to pojęcie intelektualne /podkr. - K.J./, byt stworzony dla celów myślenia drogą wyboru pewnych cech istotnych dla jakiegoś badanego zjawiska lub problemu przestrzennego, przy pominięciu innych cech uznanych za nieistotne" /por.[145], s.47 /.
W [36] Domański nazywa taki region "narzędziem badania".

2. Region ekonomiczny jako istniejąca "obiektywnie", naturalna jednostka obszaru.

Zwolennikami tej koncepcji są przede wszystkim badacze radzieccy. W analizie problemu regionu ekonomicznego wychodzą oni z podstaw filozofii marksistowskiej. Twierdzą oni, iż region ekonomiczny istnieje "obiektywnie". Powstał on w wyniku terytorialnego podziału pracy i rozwijał się spontanicznie w ciągu lat. Jest on naturalną jednostką obszaru, rządzi się swoimi prawami i jest w pewnym sensie niezależny.

Pełne omówienie tej koncepcji zawiera praca [3]. Jej autor, P. Ałampiew uważa, że "przyznanie obiektywnego występowania jednorodności regionu /a więc nie tak, jak uważał Whittlesey - jednorodnego w sensie subiektywnie wybranych cech - przyp. K.J./ oznacza równoczesne przyznanie obiektywnych praw powstania, formowania się i rozwoju regionów ekonomicznych. Te prawa mogą być zbadane..." /por.[3], s.239/. Dodajmy tutaj, że nieco zbliżone poglądy głosił w początkach wieku angielski badacz Herbertson. Były one później rozwijane. Twierdzono, że regiony tworzą w pewnym sensie całość i są zbliżone do organizmów żywych /tzw. regiony - obiekty/.

Wg Ałampiewa problem obiektywności istnienia regionów polega na tym, "czy występuje w rzeczywistości jedność ekonomiczna w granicach regionu, pozwalająca rozpatrywać ekonomiczny region jako pewną organiczną całość" /por.[3] s.239/.

Domański w [36] nazywa taki region "przedmiotem poznania".

3. Region ekonomiczny jako wyodrębniony dla potrzeb planowania i zarządzania gospodarką narodową obszar kraju.

Ta koncepcja regionu ekonomicznego rozwinęła się również w wieku XX, choć jej pewne przejawy występowały nieco wcześniej m.in. w okresie pojawienia się w Europie prądu społeczno-politycznego zwanego regionalizmem. Miał on na celu m.in. reformę istniejących w państwach europejskich podziałów administracyjnych i przystosowanie ich do potrzeb bujnie kwitnącego życia ekonomicznego.

Stworzenie centralnego planowania gospodarki narodowej wymagało dostosowania podziału administracyjnego do potrzeb tego planowania. Stąd wynika ta koncepcja. Domański w [76] nazywa taki region "narzędziem działania". Tak wyodrębniony region powinien stać się jednostką planistyczno-administracyjną /dla przykładu w Polsce jest to województwo - lub na wyższym poziomie - makroregion planowania/.

W literaturze spotyka się dla tak rozumianego regionu określenia "region planowania" i "region planowany". To drugie określenie oznacza docelowy, wynikający z planów, przyszły region ekonomiczny, powstały w wyniku przekształceń gospodarki narodowej. Oba określenia traktować będziemy tożsamościowo, zważywszy, iż w krótkim okresie, kilkuletnim, region planowania pokrywa się z regionem planowanym.

Zajmiemy się obecnie charakterystyką głównych powiązań między powyżej przedstawionymi koncepcjami. Analiza ta nie rości sobie prawa do stworzenia jedynej słusznej koncepcji regionu ekonomicznego /gdyż nie to jest celem pracy/, a posłuży jedynie jako punkt wyjścia do dalszych rozważań.

Jak wyżej wspomniano, w przeszłości między zwolennikami powyższych koncepcji /przy czym chodzi zwłaszcza o koncepcje regionu - narzędzia badania i regionu - przedmiotu poznania/ dochodziło do sporów, które jak się wydaje, wynikały przede wszystkim z przyjęcia różnych płaszczyzn rozważań. I tak np. uczeni radzieccy zarzucają Whittleseyowi subiektywizm badawczy.

Tymczasem, region ekonomiczny może być zarówno narzędziem analitycznym badania, jak i funkcjonującym w rzeczywistości "obiektem" - jest nawet tak, że jedno implikuje drugie - żeby "coś" badać, to "coś" musi istnieć.

Na fakt ten zwracają uwagę m.in. Isard twierdząc, iż region może być zarówno pojęciem, jak i konkretną rzeczywistością /por.[67]/ oraz Thomas: "każde zjawisko czy zjawiska o określonych cechach zajmują pewną część powierzchni Ziemi, można ową część nazwać regionem. Takie ujęcie pozwala na zbliżenie (...) poglądów skrajnych, gdyż zgodnie z nim region - pojęcie intelektualne - znajduje swój wyraz w rzeczywistości w treści materialnej określonej kryteriami wyróżnienia regionu" /por.[142] s.255 /. Podobnie uważa Wróbel: "kiedy region wyznaczony jest na podstawie konkretnych kryteriów odnoszących się do jednorodności cech obszaru, wtedy (...) pojęciu regionu odpowiada w rzeczywistości konkretna treść określona kryteriami wyróżniania regionu" /por.[150], s.66 /.

Jak widać z przytoczonych rozważań, powyższe trzy koncepcje regionu ekonomicznego dotyczą jakby trzech różnych etapów badań nad regionem.

Ilustruje to poniższy schemat:

Region - przedmiot poznania
↓
Region - narzędzie badania
↓
Region - narzędzie działania

Region - przedmiot poznania, tzn. istniejący w rzeczywistości, może być badany za pomocą narzędzia badawczego zwanego regionem. Efektem tych badań jest m.in. stworzenie regionu - narzędzia działania, tzn. podziału kraju na takie obszary, z których każdy będzie kompleksowo planowany i administrowany.

Sprzeczność tkwi jednak w czym innym. Podejście reprezentowane przez Whittleseya nazwać można relatywistycznym. Zakłada on możliwość formułowania różnych celów badania, a zatem występowanie różnych ugrupowań regionalnych /podkreślamy, iż obecnie abstrahujemy od tego, czy region rozumiany jest jako rzeczywistość, czy jej odbicie w myśli/. Natomiast naukowcy radzieccy uznają istnienie tylko jednego ugrupowania regionalnego - tego danego przez historię działalności gospodarczej człowieka w kraju, a zatem przez terytorialny podział pracy, przez specjalizację poszczególnych obszarów, itd.

Sprzeczność ta wynikać może z dwóch powodów. Otóż, po pierwsze mogą występować różne podziały branżowe /a więc regiony przemysłu, rolnictwa, usług/. Oczywiste jest, że podziały te nie będą się pokrywać. Whittlesey zakłada występowanie takich różnych ugrupowań /Berezowski w [8] proponuje takie wyodrębnione na podstawie kryteriów branżowości regiony nazywać rejonami/, natomiast Ałampiew abstrahuje od regionów branżowych.

Po drugie, nawet przy rozważaniu regionu ekonomicznego /rozumianego ogólnie/ Whittlesey stwierdza występowanie różnych ugrupowań regionalnych, tylu, ile celów badania zostanie postawionych. W zależności od badacza mogą zostać wyróżnione różne kryteria, które mogą dać różne regiony. I w tym przypadku mowa jest o dwóch różnych problemach. Ałampiew mówi o tym "najlepszym" podziale, tym docelowym, tym zgodnym z rzeczywistością. U Whittleseya tego podziału nie ma. Jest jedynie subiektywizm poszczególnych badaczy, którzy starają się zbliżyć do tego "prawdziwego" podziału /ale z góry wiadomo, że nie zostanie on odkryty/.

Bardzo trafnie, jak się wydaje, ujmuje ten problem Andrzej Wróbel /por.[150], s.41 /: "pytanie, czy regiony istnieją obiektywnie, można traktować więc jako równoznaczne z pytaniem czy /i jakieś najlepsze, jedynie słuszne ramy podziału terytorialnego, uzasadnione przez samo rozmieszczenie i wzajemne relacje badanych przedmiotów i zjawisk".

Podsumowująo powyższe rozważania, spór wynika raczej z przyjęcia odmiennych systemów filozoficznych, a nie z różnych poglądów na istotę regionu.
Należy się jeszcze ustosunkować do różnic między pojęciem region ekonomiczny /w sensie ogólnym/ a region przemysłu, rolnictwa, itp. Jak wyżej wspomniano, niektórzy autorzy proponują regiony branżowe nazywać rejonami, w odróżnieniu od regionu, którą to nazwę rezerwują wyłącznie do regionu ekonomicznego w sensie ogólnym.
W niniejszej pracy przyjęto nazywać wszelkie wyodrębnione obszary kraju regionami /lub obszarami typu regionalnego, o czym dalej/, zaznaczając o jaki rodzaj regionu chodzi /a więc region ekonomiozny, region usług, itd./.

1.3. Podstawowe cechy regionu ekonomicznego.

Zajmiemy się obecnie przeglądem najważniejszych cech regionu.

1. Region - zbiór jednostek administracyjnych.

Każdy kraj jest podzielony na pewne jednostki administracyjne, w celu ułatwienia zarządzania. Jako jedną z podstawowych cech regionu przyjmuje się, że powinien on składać się z tych jednostek administracyjnych, innymi słowy, dana jednostka musi należeć w całości tylko do jednego regionu. Tkwi tu milczące założenie o wewnętrznej jednolitości jednostek administracyjnych. Traktowane są one jako obiekty. Abstrahuje się zatem od wewnętrznych różnic w każdej jednostce administracyjnej.

2. Region - obszar ciągły przestrzennie.

Cecha ta, ciągłość przestrzenna, zwana również przez niektórych autorów przyległością /ang. contiguity/ oznacza, że jednostki administracyjne wchodzące w skład regionu powinny przylegać do siebie, graniczyć z sobą. Region powinien być ciągły przestrzennie.

3. Region - obszar strefowy.

Jak się wydaje, jest to najważniejsza cecha regionu. Można ją nazwać również jednolitością lub jednorodnością.
B. Winiarski /por.[146], s.175 / uważa regiony strefowe za "obszary wewnętrznie jednolite, o wyraźnej dominancie określonego rodzaju działalności społeczno-gospodarczej, wyróżniające się pod tym względem od obszarów sąsiednich". O ile teza ta ma zastosowanie tylko w odniesieniu do regionu ekonomicznego, o tyle, teza S. Berezowskiego /por.[8], s.17 / ma zastosowanie również do regionów branżowych. Wg niego "region jest to obszar, którego poszczególne części posiadają możliwie wiele różnic w stosunku do obszarów otaczających".
Strefowość oznacza podobieństwo, oznacza zbliżone wartości cech opisujących pewne zjawiska ekonomiczno-społeczne.
W odniesieniu do regionu ekonomicznego /tylko i wyłącznie/ występowanie strefowości wynika z występowania specjalizacji produkcyjnej lub innej, powodującej występowanie podobieństw między jednostkami wchodzącymi w skład regionu. Specjalizacja

w regionie ekonomicznym jest efektem terytorialnego podziału pracy. Natomiast w odniesieniu do regionów branżowych strefowość wynika często z przeszłości historyczno-ekonomicznej, z nierównomiernego rozwoju poszczególnych części kraju. Powyższe trzy cechy dotyczą zarówno regionu ekonomicznego, jak i regionów branżowych. Obecnie słów kilka poświęcimy jeszcze jednej własności regionu, przy czym własność ta występuje w zasadzie tylko w przypadku regionu ekonomicznego.

4. Region ekonomiczny - obszar kompleksowy.

Kompleksowość regionu ekonomicznego oznacza określony układ proporcji między rozwojem poszczególnych rodzajów działalności w tym regionie.
Wg Fajferka /por. [48], s.34 /, "kompleksowy rozwój regionu oznacza proporcjonalny i względnie wszechstronny rozwój działów i gałęzi określonego terytorium".
Niektórzy autorzy widzą w kompleksowości przeciwieństwo specjalizacji, o której wspomniano przy okazji rozważań o strefowości. O wzajemnych zależnościach pojęć "specjalizacja" i "kompleksowość" pisze szeroko Markos /por. [95], s.64 /, twierdząc m.in. "kompleksowość oznacza zawsze wielorakość, złożoność, ale równocześnie szczególny skład, specjalizację". Natomiast Berezowski stwierdza /por.[8] , s.57 /: "Rozwój gospodarki poszczególnych krajów przechodzi przez charakterystyczne etapy. Pierwszy z nich można by nazwać autarkicznym na niskim poziomie sił wytwórczych, niemal całkowicie w ramach gospodarki bez obrotu towarowego i przy minimalnej wymianie międzyregionalnej. Następnie wzrasta specjalizacja, po której następuje etap pełniejszego rozwoju produkcji na

potrzeby wewnątrzregionalne. Wreszcie pojawia się znowu lekkie zwiększenie specjalizacji w ramach swoistej integracji".
Z cechą kompleksowości powiązana jest własność węzłowości. Węzłowość regionu oznacza różne natężenie występowania danego zjawiska na biegunie /czyli w obszarze koncentracji/ i na peryferiach działalności. Węzłowość jest jedną z możliwych form organizacji regionu ekonomicznego. Z faktu, iż region ekonomiczny jest kompleksem, wynikają inne własności /przez niektórych autorów wymieniane odrębnie/, takie, jak:

- istnienie powiązań z innymi regionami w wyniku terytorialnego podziału pracy,
- istnienie wewnątrzregionalnych powiązań gospodarczych,
- względne zbilansowanie produkcji i konsumpcji,
- istnienie bazy surowcowej i energetycznej,
- wspólna organizacja i władza.

Wewnątrzregionalne powiązania gospodarcze, organizacja i struktura regionu mogą mieć własność węzłowości, ale tak wcale nie musi być.
Jak podkreślono wyżej, spośród powyżej wymienionych cech tylko trzy pierwsze stosują się do wszystkich typów regionów, tzn. zarówno regionu ekonomicznego, jak i regionów branżowych.

Obecnie przedstawimy te własności w postaci sformalizowanej /przy użyciu notacji stosowanej w naukach ilościowych/.
Dany jest zbiór obiektów o liczebności n - obiektami tymi są jednostki administracyjne kraju:

$$P = \{ P_1, P_2, \ldots, P_n \}$$

Na zbiorze tym określona jest rodzina podzbiorów:

$$R = \{ R1, R2, \ldots, R_k \} \qquad k < n$$

Ponadto na zbiorze P x P określona jest funkcja

$$c : P \times P \rightarrow \{0;\ 1\}$$

gdzie:

$$c\ /P_i,\ P_j/ = \begin{cases} 1, & \text{gdy jednostki administracyjne /obiekty/ } P_i \text{ i } P_j \text{ przylegają do siebie /graniczą/,} \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Na rodzinie podzbiorów - określona jest również funkcja

$$m: R \rightarrow \langle 0;\ +\infty)$$

tzw. miara jednolitości regionu o własnościach:

1. $m\ /R_j/ \geqslant 0 \qquad j = 1, 2, \ldots, k$
2. $m\ /R_i/ > m\ /R_j/ \Leftrightarrow$ region R_i jest bardziej jednolity niż region R_j, $i \neq j \qquad i, j = 1, \ldots, k$

Warunek 1 /region - zbiór jednostek administracyjnych/

a/ $R_1 \cup R_2 \cup \ldots \cup R_k = P$

b/ $R_i \cap R_j = \emptyset$, $i \neq j$, $\quad i, j = 1, \ldots, k$

c/ $\forall P_i \quad \exists!\, R_j \quad P_i \in R_j$

Warunek 2 /region - obszar przestrzennie ciągły/

$\forall P_i \in R_j \quad /j = 1, 2, \ldots, k/$ zachodzi:

$\forall P_k \in \overline{R_j} \quad \exists$ ciąg $/P_i,\ Pi_1,\ Pi_2, \ldots,\ Pi_{\overline{l}},\ P_k/$ taki, że

$$c\ /Pi,\ Pi_1/ = c\ /Pi_1,\ Pi_2/ = \ldots = c\ /Pi_{\overline{l}},\ P_k/ = 1$$

Warunek 3 /region - obszar strefowy/

$$m\ /R_j/ > m_o$$

gdzie m_o - krytyczny minimalny poziom jednolitości regionu.

Jeśli dany obszar kraju spełnia warunki 1, 2, 3 nazywać będziemy go <u>regionem</u>.

Jeśli dany obszar kraju spełnia warunki 1, 3 nazywać będziemy go obszarem typu regionalnego.
Przez regionalizację rozumieć będziemy zarówno czynność, jak i efekt wyodrębniania rodziny podzbiorów R.

Powyższe rozważania, prowadzone w sposób bardzo sformalizowany, wymagają kilku słów komentarza.
Przyjęto w nich definicję Berezowskiego, mówiącą, iż jedyną cechą regionu jest wewnętrzne podobieństwo części regionu.
W powyższej definicji abstrahuje się od celów badania. Niezależnie od celów /a zatem niezależnie od tego, czy interesuje nas badanie regionu ekonomicznego, czy regionów branżowych/ region zawsze powinien cechować się jednorodnością. Owa jednorodność czy strefowość formułowana jest w warunku 3. Należy zaznaczyć, że sposób mierzenia tej jednorodności, sposób spełnienia warunku 3 jest uzależniony od stosowania konkretnej metody regionalizacyjnej, o czym będzie traktował jeden z następnych rozdziałów.

2. PODSTAWY ANALIZY WIELOWYMIAROWEJ.

2.1. Uwagi ogólne.

W ostatnich kilkunastu latach w naukach społecznych obserwowane są coraz częstsze próby stosowania metod ilościowych. Chęć kwantyfikacji zjawisk społecznych, a także fakt, że próbują analizować je również przedstawiciele dyscyplin ilościowych powoduje, iż podejście takie staje się popularne i powszechnie akceptowane. Stosowanie metod ilościowych upraszcza opis, częstokroć również potwierdza odkryte indukcyjnie czy intuicyjnie prawa funkcjonujące w odniesieniu do zjawisk społecznych. Z drugiej strony, ich stosowanie niesie z sobą niebezpieczeństwo fetyszyzacji tych metod, pełniących rolę służebną w stosunku do analizy jakościowej zjawisk społecznych.

Szeroko stosowane są w naukach społecznych metody statystyki opisowej i statystyki matematycznej /spotyka się je np. w socjologii, psychologii, ekonomii/. Coraz bardziej popularne stają się również metody analizy wielowymiarowej.

Rozdział ten poświęcony będzie podstawowym pojęciom analizy wielowymiarowej, dokonana zostanie również próba prezentacji i systematyzacji pewnych zadań i metod służących do ich rozwiązywania.

W literaturze przedmiotu pojęcie analizy wielowymiarowej nie zostało zdefiniowane w sposób precyzyjny. Odwołując się do tej literatury, zakres pojęcia "analiza wielowymiarowa" wyznaczony jest w dużym przybliżeniu przez zakresy następujących pojęć:

w języku angielskim - multivariate analysis /por. [76],[77 /, multivariate statistical analysis /por.[5],[17],[78] ,[100]/;

w języku francuskim - analyse des données multidimensionnelles /por. [7] ,[11],[136]/;

w języku rosyjskim - mnogomiernyj statisticzeskij analiz /por.[1] , [99]/;

w języku polskim - statystyczna analiza porównawcza /por.[18], [56],[57] ,[58]/, wielowymiarowa analiza porównawcza /por. [66],[112],[114]/, analiza wielowymiarowa /por.[74]/.

Maurice Kendall pisze: "możemy zdefiniować analizę wielowymiarową jako gałąź statystyki związaną z zależnościami między zbiorami zależnych zmiennych i obiektów, w których występują". Dodaje również, że "analiza wielowymiarowa może być łatwiej zdefiniowana przez wyliczenie tematów, którymi się zajmuje, niż za pomocą formalnej definicji" /por.[77], s.9 /.

Powyższe stwierdzenia, a także prezentacja zadań analizy wielowymiarowej, która zostanie przeprowadzona w dalszej części pracy, wskazują na trudności w ścisłym zdefiniowaniu pojęcia "analiza wielowymiarowa".

Obecnie kilka słów poświęcimy podejściom dotyczącym stosowania metod analizy wielowymiarowej w praktyce. Zaznaczyć należy, że większość metod analizy wielowymiarowej są to metody bazujące na pewnych pojęciach statystyki matematycznej. Podejście, którego konsekwencją jest stosowanie metod analizy wielowymiarowej można skrótowo opisać następująco:

Dana jest populacja obiektów $\mathcal{P}$. Z populacji tej wybiera się próbkę, liczącą n obiektów /jednostek/. Oznacza się ją P. Na obiektach tej próby dokonuje się obserwacji m zmiennych losowych: X_1, X_2,..., X_m. Przez x_{ij} oznacza się wartość j-tej zmiennej, jaka została zaobserwowana w i-tym obiekcie. Należy x_{ij} zatem traktować jako i-tą realizację zmiennej losowej X_j. Mamy tu do czynienia z m-wymiarową zmienną losową $X = /X_1, X_2, ..., X_m/$, na której dokonano n obserwacji.

W opisanym powyżej podejściu abstrahuje się od błędów pomiaru. Zakłada się również ustalony moment czasu, w którym obserwacje są dokonane. Można powiedzieć, że stochastyka uwzględniona w powyższym podejściu wynika z nieograniczoności populacji P. Zdaniem autora, traktowanie zjawisk społeczno-ekonomicznych jako zmienne losowe jest zasadne wtedy, gdy istnieje możliwość powtarzania prób. W badaniach ekonomiczno-społecznych takich możliwości w zasadzie nie ma /w danych warunkach można dokonać tylko 1 obserwacji - przeciwnie niż np. w doświadczalnictwie/. Poza tym, badania za pomocą metod analizy wielowymiarowej /o czym szerzej w dalszej części pracy/, obejmują zwykle pewien wyczerpujący zbiór /np. jednostki administracyjne kraju, przedsiębiorstwa pewnej branży, robotnicy przedsiębiorstwa, itp./. Skłania to do przyjęcia tezy, iż populacja jest skończona, a próba obejmuje wszystkie jednostki /obiekty/ populacji. Można to przedstawić za pomocą następującego podejścia:

Dany jest skończony zbiór obiektów P, liczący n jednostek. Na obiektach tych dokonuje się obserwacji na zmiennych: X_1, X_2,..., X_m. Przez x_{ij} oznacza się wartość j-tej zmiennej, którą zaobserwowano w i-tym obiekcie.

W tej sytuacji mamy do czynienia z m zmiennymi /nielosowymi/, z których każda przyjmuje m wartości. Na fakt ten zwróciła uwagę T. Czyż w [32].

W powyższych rozważaniach abstrahowano od rozumienia natury zjawisk ekonomiczno-społecznych, tzn. od determinizmu i stochastyki tych zjawisk. Są to problemy natury filozoficznej, wynikające z przyjęcia określonego rozumienia pojęć konieczności i przypadkowości. Natomiast w rozważaniach powyższych losowość rozpatrywana była w ujęciu bardziej formalnym, tzn. takim, gdy wynikać może z błędów pomiaru czy niewyczerpującej próby.
Z punktu widzenia celu pracy nie ma potrzeby agitowania za określoną postawą filozoficzną, zwłaszcza, że różnice praktyczne, a także w większości teoretyczne, związane z przyjęciem stanowiska deterministycznego lub stochastycznego w rozumieniu natury zjawisk, nie występują.

2.2. Podstawowe pojęcia analizy wielowymiarowej.

2.2.1. Obiekt i zmienna.

Obecnie bardziej szczegółowo zajmiemy się niektórymi podstawowymi pojęciami analizy wielowymiarowej, większą uwagę poświęcając tym, które wykorzystywane będą w dalszej części pracy.
Jednym z nich jest pojęcie obiektu. W analizie wielowymiarowej jest ono pojęciem pierwotnym, zatem nie zostanie tutaj zdefiniowane. Chcieliśmy natomiast zwrócić uwagę na następujące zasady, które powinny być przyjęte przy badaniach za pomocą metod analizy wielowymiarowej:

1. Obiekty powinny należeć do jednej kategorii obiektów.
2. Obiekty powinny być jednolite z punktu widzenia celu badania.

Przyjęcie pierwszej zasady oznacza, że badaniu za pomocą metod analizy wielowymiarowej mogą być poddane jedynie obiekty tego samego typu, a więc np. zbiór województw Polski, zbiór przedsiębiorstw pewnej branży, zbiór robotników wydziału itp. Nie mogą się znaleźć w zbiorze obiektów obiekty różnych kategorii, a więc np. województwo, przedsiębiorstwo i robotnik, gdyż takie badanie nie ma sensu.

Zasada druga oznacza, iż np. mając na celu zbadanie poziomu rozwoju społeczno-gospodarczego województw, nie można jako zbioru obiektów przyjąć gmin. Co prawda, gmina jest obiektem bardziej jednolitym niż województwo /z punktu widzenia zjawisk obserwowanych w obiektach/, niemniej jednak z punktu widzenia celu badania województwo jest obiektem wystarczająco jednolitym.

Obiekty oznaczać będziemy: $P_1, P_2, \ldots, P_n$ /gdzie n jest liczbą naturalną/, a zbiór obiektów P.

Drugim podstawowym pojęciem analizy wielowymiarowej jest <u>zmienna</u> /cecha/. Przyjmując jako pierwotne pojęcie obiektu, zmienną można zdefiniować jako odwzorowanie:

$$X_j : P \to R \qquad j = 1, \ldots, m$$

Zmienna /ściślej, każda z m zmiennych/ jest odwzorowaniem zbioru obiektów w zbiór liczb rzeczywistych. Spotykane jest również tzw. dualne podejście, które jako pierwotne traktuje pojęcie zmiennej. Oznaczając zbiór zmiennych przez X, można obiekt zdefiniować jako:

$$P_i : X \to R \qquad i = 1, \ldots, n$$

W świetle powyższego, przy przyjęciu jako pierwotne pojęcia obiektu /zmiennej/ zbiór zmiennych /obiektów/ traktuje się jako zbiór odwzorowań.

W zależności od zbioru wartości, jakie przyjmuje zmienna, można je podzielić na /por. [4]/:

- zmienne ciągłe /zbiór wartości zmiennej jest zbiorem nieprzeliczalnym/,
- zmienne dyskretne /zbiór wartości zmiennej jest zbiorem przeliczalnym/.

Szczególnym przypadkiem zmiennych dyskretnych są tzw. zmienne binarne, czyli dychotomiczne, zwane również zerojedynkowymi /zbiór ich wartości ma liczebność 2, są one często kodowane jako 0 i 1/.

Jakkolwiek w literaturze często podkreślane jest rozróżnienie zmiennych ciągłych i dyskretnych, chcielibyśmy zwrócić uwagę na to, iż z punktu widzenia praktyki podejście takie nie zawsze jest słuszne. W zasadzie w badaniach ekonomiczno-społecznych wszystkie zmienne są dyskretne. Wynika to z faktu, iż pomiaru wartości zmiennej ciągłej dokonuje się zawsze z pewnym przybliżeniem, będącym konsekwencją niedokładności pomiaru /lub np. zaokrąglenia do pewnego miejsca po przecinku przy obliczaniu wielkości względnych/. Zatem zmienna, która teoretycznie może przyjmować nieprzeliczalną ilość wartości jest mierzona tylko za pomocą przeliczalnej ilości wartości.

Zmienna: średnia temperatura w ^{o}K jest zmienną ciągłą, jednak istnieją możliwości jej pomiaru tylko z pewną dokładnością /np. 0,1 ^{o}K/. Oczywiście, taka "udyskretniona" zmienna ciągła przyjmuje wartości ze zbioru o bardzo dużej liczebności, nie zmienia to faktu, iż praktycznie jest zmienną dyskretną. W niniejszej pracy zmienną powyżej scharakteryzowaną traktować będziemy jako ciągłą /tak jak to czynią wszyscy badacze zajmujący się stosowaniem metod analizy wielowymiarowej/, zwróciwszy uwagę, iż w praktyce nie powoduje to żadnych ujemnych konsekwencji - np. przy liczeniu odległości, itp.

W zależności od stosowanej skali prezentacji wartości zmiennych, zmienne można podzielić na /por. [4], [81], [136]/:

- zmienne prezentowane na skali nominalnej /np. zawód/,
- zmienne prezentowane na skali porządkowej /np. ilość lat nauki/,
- zmienne prezentowane na skali interwałowej /np. temperatura w ^{o}C/,
- zmienne prezentowane na skali stosunkowej /np. temperatura w ^{o}K/.

Oznaczmy przez X_i i X_j wartości przyjęte przez jakąś zmienną X w obiekcie i-tym i j-tym. Skala prezentacji wartości zmiennej X implikowana jest przez możliwość porównania wartości tej zmiennej. I tak:

1. $X_i = X_j$ lub $X_i \neq X_j$ $\Rightarrow$ X prezentowane na skali nominalnej
2. $X_i = X_j$ lub $X_i > X_j$ lub $X_i < X_j$ $\Rightarrow$ X prezentowane na skali porządkowej
3. $X_i > X_j$ /$X_i < X_j$/ to X_i większe /mniejsze/ od X_j o $X_i - X_j$ $\Rightarrow$ X prezentowane na skali interwałowej
4. $X_i > X_j$ /$X_i < X_j$/ to X_i większe /mniejsze/ od X_j o $X_i - X_j$ oraz X_i większe /mniejsze/ od X_j $\frac{X_i}{X_j}$ razy $\Rightarrow$ X prezentowane na skali stosunkowej.

Dokonując "krzyżowej" klasyfikacji ze względu na oba kryteria /jak czyni to np. Anderberg w [4]/ można wyróżnić 7 /a przy wyłączeniu ze zmiennych dyskretnych zmiennych binarnych 11 /

różnych rodzajów zmiennych.

Analiza zmiennych, jakie występują w badaniach społeczno-ekonomicznych, wskazuje na podstawowe rodzaje zmiennych spotykane w praktyce. Są to, zdaniem autora:

1. zmienne ciągłe prezentowane na skali stosunkowej,
2. zmienne dyskretne prezentowane na skali porządkowej /z pominięciem zmiennych binarnych/,
3. zmienne dyskretne prezentowane na skali nominalnej /z pominięciem zmiennych binarnych/,
4. zmienne binarne prezentowane na skali nominalnej.

A oto przykłady powyższych typów zmiennych:

1. ilość samochodów na tys. mieszkańców, przeciętna powierzchnia użytkowa mieszkania w m^2 na mieszkańca itp.
2. przeciętna ilość lat nauki robotnika przedsiębiorstwa, przeciętna jakość towaru wg ocen ekspertów, itp.
3. profil przedsiębiorstwa /np. włókiennicze, odzieżowe, itp./
4. występowanie lub nie uczelni wyższej w województwie, itp.

Wartości zmiennych dla poszczególnych obiektów przedstawia się w tzw. macierzy danych $X = \{ \overline{x_{ij}} \}$

gdzie: $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$.

Macierz ta ma wymiary n x m. Jej element x_{ij} jest to wartość j-tej zmiennej zaobserwowana w i-tym obiekcie. Oczywiście kolumny tej macierzy interpretować można jako zmienne, a wiersze jako obiekty.

W takim ujęciu zbiór obiektów P można traktować jako zbiór n punktów przestrzeni m-wymiarowej, a zbiór zmiennych jako zbiór m punktów przestrzeni n-wymiarowej.

W pracy będziemy zajmować się głównie zmiennymi ciągłymi prezentowanymi na skali stosunkowej.

Jak łatwo się zorientować, cel badań analizy wielowymiarowej jest zwykle tak sformułowany, że zmienne mogą odzwierciedlać różne, często nie powiązane ze sobą zjawiska. Dlatego wartości ich są wyrażone w różnych jednostkach miary, a ponadto są wielkościami różnych rzędów. Zachodzi potrzeba unormowania tych zmiennych, tzn. po pierwsze, sprowadzenia ich wartości do wielkości niemianowanych, po drugie, ujednolicenia rzędów wielkości. Istnieje wiele sposobów normowania zmiennych. Traktuje o nich np. [18]. Dla zmiennych ciągłych prezentowanych na skali stosunkowej normowania można dokonać za pomocą jednego z poniższych wzorów:

1. $z_{ij} = \dfrac{x_{ij} - \bar{x}_j}{s_j}$ /2.1/

2. $z_{ij} = \dfrac{x_{ij} - \bar{x}_j}{r_j}$ /2.2/

3. $z_{ij} = \dfrac{x_{ij}}{s_j}$ /2.3/

4. $z_{ij} = \dfrac{x_{ij}}{r_j}$ /2.4/

5. $z_{ij} = \dfrac{x_{ij}}{\max\limits_i x_{ij}}$ /2.5/

6. $z_{ij} = \dfrac{x_{ij}}{\min\limits_i x_{ij}}$ /2.6/

7. $z_{ij} = \dfrac{x_{ij}}{\bar{x}_j}$ /2.7/

8. $z_{ij} = \dfrac{x_{ij}}{x_{1j}}$ /2.8/

9. $z_{ij} = \dfrac{x_{ij} - \min\limits_i x_{ij}}{r_j}$ /2.9/

10. $z_{ij} = \dfrac{x_{ij}}{\sqrt{\sum\limits_{t=1}^{n} x_{tj}^2}}$ /2.10/

gdzie:

z_{ij} - wartość j-tej zmiennej w i-tym obiekcie po unormowaniu

$\bar{x}_j = \dfrac{1}{n} \sum\limits_{i=1}^{n} x_{ij}$

$$s_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} /x_{ij} - \bar{x}_j/^2}$$

$$r_j = \max_i x_{ij} - \min_i x_{ij}$$

x_{1j} - wartość j-tej zmiennej w dowolnie wybranym obiekcie.

Poniżej przedstawiamy zestawienie, porównujące kilka charakterystyk powyższych sposobów normowania zmiennych.

Sposób normowania	Średnia po unormow.	Odchylenie st. po unormow.	Rozstęp po unormow.	Różnice między dwoma wartościami cech po unormow.
/2.1/	0	1	$\frac{r_j}{s_j}$	$\frac{x_{ij} - x_{kj}}{s_j}$
/2.2/	0	$\frac{s_j}{r_j}$	1	$\frac{x_{ij} - x_{kj}}{r_j}$
/2.3/	$\frac{\bar{x}_j}{s_j}$	1	$\frac{r_j}{s_j}$	$\frac{x_{ij} - x_{kj}}{s_j}$
/2.4/	$\frac{\bar{x}_j}{r_j}$	$\frac{s_j}{r_j}$	1	$\frac{x_{ij} - x_{kj}}{r_j}$
/2.5/	$\frac{\bar{x}_j}{\max_i x_{ij}}$	$\frac{s_j}{\max_i x_{ij}}$	$\frac{r_j}{\max x_{ij}}$	$\frac{x_{ij} - x_{kj}}{\max_i x_{ij}}$
/2.6/	$\frac{\bar{x}_j}{\min_i x_{ij}}$	$\frac{s_j}{\min_i x_{ij}}$	$\frac{r_j}{\min_i x_{ij}}$	$\frac{x_{ij} - x_{kj}}{\min_i x_{ij}}$
/2.7/	1	$\frac{s_j}{\bar{x}_j}$	$\frac{r_j}{\bar{x}_j}$	$\frac{x_{ij} - x_{kj}}{\bar{x}_j}$
/2.8/	$\frac{\bar{x}_j}{x_{1j}}$	$\frac{s_j}{x_{1j}}$	$\frac{r_j}{x_{1j}}$	$\frac{x_{ij} - x_{kj}}{x_{1j}}$
/2.9/	$\frac{\bar{x}_j - \min_i x_{ij}}{r_j}$	$\frac{s_j}{r_j}$	1	$\frac{x_{ij} - x_{kj}}{r_j}$
/2.10/	$\frac{\bar{x}_j}{\sqrt{\sum_{t=1}^{n} x_{tj}^2}}$	$\frac{s_j}{\sqrt{\sum_{t=1}^{n} x_{tj}^2}}$	$\frac{r_j}{\sqrt{\sum_{t=1}^{n} x_{tj}^2}}$	$\frac{x_{ij} - x_{kj}}{\sqrt{\sum_{t=1}^{n} x_{tj}^2}}$

Analiza powyższego zestawienia pozwala stwierdzić, iż część propozycji powoduje ujednolicenie wszystkich zmiennych nie tylko pod względem rzędu wielkości, ale również pod względem zmienności. Dotyczy to propozycji /2.1/, /2.2/, /2.3/, /2.4/ i /2.9/, w których rozstęp lub odchylenie standardowe zmiennych mających przed unormowaniem różną zmienność są po unormowaniu takie same. Oznacza to wyeliminowanie zmienności zmiennych jako podstawy różnicowania obiektów. Z tego powodu godne uwagi wydają się być propozycje pozostałe, a zwłaszcza /2.7/.

2.2.1. Podobieństwo i odległość.

Kluczową rolę w badaniach za pomocą metod analizy wielowymiarowej pełni pojęcie podobieństwa. Pojęcie podobieństwa umożliwia mierzenie relacji między dowolną parą obiektów. W praktyce, podobieństwo może być mierzone przez funkcję podobieństwa lub funkcję odległości.

Funkcja podobieństwa jest to odwzorowanie:

$\bar{s} : P \times P \rightarrow R$

spełniające następujące warunki:

$\forall\, i,k = 1,2,\ldots, n$

1. $0 \leqslant \bar{s}/P_i, P_k/ \leqslant 1$
2. $\bar{s}/P_i, P_i/ = 1$
3. $P_i \neq P_k \Rightarrow \bar{s}/P_i, P_k/ < 1$
4. $\bar{s}/P_i, P_k/ = \bar{s}/P_k, P_i/$

Powyżej przedstawiona funkcja podobieństwa umożliwia badanie relacji między obiektami. Podobnie mówić można o funkcji podobieństwa zmiennych /z tym, że wtedy: $\bar{s} : X \times X \rightarrow R$/.

Funkcję podobieństwa obiektów /zmiennych/ można interpretować jako miarę podobieństwa obiektów /zmiennych/ w sensie wartości zmiennych zaobserwowanych w tych obiektach. Podobieństwo jest tym większe, im większa jest wartość funkcji podobieństwa. Podobne informacje niesie również funkcja odległości.

Jest to odwzorowanie:

$d : P \times P \to R$

spełniające następujące warunki:

$\forall\, i,k = 1,2,\dots, n$

1. $d/P_i, P_k/ \geqslant 0$
2. $d/P_i, P_i/ = 0$
3. $P_i \neq P_k \Rightarrow d/P_i, P_k/ > 0$
4. $d/P_i, P_k/ = d/P_k, P_i/$

Czasem dodaje się jeszcze jeden warunek, tzw. nierówność trójkąta

$d/P_i, P_k/ \leqslant d/P_i, P_j/ + d/P_j, P_k/$

Do badania relacji między zmiennymi służy analogicznie zdefiniowana funkcja odległości zmiennych /z tym, że wtedy $d: X \times X \to R$/.

Interpretacja funkcji odległości jest odwrotna do funkcji podobieństwa. Podobieństwo jest tym większe, im mniejsza jest wartość funkcji odległości.

Wg powyższych definicji funkcja podobieństwa przybiera odległości z przedziału $\langle 0; 1 \rangle$, a funkcja odległości z przedziału $\langle 0; +\infty)$. Obie te funkcje można stosować do mierzenia podobieństwa między obiektami i między zmiennymi.

Mając daną postać lub wartości funkcji podobieństwa, można w sposób jednoznaczny znaleźć postać lub wartości funkcji odległości i vice versa. Do tego celu wystarczy znać postać funkcji transformującej:

$$t : \langle 0; 1\rangle \to \langle 0; +\infty)$$

lub

$$t^{-1} : \langle 0; +\infty) \to \langle 0; 1\rangle$$

Do celu transformacji najbardziej przydatne mogą być funkcje:

1. $t/S/P_i, P_k// = -\log s\,/P_i, P_k/$

$$t^{-1}\,/d/P_i, P_k// = 10^{-d/P_i, P_k/}$$

2. $t/s/P_i, P_k// = \dfrac{1 - s\,/P_i, P_k/}{s\,/P_i, P_k/}$ $\qquad s\,/P_i, P_k/ \neq 0$

$$t^{-1}\,/d/P_i, P_k// = \frac{1}{1 + d/P_i, P_k/}$$

Inne propozycje funkcji transformujących można znaleźć w [136].

Mówiąc o podobieństwie obiektów /lub zmiennych/ w sensie wartości zaobserwowanych można rozpatrywać 2 rodzaje podobieństwa /mówi o tym np. Boyce w [19]:

- podobieństwo wielkości /ang. size similarity/,
- podobieństwo kształtu czyli struktury /ang. shape similarity/.

W pracy zajmować będziemy się podobieństwem wielkości. 2 obiekty /2 zmienne/ są podobne - w sensie podobieństwa wielkości - gdy posiadają zbliżone wartości zmiennych.

Natomiast 2 obiekty /2 zmienne/ są podobne - w sensie podobieństwa kształtu - gdy posiadają zbliżone proporcje wartości zmiennych. Podobieństwo kształtu rozważa się zwłaszcza w odniesieniu do zmiennych. Mówi się wtedy o zależności /korelacji/ zmiennych, stosuje najpopularniejszą miarę - współczynnik korelacji Pearsona.

W zależności od rodzajów zmiennych można stosować różne rodzaje funkcji podobieństwa lub odległości.

W pracy zajmować będziemy się zmiennymi ciągłymi mierzonymi na skali stosunkowej. W przypadku tych zmiennych do badania podobieństwa obiektów stosuje się tradycyjnie funkcję odległości. Poniżej przedstawione zostaną najczęściej stosowane funkcje odległości dla tego przypadku.

Dla innych rodzajów zmiennych postacie funkcji podobieństwa lub odległości można znaleźć w [1],[4] ,[7] ,[11],[19],[26], [29],[47],[54],[60] ,[70],[81],[109],[114],[116],[121],[125],[133], [136],[156],

Funkcje odległości obiektów dla zmiennych ciągłych prezentowanych na skali interwałowej lub stosunkowej są następujące /są to funkcje najczęściej stosowane/:

1. $$d/P_i,P_k/ = \frac{1}{m} \sum_{j=1}^{m} /x_{ij} - x_{kj}/^2 \qquad /2.11/$$

2. $$d/P_i,P_k/ = \frac{1}{m} \sum_{j=1}^{m} |x_{ij} - x_{kj}| \qquad /2.12/$$

3. $$d/P_i,P_k/ = \frac{1}{m} \sum_{j=1}^{m} /\frac{x_{ij} - x_{kj}}{x_{ij} + x_{kj}}/^2 \qquad /2.13/$$

4. $$d/P_i,P_k/ = \frac{1}{m} \sum_{j=1}^{m} \frac{|x_{ij} - x_{kj}|}{|x_{ij}|+|x_{kj}|} \qquad /2.14/$$

5. $$d/P_i,P_k/ = \frac{1}{m} \frac{\sum_{j=1}^{m} |x_{ij} - x_{kj}|}{\sum_{j=1}^{m} |x_{ij} + x_{kj}|} \qquad /2.15/$$

6. $$d/P_i,P_k/ = \max_{j} |x_{ij} - x_{kj}| \qquad /2.16/$$

7. $$d/P_i,P_k/ = \frac{1}{m} \sum_{j=1}^{m} \sum_{l=1}^{m} c_{jl}^{-1} /x_{ij}-x_{kj}/ /x_{il}-x_{kl}/ \qquad /2.17/$$

gdzie:

m - ilość zmiennych

c_{jl}^{-1} - element na przecięciu j-tego wiersza i l-tej kolumny macierzy odwrotnej do macierzy wariancji - kowariancji zmiennych

Inne przykłady miar odległości można znaleźć np. w [1],[29],[136]. Spośród powyżej przedstawionych funkcji odległości szczególnie godne uwagi /i najpowszechniej stosowane/ są odległość euklidesowa /2.11/ oraz odległość Hamminga /2.12/. Jeśli chodzi o propozycje /2.13/, /2.14/ oraz /2.15/ to przydatne mogą być przy zmiennych nienormowanych. Propozycja /2.17/ jest pewnym uogólnieniem odległości euklidesowej. Niezbyt godna polecenia jest propozycja /2.16/, gdyż opiera się tylko na jednej zmiennej, tej, dla której różnica wartości w rozpatrywanych obiektach jest największa.

W badaniach ekonomiczno-społecznych spotykane są również przypadki, gdy rozpatrywany problem opisany jest zmiennymi różnych rodzajów, występuje problem określenia funkcji podobieństwa lub funkcji odległości. Problem ten, jak dotąd, nie doczekał się zadowalających rozwiązań. Spośród propozycji występujących w tym zakresie, można wymienić dwie:

1. Sprowadzenie wszystkich zmiennych różnych rodzajów do zmiennych jednego rodzaju /por. [81], [136]/.

   Po sprowadzeniu do zmiennych jednego rodzaju można wyznaczyć wartości funkcji podobieństwa lub odległości. Niemniej jednak, wątpliwości budzi sama idea sprowadzania zmiennych. Zwykle to sprowadzanie odbywa się ze skal o większej zawartości informacji /jak interwałowa czy stosunkowa/ do skal o mniejszej zawartości informacji /jak nominalna czy porządkowa/. Traci się przez to pewną część informacji.

2. Wyznaczenie funkcji podobieństwa /lub odległości/ będącej pewną wypadkową funkcji podobieństwa /lub odległości/ dla zmiennych różnych rodzajów /por. [60]/.

2.3. Metody i zadania analizy wielowymiarowej - próba systematyzacji i krótka charakterystyka.

Aparat analizy wielowymiarowej jest bardzo szeroki i niejednorodny. Literatura przedmiotu nie doczekała się w zasadzie wielu propozycji odnośnie systematyzacji metod i zagadnień rozwiązywanych za pomocą metod analizy wielowymiarowej. Jest to zresztą problem bardzo trudny. Wspomnieć można jedynie o propozycji, zamieszczonej w [11]. Jej autorzy dzielą ogół metod analizy wielowymiarowej z punktu widzenia charakteru problemu czy zadania, które może być za pomocą tych metod rozwiązane.

I tak, wyróżniają:

- metody opisu /np. analiza czynnikowa/,
- metody strukturalizacji /np. taksonomia numeryczna/,
- metody wyjaśniania /np. analiza kanoniczna/.

Niniejszy punkt poświęcony będzie syntetycznej prezentacji podstawowych zadań i metod analizy wielowymiarowej. Ujęcie poniższe nie pretenduje do miana jedynego, czy najlepszego. Nie zawiera ono być może wszystkich metod analizy wielowymiarowej, lecz metody podstawowe, najważniejsze, najczęściej spotykane, zarówno w opracowaniach teoretycznych, jak i badaniach empirycznych.

Ogół metod analizy wielowymiarowej można podzielić na dwie podstawowe grupy:

- metody badania związków między zmiennymi,
- metody badania zbioru obiektów.

W pierwszej grupie metod pojęciem pierwotnym /p. 2.2.1/ jest zmienna, opisana przez zbiór wartości przyjętych w n obiektach.

Do grupy tej należą:

- metody czynnikowe,
- analiza kanoniczna,
- analiza regresji,
- analiza taksonomiczna zmiennych.

W drugiej grupie metod pojęciem pierwotnym jest obiekt, charakteryzowany przez wartości m zmiennych. Do grupy tej należą:

- metody porządkowania liniowego obiektów,
- analiza dyskryminacyjna,
- metody klasyfikacji.

Dokonamy obecnie syntetycznej prezentacji tych grup metod, przy czym ze względu na problematykę pracy nieco więcej miejsca poświęcimy metodom klasyfikacji, a w odniesieniu do pozostałych grup metod przedstawimy jedynie zagadnienie, które może być za pomocą tych metod realizowane.

2.3.1. Metody badania związków między zmiennymi.

Metody czynnikowe.

Metody te występują w dwóch postaciach, jako:

- analiza czynnikowa,
- analiza głównych składowych.

Analiza czynnikowa.

Zagadnienie rozwiązywane za pomocą metod analizy czynnikowej jest następujące:

Dane jest m zmiennych: $X_1, X_2, \ldots, X_m$, Zbiór ten należy przekształcić w zbiór k/k $<$ n/ zmiennych /zwanych czynnikami/: $F_1, F_2, \ldots, F_k$, tak, aby zaochodziły warunki:

1. $X_j = a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jk}F_k + d_jU_j \qquad j = 1,2,\ldots,m$

2. $\mathrm{cor}\ /F_i, F_j/ = 0 \qquad i \neq j,\ i,j = 1,2,\ldots,k, \qquad \mathrm{cor}\ /U_i, F_j/ = 0$

$i = 1,\ldots,m$

gdzie: $j = 1,\ldots,k$

cor $/F_i, F_j/$ - współczynnik korelacji między zmienną F_i i F_j, natomiast U_j jest zmienną zwaną czynnikiem swoistym.

Zadaniem metod analizy czynnikowej jest wyznaczenie wartości współczynników $a_{11}, a_{12}, \ldots, a_{1k}, a_{21}, a_{22}, \ldots, a_{2k}, \ldots, a_{m1}, a_{m2}, \ldots, a_{mk}$.

Zagadnienie rozwiązywane za pomocą metod analizy czynnikowej sprowadza się do uproszczenia zbioru zmiennych za pomocą zbiorru nieskorelowanych zmiennych o mniejszej liczebności. Zmienne te nazywane są czynnikami wspólnymi i wyjaśniają pewną część zmienności każdej ze zmiennych. Udział j-tego czynnika wspólnego w wyjaśnieniu zmienności i-tej zmiennej określany jest przez współczynnik a_{ij}. Pozostała część zmienności każdej zmiennej wyjaśniana jest przez zmienną zwaną czynnikiem swoistym /swoistym dla każdej zmiennej X_i/.

Możliwości rozwiązań powyższego zagadnienia jest wiele. Spośród najczęściej stosowanych wyróżnić należy tzw. metodę czynników głównych oraz metodę centroidalną.

Najbardziej wyczerpującą pracą z tej dziedziny jest monografia

[64] natomiast bardzo poglądowe są prace [101],[107],[114]. Oprócz tego problemy teoretyczne, a także ciekawsze badania empiryczne za pomocą analizy czynnikowej, a także analizy głównych składowych można znaleźć w [5],[7],[25],[26],[27],[31],[32], [34],[58],[66],[69],[77],[90],[99],[100],[108],[116].

Analiza głównych składowych.

Analiza głównych składowych jest w konkretnych, metodologicznych rozwiązaniach podobna do analizy czynnikowej, bowiem zagadnienie analizy głównych składowych rozwiązuje się w zasadzie za pomocą metody głównych czynników. W swej istocie jest to jednak odrębna grupa metod /a ściślej, metoda/.

Zagadnienie analizy głównych składowych jest następujące:

Dany jest zbiór m zmiennych: $X_1, X_2, \ldots, X_m$. Zbiór ten należy transformować w inny zbiór m zmiennych: $F_1, F_2, \ldots, F_m$ /zwanych składowymi/ tak, aby spełnione były warunki:

1. $X_j = a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jm}F_m, \qquad j = 1, \ldots m$
2. $\operatorname{cor}(F_i, F_j) = 0 \qquad i \neq j\ , \quad i, j = 1, 2, \ldots, m$

Jak widać, w przypadku stosowania metod analizy czynnikowej zbiór zmiennych jest przekształcany na zbiór nieskorelowanych /ortogonalnych/ czynników o mniejszej /z reguły dużo/ liczebności, natomiast w przypadku analizy głównych składowych zbiór zmiennych zostaje jedynie ortogonalnie transformowany na zbiór składowych o tej samej liczebności.

Dodać należy, iż nie wszyscy badacze zdają sobie z tego sprawę, utożsamiając analizę głównych składowych z analizą czynnikową, a składowe z czynnikami. Najbardziej konsekwentnie zróżnicowanie to jest przeprowadzone w pracach: [25],[64].

W literaturze istnieje wiele propozycji odnośnie analizy głównych składowych, wspomniano o nich wyżej.

Analiza taksonomiczna.

Nazwę tę przyjęto w pracy dla grupy metod realizujących następujące zagadnienie:

Dany jest zbiór m zmiennych: $X_1, X_2, \ldots, X_m$. Ze zbioru tego należy wybrać k zmiennych /k < n/, które spełniają następujące warunki:

1. są słabo skorelowane między sobą,
2. są silnie skorelowane ze zmiennymi, które nie zostały wybrane,
3. charakteryzują się dużą zmiennością.

Przez niektórych badaczy zagadnienie to nazywane jest doborem zmiennych diagnostycznych /por. [114],[116]/.

Cel stosowania metod analizy taksonomicznej jest podobny jak analizy czynnikowej. I w tym przypadku chodzi o zmniejszenie liczebności zbioru zmiennych, tak, aby strata informacji była jak najmniejsza, tzn. aby wybrane zmienne były jak najlepszymi reprezentantami całego zbioru zmiennych. Z tym, że w analizie czynnikowej dokonuje się tego przez przekształcenie zbioru zmiennych w zbiór innych zmiennych /czynników/, natomiast w analizie taksonomicznej przez wybór części zmiennych. Niektóre z metod analizy taksonomicznej realizują powyższe zagadnienie dwuetapowo: jako klasyfikację zmiennych /tzn. podział zbioru zmiennych na klasy zmiennych "podobnych" - zwykle w sensie współczynnika korelacji - można tu wykorzystać większość metod klasyfikacji, o których mowa w 3.2/, a następnie wybór reprezentantów /tzn. z każdej klasy zmiennych podobnych wybór jednej, najlepszego reprezentanta klasy/.

Propozycje metod analizy taksonomicznej, a także rozważania metodologiczne na ich temat można znaleźć w pracach [11],[28], [56],[57],[65],[66],[110],[112],[114].

Przedstawiono powyżej dwie grupy metod, tzn. analiza taksonomiczna oraz metody czynnikowe, bazują na zależnościach "wewnątrz" zbioru zmiennych,
Poniżej zaprezentujemy dwie grupy metod, które bazują na zależnościach "między" dwoma zbiorami zmiennych /przy czym w przypadku analizy regresji, jeden z tych zbiorów jest jednoelementowy/.

Analiza regresji.

Dany jest zbiór m zmiennych: $X_1, X_2, \ldots, X_m$ oraz zmienna Y, na którą każda ze zmiennych $X_1, X_2, \ldots, X_m$ ma wpływ. Należy znaleźć kombinację liniową zmiennych:

$$a_1X_1 + a_2X_2 + \ldots + a_mX_m + a_o$$

tak, aby:

$\sum_{i=1}^{n} e_i^2$ była jak najmniejsza, gdzie:

$$e_i = y_i - a_1x_{i1} - a_2x_{i2} - \ldots - a_mx_{im} - a_0$$

Metody analizy regresji są najczęściej stosowanymi metodami analizy wielowymiarowej. W ujęciu stochastycznym metody te wykorzystywane są przy budowie modeli ekonometrycznych. Ujęcie deterministyczne, czyli tzw. regresja opisowa, przedstawione są m.in. w pracach [11],[53].

Analiza kanoniczna

Zagadnienie analizy kanonicznej jest następujące:
Dane są 2 zbiory zmiennych: $X = \{X_1, X_2, \ldots, X_m\}$ oraz $Y = \{Y_1, Y_2,$

$\ldots, Y_l\}$. Zbiór X należy sekwencyjnie transformować w zbiór $X' = \{X'_1, X'_2 \ldots, X_k\}$ /k < m/, natomiast zbiór Y w $Y' = \{Y'_1, Y'_2, \ldots, Y'_k\}$ /k < l /, tak, aby spełnione były warunki:

1. $X'_j = a_{j1}X_1 + a_{j2}X_2 + \ldots + a_{jm}X_m \qquad j = 1,2,\ldots, k$

2. $Y'_j = b_{j1}Y_1 + b_{j2}Y_2 + \ldots + b_{jm}Y_m \qquad j = 1,2,\ldots, k$

3. $\mathrm{cor}\ /X'_i, X'_j/ = 0 \qquad i \neq j,\ i,j = 1,2,\ldots, k$

4. $\mathrm{cor}\ /Y'_i, Y'_j/ = 0 \qquad i \neq j,\ i,j = 1,2,\ldots, k$

5. $\mathrm{cor}\ /X'_i, Y'_i/ = \max\limits_{t_i} \{ \mathrm{cor}\ /X'_{t_i}, Y'_{t_i}/ \}$

Zmienne X'_i i Y'_i /i = 1,2,...,k/ nazywane są zmiennymi kanonicznymi związanymi ze zmiennymi X_i i Y_i. Zbiory zmiennych zostają przekształcone w inne zbiory, tak aby powiązania między tymi zbiorami były jak największe /współczynniki korelacji między zmiennymi tych dwóch zbiorów były jak największe/. Metody analizy kanonicznej przedstawione są w pracach [5], [11], [77], [100].

2.3.2. Metody badania zbioru obiektów.

Ze względu na tematykę pracy, przy prezentacji metod badania zbioru obiektów większa uwaga poświęcona będzie analizie dyskryminacyjnej i metodom klasyfikacji /idea ich jest w pewnym sensie podobna/.

Metody porządkowania liniowego zbioru obiektów.

Zagadnienie realizowane przez metody porządkowania liniowego zbioru obiektów, przedstawić można w sposób następujący: W danym zbiorze obiektów, opisanych za pomocą m zmiennych, należy wprowadzić relację porządku liniowego /bazując na wartościach zmiennych/. Relacja ta powinna być zwrotna, antysymetryczna, przechodnia i spójna. Mówiąc innymi słowy, obiekty należy " uszeregować " ze względu na wartości zmiennych. Zaznaczyć wypada, iż metody te mogą funkcjonować w zasadzie wtedy, gdy wszystkie zmienne będą stymulantami lub destymulantami /por. [65] / tzn. wzrost lub odpowiednio spadek wartości zmiennej świadczy o wyższej " randze " obiektu/. Metod tych jest wiele. Można wspomnieć np. o metodzie wzorca rozwoju /por. [65] /, rangowej /por. [66] /, antywzorca rozwoju /por. [112] /. Inne metody tej grupy zaprezentowane są w pracach [28] , [58] .

Analiza dyskryminacyjna i metody klasyfikacji.

Jako ostatnia grupa metod zostaną omówione metody analizy dyskryminacyjnej i metody klasyfikacji. M. Kendall tak charakteryzuje dyskryminację i klasyfikację:
" Dyskryminacja. Dane są dwie próbki pochodzące z dwóch różnych populacji. Dla każdego obiektu dane są wartości p zmiennych. Należy skonstruować metodę przydzielania nowego /nie należącego do tych dwóch próbek - K.J/. obiektu do właściwej populacji na bazie wartości p zmiennych, które przyjmuje ten obiekt.

Klasyfikacja. Dana jest próbka obiektów z tej samej lub różnych populacji. Dla każdego obiektu dane są wartości p zmiennych. Należy skonstruować metodę dzielącą obiekty na grupy." /por. [76], s.165/

W ujęciu prezentowanym w pracy /patrz 2.1./, a więc deterministycznym, można oba te pojęcia zdefiniować następująco:

Dyskryminacja. Dana jest k skupień obiektów: $S_1, S_2, \ldots, S_k$ o liczebnościach $n_1, n_2, \ldots, n_k$. Dla każdego obiektu $P_i \in P$ należy wyznaczyć klasę S_j /$1 \leq j \leq k$/ tak, aby po przydzieleniu do klas wszystkich n obiektów należących do zbioru n obiektów P, spełnione były warunki:

- obiekty znajdujące się w tych samych klasach były jak najbardziej podobne;
- obiekty znajdujące się w różnych klasach były jak najmniej podobne.

Klasyfikacja. Dany jest zbiór n obiektów. Zbiór ten należy podzielić na k klas /przy czym liczba k jest dana lub nie/ tak, aby spełnione były warunki:

- obiekty znajdujące się w tych samych klasach były jak najbardziej podobne;
- obiekty znajdujące się w różnych klasach były jak najmniej podobne.

Jak zatem widać, w przypadku dyskryminacji podziału obiektów na klasy dokonuje się na bazie istniejących już klas, które stanowią pewien wzór podziału. W przypadku klasyfikacji takiego " wzoru " nie ma.

Pojęcia dyskryminacji i klasyfikacji można również interpretować na gruncie teorii rozpoznawania obrazów /ang. pattern recognition, pattern classification/. Ta dyscyplina naukowa wyrosła na bazie automatyki. W ostatnich latach jej algorytmy, a także zastosowanie, również w naukach społecznych, rozwinęły się burzliwie.

Zadanie rozpoznawania obrazów składa się z dwóch części. Pierwsza część, tzw. uczenie, polega na tym, że " nauczyciel " przekazuje " uczniowi " pewne informacje. Dotyczą one " obrazów ", do jakich należą " obiekty ". Przez " obraz " należy rozumieć klasę obiektów. Informacje te polegają na tym, że dla każdego " obiektu " nauczyciel wskazuje " obraz ", do którego należy ten " obiekt ". Druga część, tzw. rozpoznawanie lub egzamin, polega na tym, że " uczeń " korzystając z informacji przekazanych mu w pierwszej części stara się dla każdego z nowo rozpatrywanych " obiektów " rozpoznać " obraz ", do którego dany " obiekt " należy /stąd nazwa teorii/.

Przykładów rozpoznawania obrazów jest wiele. W klasycznym przypadku, zaczerpniętym z automatyki - " uczniem " jest maszyna cyfrowa, " nauczycielem " człowiek wprowadzający dane do maszyny, " obiektami " - litery napisane " obrazami " - - nazwy tych liter.

W życiu codziennym z rozpoznawaniem obrazów spotykamy się np. przy nauce czytania przez dzieci. " Uczniem " jest dziecko, " nauczycielem "- nauczyciel, starszy brat, mama, babcia, itd., " obiektami " - litery, " obrazami "- nazwy liter.

W teorii rozpoznawania obrazów spotyka się również tzw. " rozpoznawanie bez nauczyciela ". Składa się ono tylko z drugiej części - rozpoznawania, nie uwzględnia uczenia, tzn. " uczeń " nie posiada żadnych informacji o " obrazach ".

Analiza dyskryminacyjna jest przykładem rozpoznawania obrazów. " Nauczycielem ", a właściwie informacjami " nauczyciela " są w tym przypadku informacje , dotyczące klas S_1, S_2,..., S_k

" uczniem " - dokonujący dyskryminacji, " obrazami " klasy, a " obiektami " - zbiór P, liczący n obiektów. Analiza dyskryminacyjna obejmuje w zasadzie tylko drugą część procesu rozpoznawania.

Natomiast metody klasyfikacji to przykład rozpoznawania obrazów " bez nauczyciela ", gdyż nie ma w tym przypadku żadnych informacji o klasach. " Uczniem " jest dokonujący klasyfikacji, " obrazami " klasy, a " obiektami " - zbiór P, liczący n obiektów. Metody klasyfikacji obejmują zatem drugą część procesu rozpoznawania, a poza tym jest to rozpoznawanie bez nauczyciela.

Jak widać, rozpoznawanie " bez nauczyciela " jest bez porównania trudniejsze.

W pracy zajmować się będziemy jedynie metodami klasyfikacji.

W badaniach ekonomiczno - społecznych rozpoznawanie " bez nauczyciela " jest spotykane znacznie częściej.

Najciekawsze prace traktujące o analizie dyskryminacyjnej są to [11], [39], [76], [81], natomiast o zagadnieniach teorii rozpoznawania obrazów traktują [2], [6], [39], [52], [55], [71], [103], [125].

Bogatej literatury doczekały się również metody klasyfikacji.

W pracy przez metody klasyfikacji rozumie się również te metody ilościowe, które są określane jako:

- klasyfikacja lub klasyfikacja automatyczna /fr.classification automatique, niem. automatische Classification, ros. awtomaticzeskaja kłassifikacija/, por.[1], [29], [39], [76], [97], [99], [136], [156].
- analiza skupień /ang.cluster analysis, clustering, niem.

Cluster - Analyse/, por. [16], [40], [54], [72], [73], [79], [84], [96], [120], [132].

- taksonomia numeryczna /ang. numerical taxonomy, mathematical taxonomy, fr. taxinomie/ por. [4], [7], [49], [50], [70], [114], [133], [134], [138].
- grupowanie, por. [17], [30], [51], [71], [144].
- analiza typu Q, por. [26], [107], [108].

Ta różnorodność terminologiczna wynika z faktu, iż metody klasyfikacji tworzone i stosowane są przez przedstawicieli różnych dziedzin /np. biologowie, antropologowie używają zwrotu taksonomia numeryczna, psychologowie - analiza typu Q/.

3. METODY KLASYFIKACJI I REGIONALIZACJI.

3.1. Uwagi ogólne o metodach klasyfikacji i regionalizacji.

Rozdział trzeci poświęcony będzie teoretycznym problemom jednej z grup metod analizy wielowymiarowej, mianowicie metodom klasyfikacji, a także metodom regionalizacji, stanowiącym, jak pokażemy, pewne rozszerzenie metod klasyfikacji. Na wstępie zostaną zdefiniowane zagadnienia klasyfikacji i zagadnienie regionalizacji.

Zagadnienie klasyfikacji.

Dany jest zbiór P, liczący n obiektów: $P_1, P_2, \ldots, P_n$. Każdy z tych obiektów charakteryzowany jest przez wartości m zmiennych: $X_1, X_2, \ldots, X_m$. Zbiór ten należy podzielić na k / przy czym liczba k jest dana lub nieznana / podzbiorów, tzw. klas: $S_1, S_2, \ldots, S_k$, tak, aby spełnione były warunki:

1. $S_1 \cup S_2 \cup \ldots \cup S_k = P$
2. $S_i \cap S_j = \emptyset$, $i \neq j$, $i,j = 1,2,\ldots,k$
3. Obiekty znajdujące się w tych samych klasach są jak najbardziej podobne.
4. Obiekty znajdujące się w różnych klasach są jak najmniej podobne.

Zauważmy, iż kluczowymi dla rozwiązania zagadnienia klasyfikacji warunkami są warunki 3 i 4. Zgodnie z ich treścią, klasyfikacja bazuje na wzajemnym podobieństwie obiektów, przy czym przez obiekty podobne rozumie się tu obiekty o zbliżonych wartościach opisujących je zmiennych. Warunki te, w przeciwieństwie do dwóch pierwszych, określone są w sposób mniej formalny i ścisły. Dopuszcza to możliwość stosowania różnych rozwiązań w zakresie metod klasyfikacji.

Słów kilka poświęcimy bliższemu sprecyzowaniu pojęcia klasy. W literaturze spotyka się wiele nazw, jak również definicji - klasa, skupienie, podzbiór, typ, w literaturze anglosaskiej powszechny jest termin cluster /jest to określenie trudno przetłumaczalne, oznacza "kiść", często " cluster " tłumaczy się jako " skupienie "/.

Spora różnorodność występuje w zakresie definicji klasy. Najczęściej spotykane są następujące definicje /por. [58]/:

1. taka zbiorowość, że odległość między dowolną parą obiektów z tej zbiorowości jest mniejsza niż między jakąkolwiek parą z różnych klas,
2. obszar o dużej gęstości przedzielony obszarem o małej gęstości,
3. taka zbiorowość obiektów, do których obiekty najbardziej podobne są w tej samej klasie.

Formalnie najbardziej poprawna jest definicja 1, niemniej jednak dla celów praktycznych nie zawsze jest użyteczna. Dotyczy to często spotykanych sytuacji, gdy nie występuje klasyfikacja naturalna, tzn. gdy nie ma wyraźnych granic między klasami. Wtedy zgodność z definicją zapewnia tylko klasyfikacja składająca się z dużej ilości klas o małej liczebności. Z kolei definicja 2 ze względu na swą nieprecyzyjność jest również mało przydatna dla celów praktycznych, natomiast definicja 3 wyklucza całkowicie możliwość występowania klas jednoelementowych.

Pewnym rozszerzeniem problemu klasyfikacji jest zagadnienie regionalizacji.

Zagadnienie regionalizacji.

Dany jest zbiór P, liczący n obiektów: $P_1, P_2, \ldots, P_n$. Każdy z tych obiektów charakteryzowany jest przez wartości m zmiennych: $X_1, X_2, \ldots, X_m$. Ponadto dane są wartości tzw. funkcji przyległości /sąsiedztwa/ obiektów:

$$c : P \times P \rightarrow \{0;1\}$$

gdzie:

$$c\,/P_i, P_j/ = \begin{cases} 1, \text{gdy obiekty } P_i \text{ i } P_j \text{ sąsiadują /przylegają do siebie/;} \\ 0, \text{w przeciwnym przypadku} \end{cases}$$

Zbiór P należy podzielić na k /gdzie liczba k jest dana lub nieznana/ podzbiorów tzw. regionów $S_1, S_2, \ldots, S_k$, tak aby spełnione były warunki:

1. $S_1 \cup S_2 \cup \ldots \cup S_k = P$
2. $S_i \cap S_j = \emptyset \quad i \neq j, \; i,j = 1,2,\ldots, k$
3. Obiekty znajdujące się w tych samych regionach są jak najbardziej podobne.
4. Obiekty znajdujące się w różnych regionach są jak najmniej podobne.
5. $\forall P_i \in S_j$ $/j = 1,2,\ldots,k/$ zachodzi: $\forall P_k \in S_j \; \exists$ ciąg $/P_i, P_{i_1}, P_{i_2}, \ldots, P_{i_l}, P_k/$ taki że $c\,/P_i, P_{i_1}/ = c\,/P_{i_1}, P_{i_2}/ = \ldots = c\,/P_{i_l}, P_k/ = 1$ lub card $/S_j/ = 1$

Wynika z powyższego określenia, że regionalizacja różni się od klasyfikacji jedynie wprowadzeniem pewnego dodatkowego warunku, wyrażonego przez funkcję przyległości. Warunek ten powoduje, że pewne obiekty nie mogą znaleźć się w tej samej klasie /zwana jest ona regionem/, nawet choćby były bardzo podobne.

Warunek ten został nazwany warunkiem przyległości /a zagadnienie - regionalizacją/ ze względów tradycyjnych, jak również z powodu problematyki pracy. Niemniej jednak zagadnienie regionalizacji można rozumieć szerzej, nie tylko w sensie przestrzennym. Na przykład, na podobnych zasadach oparta jest periodyzacja /por.[101]/. Mamy wtedy do czynienia z obiektem rozpatrywanym w kolejnych jednostkach czasu /a nie jak powyżej z obiektami rozpatrywanymi w jednym momencie czasowym/. Przy periodyzacji, tzn. podziale badanego okresu na podobne podokresy, funkcja przyległości /rozumiana w sensie czasowym/ określana jest następująco:

$$c\,/\,P_i,\,P_k\,/ \equiv \begin{cases} 1, & \text{gdy } |i-k| = 1 \\ 0, & \text{gdy } |i-k| \neq 1 \end{cases}$$

przy czym i oraz k są indeksami jednostek czasu, natomiast P_i oraz P_k oznacza tu obiekt rozpatrywany w jednostkach czasu o indeksach i oraz k.

O zagadnieniu regionalizacji i jej związku z klasyfikacją traktuje m.in. praca [26]. Jej autorzy stwierdzają /por.[26], s.10 /, że " między regionalizacją i klasyfikacją zachodzą istotne podobieństwa, jednak procedury i pojęcia tych dwóch dziedzin nie pozostają w jednoznacznej odpowiedniości.Regionalizacja, ze względu na przedmiot badania i naukowy cel poznania, jest z natury geograficzna i tylko pod względem formalnym utożsamiana jest z klasyfikacją /a ściślej z pewnym szczególnym rodzajem klasyfikacji - przyp. K.J./. Klasyfikacja jako czynność bardziej precyzyjna lub lepiej znana i zorganizowana, teoretycznie może wystąpić w roli modelu teoretycznego regionalizacji. Akceptacja tego poglądu upoważnia do rozważania regionalizacji w terminach klasyfikacji ".

Niech C /n, k/ oznacza ilość możliwych klasyfikacji zbioru n obiektów na k klas.

Zachodzi:

$$C/n,k/ = \left[k^n - \sum_{i=1}^{k-1} /{}^{k}_{i}/\ C/n,i/ \right] : k!$$

Dowód. Zauważmy, że klasyfikacja n obiektów na k klas można traktować w sensie formalnym jako przyporządkowanie każdemu obiektowi klasy. Każdej klasyfikacji odpowiada n - elementowy zbiór uporządkowany /ciąg/ liczb naturalnych z przedziału $\langle 1, k \rangle$. Ilość możliwych przyporządkowań, tzn. ilość możliwych ciągów jest to po prostu ilość wariacji z powtórzeniami n elementowych z k elementów. Jest ona równa jak wiadomo k^n. Zauważmy jednak, że wśród tych ciągów mogą znaleźć się również takie, w których występować będą nie wszystkie liczby naturalne z przedziału $\langle 1, k \rangle$, lecz tylko część z nich. Wtedy jednak mamy do czynienia z klasyfikacją o odpowiednio mniejszej ilości klas /ilość klas jest równa ilości liczb z przedziału $\langle 1, k \rangle$/. Przy czym dla danej ilości liczb, np. k_0, tych ciągów jest dokładnie tyle co ilość przyporządkowań k_0 liczb dla n elementów pomnożona przez $/{}^{k}_{k_0}/$ /bo tyle jest możliwych kombinacji k_0 liczb spośród k liczb - po prostu należy rozważyć każdą możliwą kombinację części liczb z możliwych k liczb/. Dlatego od k^n należy odjąć wszystkie przyporządkowania, w których występuje tylko 1,2,3,..., k - 1 liczb. Jest ich dokładnie

$$k^n - \sum_{i=1}^{k-1} /{}^{k}_{i}/\ C/n,i/$$

Zauważmy ponadto, że rozpatrywaną powyżej ilość przyporządkowań /ciągów/ należy jeszcze podzielić przez ilość możliwych permutacji k elementów. Jest tak, bowiem nie są ważne numery klas, jakie otrzymują poszczególne obiekty, lecz jedynie ich rozdział na klasy.

Dlatego np. ciąg 1 2 3 1 2 oraz 2 3 1 2 3 z punktu widzenia klasyfikacji nie różnią się niczym /klasyfikacja jest następująca: $S = \{ \{s_1, s_4\}, \{s_2, s_5\}, \{s_3\} \}$ /, a " takich samych " klasyfikacji jest dokładnie tyle, ile permutacji k elementów.

Stąd:

$$C/n, k/ = \left[k^n - \sum_{i=1}^{k-1} / {k \atop i} / \, C/n, i/ \right] : k! \quad \text{c. b. d. o.}$$

Zauważmy, że np. dla n = 10:

C /n, 1/ = 1

C /n, 2/ = 511

C /n, 3/ = 18 660

C /n, 4/ = 970 866 itd.

Jak widać, nawet dla zbiorów o stosunkowo małej ilości klas, ilość możliwych klasyfikacji rośnie gwałtownie ze wzrostem k. Powoduje to, iż niemożliwe jest przebadanie wszystkich możliwych klasyfikacji.

3.2. Metody klasyfikacji - charakterystyka.

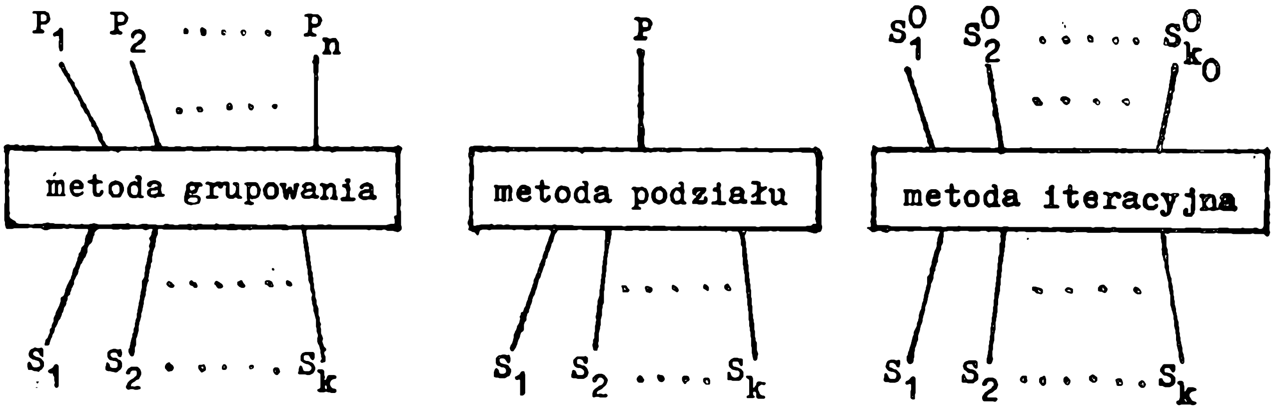
Metod klasyfikacji, tzn. metod, które realizują zdefiniowane powyżej zagadnienie klasyfikacji, jest olbrzymia ilość. W niniejszym punkcie dokonana zostanie próba usystematyzowania najważniejszych i najczęściej stosowanych metod klasyfikacji. W literaturze spotyka się próby systematyzacji tych metod. Do najciekawszych należy zaliczyć: [1], [4], [29]. W pracy głównie zajmiemy się metodami taksonomii numerycznej.

Oprócz tego jeden z punktów dotyczył będzie innych metod czy podejść przydatnych do rozwiązania zagadnienia klasyfikacji /zaliczyć tu można np. niektóre metody programowania matematycznego czy analizę czynnikową/.

Ogół metod taksonomii numerycznej można podzielić na:

- metody grupowania,
- metody podziału,
- metody iteracyjne.

Zasadniczą różnicę między poszczególnymi grupami metod prezentuje poniższy schemat:



Jak widać, w metodach grupowania poszczególne obiekty są łączone /grupowane/ w zbiory obiektów, tworzące klasy; w metodach podziału cały zbiór obiektów jest dzielony na mniejsze podzbiory, tworzące klasy; w metodach iteracyjnych pewne klasyfikacja początkowa S^0 jest w wyniku pewnych procedur przekształcana /"poprawiana"/, aż do otrzymania klasyfikacji końcowej S o tej samej lub innej ilości klas.

Obecnie dokonana zostanie charakterystyka poszczególnych grup metod.

3.2.1. Metody grupowania.

Ogół metod grupowania można podzielić na dwie podstawowe grupy: metody hierarchiczne /sekwencyjne/ i niehierarchiczne.

Hierarchiczne metody grupowania - algorytm:

1. Dana jest klasyfikacja początkowa:

   $S^o = \{ S_1^o, S_2^o, \ldots, S_n^o \}$, gdzie $S_i^o = P_i$ $/i = 1,2,\ldots, n/$

2. W i-tym etapie hierarchicznego grupowania dokonuje się grupowania dwóch klas w jedną. Są to klasy S_j^{i-1} i S_k^{i-1}, które spełniają warunek:

$$d\,/S_j^{i-1}, S_k^{i-1}/ = \min_{\substack{S_p^{i-1},\, S_r^{i-1} \\ p \neq r}} d\,/S_p^{i-1}, S_r^{i-1}/$$

   Po zakończeniu i-tego etapu hierarchicznego grupowania otrzymuje się klasyfikację:

$$S^i = \{ S_1^i, S_2^i, \ldots, S_{n-i}^i \} \quad , \text{ gdzie:}$$

$$S_l^i = S_l^{i-1} \qquad l = 1,2,\ldots, j-1, j+1,\ldots, k-1$$

$$S_j^i = S_j^{i-1} \vee S_k^{i-1}$$

$$S_l^i = S_{l+1}^{i-1} \qquad l = k, k+1,\ldots, n-i$$

3. Po zakończeniu n-1 etapów grupowania hierarchicznego otrzymuje się ciąg klasyfikacji:

$$S^o = \{ S_1^o, S_2^o, \ldots, S_n^o \}$$

$$S^1 = \{ S_1^1, S_2^1, \ldots, S_{n-1}^1 \}$$

$$\ldots\ldots\ldots\ldots\ldots$$

$$S^{n-2} = \{ S_1^{n-2}, S_2^{n-2} \}$$

$$S^{n-1} = \{ S_1^{n-1} \} = P$$

4. Z tego ciągu klasyfikacji wybiera się jedną:
$\{S_1, S_2, \ldots, S_k\}$ /o sposobie wyboru tej klasyfikacji traktuje jeden z dalszych punktów/.

Przedstawiony powyżej algorytm ma zastosowanie dla klasycznej wersji hierarchicznych metod grupowania, zwanej również metodą najbliższej pary zwrotnej /por. [26]/. Oprócz tego wyróżnić można trzy inne warianty hierarchicznych metod grupowania:

- hierarchiczna metoda grupowania - wariant elementarnego połączenia.

Metoda ta różni się od opisanej powyżej tylko w punkcie 2, gdyż w każdym etapie grupowania dokonuje się połączenia każdej klasy S_t^i z klasą S_p^i spełniającą warunek:

$$d\,/\,S_t^i, S_p^i/ = \min_{S_r^i} \quad d\,/S_t^i, S_r^i/$$

Innymi słowy, w każdym etapie, każda klasa zostaje połączona z najbliższą klasą;

- hierarchiczna metoda grupowania - wariant par zwrotnych.

W każdym etapie grupowania zostają połączone parami te klasy S_j^i, S_k^i, dla których spełniony jest warunek:

$$d\,/\,S_j^i, S_k^i\,/ = \min_{S_r^i} \quad d\,/\,S_j^i, S_r^i\,/ = \min_{S_s^i} \quad d\,/\,S_s^i, S_k^i\,/$$

W każdym etapie grupuje się te pary lub klasy, które są wzajemnie dla siebie najbliższe;

- hierarchiczna metoda grupowania - wariant porządku kolejnościowego.

W i-tym etapie hierarchicznego grupowania zostają połączone te klasy S_j^{i-1}, S_k^{i-1}, dla których spełniony jest warunek

$$d\,/\,S_j^{i-1}, S_k^{i-1}\,/ \leqslant d_{i-1}$$

Te trzy warianty różnią się od klasycznego tym, iż w każdym etapie grupowania łączy się więcej niż dwie klasy /charakteryzujące się minimalną odległością/. Powoduje to, iż całość grupowania wszystkich obiektów w zbiór P odbywa się w trakcie mniejszej ilości etapów niż n-1. Skoncentrujmy się na wariancie klasycznym metod grupowania. W zależności od stosowanej miary odległości między klasami wyróżnić można różne hierarchiczne metody grupowania:

1. metoda najbliższego sąsiada/por. [1] ,[4] ,[29] ,[47], [50] , [65] , [73] ,[84] ,[85] , [116] , [133] , [136] ,[137] , [147] , [148] /.

   W metodzie tej:

   $$d\,/S_p^{i-1},\ S_r^{i-1}/ = \min_{\substack{P_s \in S_p^{i-1} \\ P_t \in S_r^{i-1}}} d\,/P_s,\ P_t/$$

Odległość między klasami jest równa najmniejszej odległości między obiektami należącymi do tych dwóch /różnych/ klas.

2. metoda najdalszego sąsiada /por. [1] ,[4] ,[29] , [47] , [73] , [84] , [85] , [133] , [136] , [147] , /.

   W metodzie tej:

   $$d\,/S_p^{i-1},\ S_r^{i-1}/ = \max_{\substack{P_s \in S_p^{i-1} \\ P_t \in S_r^{i-1}}} d\,/P_s,\ P_t/$$

Odległość między dwoma klasami jest równa największej odległości pomiędzy obiektami należącymi do tych dwóch /różnych/ klas.

3. metoda średniej międzygrupowej /por. [1], [4], [9], [21], [26], [29], [47], [84], [85], [73], [133], [136], [147] /.

W metodzie tej /będącej pewnym kompromisem pomiędzy dwoma powyższymi/ odległość między dwoma klasami jest równa średniej arytmetycznej wszystkich odległości między obiektami należącymi do dwóch różnych klas, tzn:

$$d\ /S_p^{i-1},\ S_r^{i-1}/ = \frac{1}{n_p^{i-1} \cdot n_r^{i-1}} \sum_{\substack{P_s \in S_p^{i-1} \\ P_t \in S_r^{i-1}}} d\ /P_s,\ P_t/$$

gdzie n_i - liczebność i-tej klasy.

4. metoda średniej wewnątrzgrupowej /por. [84] /.

W metodzie tej:

$$d\ /S_p^{i-1},\ S_r^{i-1}/ = \frac{2}{/n_p^{i-1} + n_r^{i-1}/\ /n_p^{i-1} + n_r^{i-1} - 1/} \left[\sum_{\substack{P_s \in S_p^{i-1} \\ P_t \in S_r^{i-1}}} d\ /P_s,\ P_t/ + \sum_{\substack{P_s \in S_p^{i-1} \\ P_t \in S_p^{i-1}}} d\ /P_s,\ P_t/ + \sum_{\substack{P_s \in S_r^{i-1} \\ P_t \in S_r^{i-1}}} d\ /P_s,\ P_t/ \right]$$

Interpretacja powyższego, skomplikowanego wzoru, jest oczywista: odległość między dwoma klasami jest równa średniej arytmetycznej wszystkich odległości między obiektami należącymi do tych dwóch klas. Zatem, stosując tę metodę, na każdym etapie grupowania dokonuje się grupowania takich dwóch klas, aby w efekcie otrzymać klasę, dla której średnia odległość między obiektami jest najmniejsza.

5. metoda centroidalna /por. [1] , [4] ,[29] ,[47] , [73], [84] ,[85] ,[133] ,[136] , [147] /.

W metodzie tej odległość między klasami jest równa odległości między środkami ciężkości tych klas:

$$d\ /S_p^{i-1},\ S_r^{i-1}/ = d\ /g\ /S_p^{i-1}/,\ g\ /S_r^{i-1}//$$

6. metoda Warda /por. [1], [4] , [26], [29], [47], [133],[136], [144], [147] /.

$$d\ /S_p^{i-1},\ S_r^{i-1}/ = \frac{n_r^{i-1} \cdot n_p^{i-1}}{n_p^{i-1} + n_r^{i-1}} \quad d\ /g\ /S_p^{i-1}/,\ g\ /S_r^{i-1}/\ /$$

Oprócz tych sześciu metod spotyka się jeszcze metodę skupiania parami, metoda medianowa i metoda kombinatoryczna /giętka/ /por. [1] ,[47] ,[84] , [136] /. Dla metod tych nie można jednak podać ogólnego wzoru na odległość między klasami, gdyż zmienia się ona w kolejnych etapach grupowania i nie jest uzależniona od liczebności klas.

Większość z podanych powyżej hierarchicznych metod grupowania /oprócz metody średniej wewnątrzgrupowej/ posiada ciekawą własność /por. [1] , [144] /. Jak wiadomo, po połączeniu dwóch klas w jedną należy odległości tych dwóch klas od pozostałych /nie uczestniczących w grupowaniu w danym etapie/, zastąpić odległościami nowo powstałej klasy od pozostałych. Można wykazać, że spełniona jest własność:

$$d\ /S_t^{i},\ S_r^{i}/ = a_1 \cdot d\ /S_j^{i-1},\ S_r^{i-1}/ + a_2 \cdot d\ /S_k^{i-1},\ S_r^{i-1}/ +$$
$$+ a_3 \cdot d\ /S_j^{i-1},\ S_k^{i-1}/ + a_4 \left| d\ /S_j^{i-1},\ S_r^{i-1}/ - d\ /S_k^{i-1},\ S_r^{i-1}/ \right|$$

przy czym S_j^{i-1} i S_k^{i-1} są to klasy grupowane w danym, i-tym etapie.

W przypadku stosowania trzech, wymienionych na końcu, metod, korzystanie z powyższego wzoru jest jedynym możliwym sposobem. Natomiast dla metody centroidalnej i Warda, jak można sprawdzić, wzór ten jest prawdziwy jedynie w przypadku, gdy odległość między obiektami jest określona jako kwadrat odległości euklidesowej.

Wartości współczynników a_1, a_2, a_3 i a_4 dla poszczególnych hierarchicznych metod grupowania przedstawiane są poniżej:

dla metody najbliższego sąsiada:

$a_1 = 0.5$, $a_2 = 0.5$, $a_3 = 0$, $a_4 = -0.5$,

dla metody najdalszego sąsiada:

$a_1 = 0.5$, $a_2 = 0.5$, $a_3 = 0$, $a_4 = 0.5$,

dla metody średniej międzygrupowej:

$a_1 = n_j^{i-1} \quad /n_j^{i-1}+n_k^{i-1}/$, $a_2 = n_k^{i-1} \quad /n_j^{i-1}+n_k^{i-1}/$, $a_3 = 0$, $a_4 = 0$,

dla metody centroidalnej:

$a_1 = n_j^{i-1} \quad /n_j^{i-1}+n_k^{i-1}/$, $a_2 = n_k^{i-1} \quad /n_j^{i-1}+n_k^{i-1}/$, $a_3 = -n_j^{i-1}n_k^{i-1}$
$/n_j^{i-1}+n_k^{i-1}/^2$, $a_4 = 0$,

dla metody Warda:

$a_1 = /n_j^{i-1}+n_r^{i-1}/ \quad /n_j^{i-1}+n_k^{i-1}+n_r^{i-1}/$, $a_2 = /n_k^{i-1}+n_r^{i-1}/ \quad /n_j^{i-1}+$
$n_k^{i-1}+n_r^{i-1}/$, $a_3 = -n_r^{i-1} \quad /n_j^{i-1}+n_k^{i-1}+n_r^{i-1}/$, $a_4 = 0$,

dla metody skupiania parami:

$a_1 = 0.5$, $a_2 = 0.5$, $a_3 = 0$, $a_4 = 0$,

dla metody medianowej:

$a_1 = 0.5$, $a_2 = 0.5$, $a_3 = -0.25$, $a_4 = 0$,

dla metody kombinatorycznej:

$a_1 + a_2 + a_3 = 1$, $a_1 = a_2$, $a_4 = 0$.

Dokonując analizy przedstawionych powyżej hierarchicznych metod grupowania polecić wypada szczególnie metody średniej międzygrupowej, średniej wewnątrzgrupowej, centroidalną i Warda. Pozostałe metody posiadają pewne wady. Np. metoda najbliższego sąsiada bazuje na najmniejszej odległości między obiektami należącymi do dwóch różnych klas. Może to prowadzić do sytuacji, że zostają grupowane klasy, w których niektóre obiekty są bardzo odległe od siebie /a więc mało podobne/.
Stosowanie tej metody prowadzi do dołączania obiektów do klas już istniejących /gdyż pod uwagę brana jest tylko odległość do najbliższego obiektu/. Rzadziej natomiast powstają nowe podzbiory. Zjawisko to nazywane jest efektem łańcucha /effet de Chainage - por.[136]/.

Z kolei stosowanie metody najdalszego sąsiada prowadzi do tego, że powstaje dużo mało liczebnych klas, często bardzo do siebie podobnych.
Również ostatnie trzy metody nie są pozbawione wad. Wynikają one z faktu, że w schemacie rekurencyjnym przy obliczaniu odległości klasy o różnych liczebnościach traktowane są równoprawnie.
W wyniku stosowania hierarchicznych metod grupowania otrzymuje się ciąg klasyfikacji składający się z n elementów, przy czym w kolejnych etapach ilość klas maleje od n do 1. W zagadnieniu klasyfikacji chodzi o wyznaczenie k niepustych i rozłącznych klas, należy zatem z ciągu klasyfikacji wybrać jedną. W przypadku, gdy liczba k jest dana /np. wynika z celu badania/ należy wybrać klasyfikację $S^{n-k} = \{S_1^{n-k}, S_2^{n-k}, \ldots, S_k^{n-k}\}$.
Osobnego komentarza wymaga przypadek, gdy ilość klas k nie

jest dana. Należy wtedy zastosować jakieś kryterium, które pozwoli z ciągu klasyfikacji wybrać najlepszą. Literatura przedmiotu nie zawiera wielu propozycji w tym zakresie. Jedno z kryteriów podane jest w [26], jest to tzw. miernik straty informacji Berry'ego. W myśl tego kryterium, przez połączenie dwóch klas w jedną traci się pewną ilość informacji o tych klasach, konkretnie tę część informacji, która traktuje o różnicach między tymi klasami. Różne klasy przez zgrupowanie zostają "utożsamione".

Jako klasyfikację ostateczną proponuje przyjąć się tę klasyfikację S^1, dla której strata informacji, określona jako:

$$\frac{m_{1+1} - m_1}{m_1}, \quad \text{gdzie:}$$

$$m_1 = \sum_{S^1_j \in S^1} \; \sum_{\substack{k,\, l \\ P_k, P_l \in S^1_j}} d/P_k, P_l/$$

osiąga wartość maksymalną.

Jak wynika z definicji zagadnienia klasyfikacji, dobra klasyfikacja powinna zawierać klasy o dużym wewnętrznym podobieństwie.

Wewnętrzne podobieństwo w klasie może być charakteryzowane przez średnicę klasy:

$$r_t = \max_{P_i, P_j \in S_t} d/P_i, P_j/$$

Z kolei, wewnętrzne podobieństwo klasyfikacji można zdefiniować:

$$W_p = \max_t \; r_t = \max_t \; \max_{P_i, P_j \in S_t} d/P_i, P_j/$$

Wewnętrzne podobieństwo klasyfikacji zostało tu zdefiniowane jako maksymalna odległość między obiektami należącymi do tej samej klasy.

Innym miernikiem, jak się wydaje bardziej przydatnym /z tego powodu, że uwzględnia wszystkie odległości wewnątrzgrupowe/, jest:

$$W_p' = \frac{2}{\sum_{t=1}^{k} n_t \cdot /n_t - 1/} \sum_{t=1}^{k} \sum_{P_i, P_j \in S_t} d/P_i, P_j/$$

Jest to średnia odległość wewnątrzgrupowa.

Dla wyboru z ciągu klasyfikacji jednej w oparciu o kryterium wewnętrznego podobieństwa można zastosować następującą procedurę:

- z reguły cel badania implikuje dolny i górny koniec przedziału, w którym powinna znaleźć się ilość klas,
- wewnętrzne podobieństwo W_p lub W_p' jest dla hierarchicznych metod grupowania wielkością rosnącą /oznacza to, że w miarę wzrostu ilości etapów grupowania rośnie również średnia czy maksymalna odległość wewnątrzgrupowa/, zatem grupowanie należy przerwać po tym etapie, po którym uzyskuje się relatywnie wysoki przyrost W_p lub W_p' /co świadczy o dużym spadku podobieństwa wewnątrzklasowego/; oczywiście rozpatrywane są jedynie klasyfikacje o ilości klas spełniającej warunek sformułowany w punkcie poprzednim.

Niehierarchiczne metody grupowania

Algorytm:

1. Dana jest klasyfikacja początkowa $S^o = \{S_1^o, S_2^o, \ldots, S_n^o\}$, gdzie $S_i^o = P_i$ $/i = 1, 2, \ldots, n/$

2. W jednym lub więcej etapach dokonuje się łączenia /grupowania/ kilku klas w jedną. Sposób grupowania zależy od stosowanej metody.
3. Po zakończeniu grupowania otrzymuje się klasyfikację $S = \{S_1, S_2, \ldots, S_k\}$, gdzie liczba k jest dana lub /częściej/ wynika z metody.

Do tej grupy metod zaliczane są wszystkie pozostałe /oprócz hierarchicznych/ metody grupowania. W wyniku stosowania tych metod otrzymuje się klasyfikację ostateczną /a nie ciąg klasyfikacji, z których należy dokonać wyboru w oparciu o jakieś kryterium/.
Grupa tych metod jest liczna i różnorodna. Poniżej zaprezentowane zostanie kilka najbardziej godnych uwagi i najpowszechniej stosowanych metod.

1. Metoda progowa /por.[121]/.
   Metoda ta jest realizowana w jednym etapie. Zostają zgrupowane /łączone/ te obiekty, dla których:

   $$d/P_i, P_j/ \leqslant d_o.$$

   Wielkość d_o jest wielkością progową. Za podobne zostają uznane /a tym samym zgrupowane w jednej klasie/ obiekty odległe od siebie o odległość równą co najwyżej odległości progowej. Wadą tej metody jest pewna arbitralność przy wyborze wielkości progowej. Zaznaczyć należy, że w literaturze brak jest propozycji w tym zakresie. Wydaje się, że wartość progowa powinna być uzależniona od wszystkich odległości między obiektami, np.

   $$d_o = \frac{2}{n/n-1/} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d/P_i, P_j/$$

Ogólnie powinno być:

$$\bar{d} - s_d \leqslant d_o \leqslant \bar{d} + s_d$$

Dużą zaletą metody jest jej prostota i intuicyjność.

2. Metoda dendrytowa /metoda elementarnego połączenia/ /por. [26]/. W metodzie tej każdy obiekt P_i zostaje połączony z takim obiektem P_j, dla którego:

$$d/P_i,P_j/ = \min_{P_l \neq P_i} d/P_i,P_l/$$

Każdy obiekt zostaje połączony ze swoim "najbliższym sąsiadem". Procedura ta dokonywana jest w jednym etapie.
Wadą tej metody jest uzależnienie grupowania tylko od maksymalnego podobieństwa /minimalnej odległości/ dla każdego obiektu. Powoduje to, że bezpośrednio nie uwzględniane są podobieństwa obiektów mniejsze od maksymalnych. Może się zdarzyć, że obiekty bardziej podobne znajdą się w różnych klasach, a mniej podobne w tej samej; /efekt łańcucha - przez łańcuch rozumie się ciąg maksymalnych podobieństw, "najbliższych sąsiadów", który może połączyć obiekty mało podobne - por. [136],[137]/.
Dużą zaletą metody jest prostota.

3. Metoda Daceya /por. [26] /
Jest to pewien wariant metody dendrytowej. Tutaj obiekt P_i zostaje połączony z takim obiektem P_j, dla którego:

$$\delta_{ij} = \left| d/P_iP_j/ - d/P_i,P_{i_o}/ \right| - \left| d/P_j,P_i/ - d/P_j,P_{j_o}/ \right|$$

gdzie: $d/P_j,P_{j_o}/ = \min\limits_{l \neq j} d/P_j,P_l/$

$d/P_i,P_{i_o}/ = \min\limits_{l \neq i} d/P_i,P_l/$

Obiekt zostaje połączony ze swoim najbliższym sąsiadem, ale nie w sensie odległości, lecz różnicy odległości od odległości minimalnej. Metoda ta powoduje tworzenie większej ilości klas niż metoda dendrytowa i w mniejszym stopniu ulega efektowi łańcucha. W tym tkwi jej zaleta.

4. Wrocławska metoda kul /por.[20]/.

W metodzie tej, podobnie jak kilku następnych, klasami są kule. Oznacza to, że w danej klasie znajdują się obiekty, dla których odległość od środka kuli /czyli pewnego punktu/ jest nie większa od promienia, tzn.

$$d/P_j, A_i/ \leqslant r_i$$

gdzie:

A_i - środek i-tej kuli,

r_i - promień i-tej kuli.

We wrocławskiej metodzie kul promień jest jednakowy dla wszystkich kul i zadany z góry /jest to niewątpliwie wada tej metody/, np.

$$r = \max_{P_j} \min_{P_k} d/P_j, P_k/$$

lub $r = \bar{d} + 2 \cdot s_d$

gdzie:

$$\bar{d} = \frac{2}{n/n-1/} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} d/P_j, P_k/$$

$$\bar{s}_d = \sqrt{\frac{2}{n/n-1/} \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} /d\,/P_j, P_k/ - \bar{d}/^2}$$

Wyodrębnienie kul /klas/ odbywa się etapami, przy czym w każdym etapie otrzymuje się jedną klasę. Jako pierwszy środek kuli wybiera się obiekt P_1, dla którego

$$\mathrm{card}\left\{P_j \quad d/P_i,P_j/ \leqslant r\right\} = \max_{P_l} \quad \mathrm{card}\left\{P_j \quad d/P_l,P_j/ \leqslant r\right\},$$

czyli obiekt, wokół którego skonstruowana kula jest najliczniejsza. Kula ta jest pierwszą klasą.

Postępowanie to powtarza się /z pominięciem obiektów, które znalazły się w skonstruowanych klasach/ aż do przydzielenia wszystkich obiektów do kul /klas/.

Metoda ta wydaje się być przydatna, gdy w zbiorze obiektów istnieją wyraźne skupiska "gęste", tzn. grupy skoncentrowanych obiektów. Jest ona w decydującym stopniu uzależniona od zadanej wielkości promienia i zwykle stosowanie jej powoduje powstanie jednej bardzo licznej i kilku mało licznych klas. Jest to jej wada.

5. Metoda kul Wisharta /por. [1], [148]/.

Algorytm tej metody zbliżony jest do metody wrocławskiej, z tym, że promień kuli zmienia się. Dla każdego obiektu tworzone są kule o środku w punkcie o współrzędnych równych wartościom zmiennych dla tego obiektu oraz promieniu równym odległości do l-tego /gdzie liczba l jest zadana z góry/, w kolejności rosnących odległości od innych obiektów, obiektu. Oczywiście promienie będą różne dla każdego obiektu.

Dalszy sposób postępowania jest taki sam jak w metodzie wrocławskiej, tzn. klasy konstruowane są kolejno, począwszy od obiektu, który jest środkiem najliczniejszej kuli. Postępowanie powtarza się z pominięciem obiektów już przydzielonych.

Jak widać, i w tym przypadku pierwsza klasa jest najliczniejsza, liczy l+1 obiektów. Zaliczyć to należy do wad metody.

6. Metoda Sebestyena /por. [129] /.

W metodzie tej obiekty rozpatrywane są w ustalonej kolejności, a więc np. w kolejności numerów obiektów lub kolejności losowej. Promień jest jednakowy dla wszystkich kul. Pierwszy rozpatrywany obiekt jest środkiem pierwszej kuli. Kolejno rozpatrywane obiekty mogą znaleźć się w pierwszej kuli /klasie/ - w przypadku, gdy ich odległość od środka kuli jest mniejsza od promienia, lub utworzyć nową kulę /klasę/, stając się jej środkiem, w przypadku przeciwnym. Po dołączeniu obiektu do istniejącej klasy środkiem kuli staje się środek ciężkości obiektów znajdujących się w danej klasie. W przypadku, gdy warunek

$$d/P_i, A_j/ \leqslant r$$

spełniony jest dla kilku klas, proponujemy obiekt P_i dołączać do tej klasy, której środek jest najbliższy danemu obiektowi.

7. Metoda TAXMAP /por. [47] /.

W metodzie tej /którą podajemy w postaci nieco zmodyfikowanej, tzn. w oparciu o macierz odległości, a nie macierz podobieństw/ w pierwszej kolejności łączy się obiekty najmniej odległe. W następnych etapach rozważane są kolejno obiekty najmniej odległe od już skonstruowanych klas. Warunkiem dołączenia obiektu do jednej z klas jest relatywnie niski spadek przeciętnej odległości wewnątrzgrupowej.

Analizując powyższe metody trzeba stwierdzić, że są one w większości proste i intuicyjne, jednak rezultaty otrzymane za pomocą tych metod są przydatne w przypadkach, gdy istnieje klasyfikacja naturalna, tzn. gdy są ostro zarysowane granice między klasami /dotyczy to zwłaszcza metod "kulowych"/.

3.2.2. Metody podziału.

Ogół metod podziału można podzielić na 2 podstawowe grupy: metody hierarchiczne /sekwencyjne/ i niehierarchiczne.

Hierarchiczne metody podziału.

Algorytm:

1. Dana jest klasyfikacja początkowa $S^o = \{S_1^o\} = P$.
2. W r-tym etapie hierarchicznego podziału dokonuje się podziału jednej z klas. Jest to ta klasa S_t^{i-1}, dla której spełniony jest warunek:

   $$d/P_r, P_S/ = \max_{S_j^{i-1}} \max_{P_k, P_l \in S_j^{i-1}} d/P_k, P_l/$$

   Klasa S_t^{i-1} zostaje podzielona na 2 klasy: S_t^i oraz S_{t+1}^i, przy czym sposób podziału zależy od stosowanej metody.

   Po zakończeniu i-tego etapu hierarchicznego podziału otrzymuje się klasyfikację: $S^i = \{S_1^i, S_2^i, \ldots, S_{i+1}^i\}$, gdzie:

   $$S_j^i = S_j^{i-1} \qquad j = 1, 2, \ldots, t-1$$

   $$S_j^i = S_{j+1}^{i-1} \qquad j = t \quad 2, \ldots, i+1$$

3. Po zakończeniu n-1 etapów hierarchicznego podziału otrzymuje się ciąg klasyfikacji:

   $$S^o = S_1^o = P$$

   $$S^1 = \{S_1^1, S_2^1\}$$

   $$S^{n-1} = \{S_1^{n-1}, S_2^{n-1}, \ldots, S_n^{n-1}\}$$

4. Z tego ciągu klasyfikacji wybiera się jedną. Jest to klasyfikacja $S = \{S_1, S_2, \ldots, S_k\}$.

W zależności od stosowanego sposobu podziału wyróżnionej klasy wyróżnić można różne hierarchiczne metody podziału:

1. metoda odległości maksymalnej /por. [84]/.

   Obiekt P_r przydzielany jest do klasy S_t^i, a obiekt P_S do klasy S_{t+1}^i. Pozostałe obiekty klasy S_t^{i-1} rozdzielane są sekwencyjnie do jednej z dwóch klas wg zasady:

   wybierany jest obiekt P_j, dla którego spełniony jest warunek:

   $$\max\left\{d/P_j,S_t^i/,\ d/P_j,S_{t+1}^i/\right\} = \max_{P_l\in S_t^{i-1}} \max\left\{d/P_l,S_t^i/,\ d/P_l,S_{t+1}^i\right.$$

   Obiekt ten przydzielany jest do tej klasy, dla której realizowane jest $\min\limits_{S_t^i,S_{t+1}^i}\left\{d/P_j,\ S_t^i/,\ d/P_j,S_{t+1}^i/\right\}$

2. metoda odległości minimalnej /por. [84]/.

   Obiekt P_r przydzielany jest do klasy S_t^i, a obiekt P_S do klasy S_{t+1}^i.

   Pozostałe obiekty klasy S_t^{i-1} rozdzielane są sekwencyjnie do jednej z dwóch klas wg zasady:

   wybierany jest obiekt P_j, dla którego spełniony jest warunek:

   $$\min\left\{d/P_j,S_t^i/,\ d/P_j,S_{t+1}^i/\right\} = \min_{P_l\in S_t^{-1}} \min\left\{d/P_l,S_t^i/,\ d/P_l,S_{t+1}^i/\right\}.$$

   Obiekt ten przydzielany jest do tej klasy, dla której realizowane jest $\min\limits_{S_t^i,S_{t+1}^i}\left\{d/P_j,S_t^i/,\ d/P_j,S_{t+1}^i/\right\}$

3. metoda rozdziału jednoczesnego

   Obiekt P_r przydzielany jest do klasy S_t^i, a obiekt P_S do klasy S_{t+1}^i. Pozostałe obiekty klasy S_t^{i-1} rozdzielane są jednocześnie wg zasady:

Obiekt P_j przydzielany jest do tej klasy, dla której jest realizowane:

$$\min_{S_t^i, S_{t+1}^i} \left\{ d/P_j, S_t^i/, \quad d/P_j, S_{t+1}^i/ \right\} = \min \left\{ d/P_j, P_r/, \quad d/P_j, P_s/ \right\}$$

We wszystkich z powyższych metod obiekty wyróżnionej klasy są rozdzielane na dwie klasy. W dwóch pierwszych metodach dzielenie odbywa się stopniowo, oczywiście za każdym razem zmianie ulegają odległości między nie przydzielonymi obiektami a nowotworzonymi klasami. W zależności od sposobu liczenia odległości między obiektem a klasą wyróżnić można różne warianty dwóch pierwszych metod. Jedyna różnica między obu metodami polega na wyborze obiektu, który ma zostać przydzielony.
W przypadku metody trzeciej obiekty rozdzielane są jednocześnie. Pewną wadą hierarchicznych metod podziału jest to, iż w każdym etapie podział determinowany jest odległościami obiektów jedynie od dwóch obiektów, tych, między którymi odległość wewnątrzgrupowa jest największa, natomiast odległości pozostałe nie są uwzględniane.

Niehierarchiczne metody podziału.

Algorytm·

1. Dana jest klasyfikacja początkowa $S^o = S_1^o = P$.
2. W jednym lub kilku etapach dokonuje się podziału klasy na klasy o mniejszej liczebności. Sposób podziału zależy od stosowanej metody.
3. Po zakończeniu podziału otrzymuje się klasyfikację:

$$S = \left\{ S_1, S_2, \ldots, S_k \right\}$$

Literatura przedmiotu nie doczekała się wielu propozycji niehierarchicznych metod podziału. Godną polecenia jest tzw.

metoda analizy niepodobieństwa, zaproponowana w [93]. W metodzie tej w etapie podziału jedna klasa S_j^{i-1} jest dzielona na dwie klasy S_j^i oraz S_{j+1}^i, przy czym jedna z tych klas, S_j^i powstaje przez sekwencyjne dołączanie obiektów klasy S_j^{i-1} tak, aby średnia odległość dołączanego obiektu od pozostałych w klasie S_j^{i-1} obiektów minus średnia odległość dołączanego obiektu od obiektów klasy S_j^i była maksymalna. Proces dołączania jest kontynuowany, dopóki powyższa różnica jest dodatnia. Pozostałe obiekty tworzące uprzednio klasę S_j^{i-1} tworzą klasę S_{j+1}^i.

3.2.3. Metody iteracyjne.

Ogół metod iteracyjnych można podzielić na 2 podstawowe grupy:

- metody bazujące na funkcji jakości klasyfikacji,
- metody bazujące na punktach skupienia.

Przy formułowaniu zagadnienia klasyfikacji podano ogólne kryterium "dobroci" klasyfikacji. Traktowało ono o dużym podobieństwie obiektów należących do tych samych klas oraz małym podobieństwie obiektów należących do różnych klas. Na kryterium tym opierają się zarówno funkcje jakości klasyfikacji, jak również /pośrednio/ punkty skupienia.

Przez funkcję jakości klasyfikacji rozumieć będziemy funkcję:

$$h_k : \mathcal{S}_k \rightarrow R$$

gdzie $\mathcal{S}_k$ - zbiór wszystkich możliwych klasyfikacji składających się z k klas, jakie mogą być otrzymane z elementów zbioru obiektów P.

Dla ustalenia uwagi przyjmiemy, że:

$$h_i \,/\left\{S_1^1,\ S_2^1,\ \ldots,\ S_i^1\right\}/ < h_j \,/\left\{S_1^2,\ S_2^2,\ \ldots,\ S_j^2\right\}/ \Leftrightarrow$$

klasyfikacja S^1 jest lepsza od klasyfikacji S^2.

Należy zwrócić uwagę, iż większość funkcji jakości /jak zobaczymy poniżej/ umożliwia porównywanie klasyfikacji składających się z jednakowej ilości klas, natomiast dla klasyfikacji o różnych ilościach klas nie jest to możliwe.

Również, choć bardziej pośrednio, z ogólnego kryterium klasyfikacji wynika zasadność stosowania podejścia opartego na tzw. punktach skupienia. Wynika ono z tego, że dla " dobrej " klasyfikacji klasy zawierają obiekty podobne do siebie, tzn. skupione wokół siebie /w sensie interpretacji geometrycznej/.

Punkty skupienia /mogą być nimi niektóre z badanych obiektów lub obiekty skonstruowane za pomocą pewnych procedur formalnych/ są wyróżnikami czy reprezentantami klas obiektów. Wokół nich koncentrują się obiekty danej klasy.

Zarówno metody bazujące na funkcji jakości klasyfikacji, jak i te bazujące na punktach skupienia można podzielić na 2 podstawowe grupy:

- metody o stałej ilości klas,
- metody o zmiennej ilości klas /w kolejnych iteracjach ilość klas może ulegać zmianie/.

Metody iteracyjne bazujące na funkcji jakości klasyfikacji.

Algorytm:

1. Dana jest klasyfikacja początkowa $S^o = \left\{S_1^o,\ S_2^o, \ldots,\ S_k^o\right\}$,
2. W i-tej iteracji bada się kolejno obiekty P_1, $P_2, \ldots$, itd., aż do znalezienia obiektu $P_j \in S_r^{i-1}$, dla którego istnieje klasa S_l^{i-1}, taka, że:

a/ dla metod o stałej ilości klas:

$$h_k / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_l^{i-1} \cup P_j, \ldots, S_r^{i-1} - P_j, S_k^{i-1} \right\} / <$$

$$h_k / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_k^{i-1} \right\} /$$

b/ dla metod o zmiennej ilości klas:

b1/ $h_k / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_l^{i-1} \cup P_j, \ldots, S_r^{i-1} - P_j, \ldots, S_k^{i-1} \right\} /$

$< h_k / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_k^{i-1} \right\} /$ lub:

b2/ $h_{k+1} / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots, S_k^{i-1}, S_l^{i-1} \right\} / <$

$h_k / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_k^{i-1} \right\} /$ gdzie: $S_l^{i-1} = P_j$

W przypadku, gdy zachodzi warunek a/ obiekt P_j zostaje przydzielony do tej klasy S_s^{i-1}, która spełnia warunek:

$$h_k / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots, S_s^{i-1} \cup P_j, \ldots, S_k^{i-1} \right\} / =$$

$$= \min_l h_k / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots, S_l^{i-1} \cup P_j, \ldots, S_k^{i-1} \right\} /$$

i tworzy się nową klasyfikację: $S^i = \left\{ S_1^i, S_2^i, \ldots, S_k^i \right\}$;

gdzie:

$S_r^i = S_r^{i-1} - P_j$, $S_s^i = S_s^{i-1} \cup P_j$; $S_l^i = S_l^{i-1}$ $/ l \neq r, \; l \neq s /$

Natomiast w przypadku gdy zachodzi warunek b1 lub b2, obiekt P_j zostaje przydzielony do klasy S_s^{i-1} jeśli:

$$h_k / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots, S_s^{i-1} \cup P_j, \ldots, S_k^{i-1} \right\} / =$$

$$\min_l h_k / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots, S_l^{i-1} \cup P_j, \ldots, S_k^{i-1} \right\} / <$$

$h_{k+1} / \left\{ S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots, S_k^{i-1}, S_l^{i-1} \right\} /$ a do klasy S_l^{i-1} /gdzie $S_l^{i-1} = P_j$ /, gdy zachodzi nierówność o przeciwnym zwrocie.

Tworzy się następnie nową klasyfikację $S^i = \left\{S_1^i, S_2^i, \dots, S_k^i\right\}$ lub $S^i = \left\{S_1^i, S_2^i, \dots, S_{k+1}^i\right\}$ /w zależności od zwrotu nierówności/, gdzie:

$S_r^i = S_r^{i-1} - P_j$; $S_s^i = S_s^{i-1} \cup P_j$; $S_l^i = S_l^{i-1}$ /$l \neq r$, $l \neq s$/ ; $S_{k+1}^i = P_j$

3. Procedurę opisaną w punkcie 2 kontynuuje się, aż do momentu, gdy nie jest spełniony żaden z warunków a, b1, b2.

Jak wynika z opisanego powyżej algorytmu, w metodach iteracyjnych bazujących na funkcji jakości klasyfikacji, startuje się od pewnej, zadanej klasyfikacji /składającej się z ustalonej ilości klas/ i w kolejnych iteracjach klasyfikacja ta jest poprawiana, przy czym probierzem poprawy klasyfikacji jest spadek wartości funkcji klasyfikacji.

Zmiany, czyli poprawy klasyfikacji dokonuje się przez " przeniesienie " obiektu z jednej klasy do innej, a w przypadku metod o zmiennej ilości klas dopuszczane jest również, " odłączenie " obiektu od klasy i utworzenie przez niego danej iteracji samodzielnej klasy.

Oczywiście metoda realizowana jest do momentu, gdy żadna zmiana klasyfikacji nie spowoduje spadku wartości funkcji jakości klasyfikacji.

Jak wynika z powyższych rozważań, można wyróżnić różne metody iteracyjne bazujące na funkcji jakości klasyfikacji w zależności od:

- stosowanej funkcji jakości klasyfikacji,
- zadanej klasyfikacji początkowej.

Wybór funkcji jakości klasyfikacji.

Istnieje duża ilość możliwych funkcji klasyfikacji. Na funkcje te nałożony jest w zasadzie tylko jeden warunek: wynika z ogólnego kryterium podanego w zagadnieniu klasyfikacji. A oto kilka najciekawszych funkcji jakości klasyfikacji:

$$h_k /\{S_1, S_2, \ldots, S_k\}/ = \sum_{j=1}^{k} \frac{1}{n_j} \sum_{P_i, P_l \in S_j} d/P_i, P_l/ \qquad /3.1/$$

$$h_k /\{S_1, S_2, \ldots, S_k\}/ = \frac{1}{n} \sum_{j=1}^{k} \sum_{P_i \in S_j} d/P_i, g/S_j// \qquad /3.2/$$

$$h_k /\{S_1, S_2, \ldots, S_k\}/ = -\frac{1}{k} \sum_{j=1}^{k} d/g/S_j/, g/P// \qquad /3.3/$$

$$h_k /\{S_1, S_2, \ldots, S_k\}/ = \sum_{j=1}^{k} \max_{P_i, P_l \in S_j} d/P_i, P_l/ - \sum_{j_1=1}^{k-1} \sum_{j_2=j_1+1}^{k} \min_{\substack{P_i \in S_{j_1} \\ P_l \in S_{j_2}}} d/P_i, P_l/ \qquad /3.4/$$

$$h_k /\{S_1, S_2, \ldots, S_k\}/ = \sum_{j=1}^{k} \sum_{P_i, P_l \in S_j} d/P_i, P_l/ - \sum_{j_1=1}^{k} \sum_{j_2=1}^{k} \sum_{\substack{P_i \in S_{j_1} \\ P_l \in S_{j_2}}} d/P_i, P_l/ \qquad /3.5/$$

gdzie:

n_i - liczebność i-tej klasy.

Przykłady innych funkcji jakości klasyfikacji /w tym również funkcji dla porównywania klasyfikacji składających się z różnych ilości klas/ można znaleźć w [1],[51],[97],[121],[136], [156].

Z powyżej przedstawionych, funkcje dane wzorami /3.1/ i /3.2/ mierzą podobieństwo wewnątrzklasowe /tzn. podobieństwo obiektów znajdujących się w tych samych klasach/. Funkcja /3.3/ mierzy zewnątrzklasowe podobieństwo, a ściślej mówiąc niepodobieństwo.

Pewnym powiązaniem kryteriów wewnątrzklasowego i zewnątrzklasowego podobieństwa są funkcje 3.4 i 3.5, które mierzą różnicę między wewnątrzklasowym i zewnątrzklasowym podobieństwem.

Wybór klasyfikacji początkowej.

Istnieje wiele możliwych sposobów wyznaczania klasyfikacji początkowej:

1. klasyfikacja wyznaczona subiektywnie, w sposób intuicyjny przez badacza,
2. klasyfikacja losowa, tzn. przydzielenie obiektów do klas w sposób losowy,
3. klasyfikacja otrzymana w wyniku zastosowania metody klasyfikacji, np. metody grupowania, czy podziału ,
4. klasyfikacja quasi - losowa, tzn. przydzielenie obiektów do klas w sposób nielosowy, lecz nie związany z naturą badanych zjawisk, np.

$$S_1^o = \left\{P_1, P_{k+1}, \dots, P_{\left[\frac{n}{k}\right]+1}\right\}$$

$$S_2^o = \left\{P_2, P_{k+2}, \dots, P_{\left[\frac{n}{k}\right]+2}\right\}$$

$$\dots\dots\dots$$

$$S_k^o = \left\{P_k, P_{2k}, \dots, P_{\left[\frac{n}{k}\right]}\right\}$$

5. klasyfikacja otrzymana przez przydzielenie obiektów do klas skoncentrowanych wokół pewnych punktów skupienia $G_1, G_2, \dots, G_k$ /o punktach skupienia traktują dalsze rozważania/ wg. zasady:

$$S_1^o = \left\{P_{i_1} \quad d\,/P_{i_1}, G_1/ = \min_{G_j} \overline{d\,/P_{i_1}, G_j/}\right\}$$

$$S_2^o = \left\{P_{i_2} \quad d\,/P_{i_2}, G_2/ = \min_{G_j} \overline{d\,/P_{i_2}, G_j/}\right\}$$

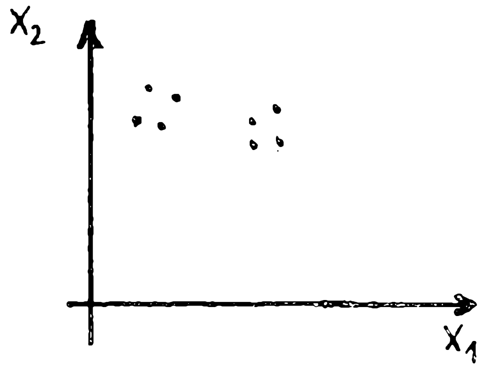
$$\dots\dots\dots\dots\dots$$

$$S_k^o = \left\{P_{i_k} : d\,/P_{i_k}, G_k/ = \min_{G_j} \overline{d/P_{i_k}, G_j/}\right\}$$

Oczywiście można zaproponować inne, bardziej skomplikowane metody wyznaczania klasyfikacji początkowej /traktuje o nich np. [41]/, ale jak łatwo zauważyć, zaliczyć je można do sposobu 3, gdyż wtedy wybór klasyfikacji początkowej sprowadza się do stosowania mniej lub bardziej skomplikowanej procedury, którą można utożsamić z metodą klasyfikacji.
Analiza powyższych propozycji wyboru klasyfikacji początkowej skłania do stwierdzenia, że najbardziej sensowna jest propozycja 3. Wydaje się ona być najbardziej "obiektywna". Ponadto z reguły gwarantuje uzyskanie klasyfikacji najlepszej /tzn. takiej, dla której funkcja jakości klasyfikacji osiąga minimum/ po niewielkiej ilości iteracji. Ma ona jedną wadę, iż /w zależności od stosowanej procedury/ może być pracochłonna. Również polecić można propozycję 1, pod warunkiem, że badacz wyznaczający tę subiektywną klasyfikację dysponuje dostatecznie dużą wiedzą merytoryczną o badanym zjawisku. W przeciwnym bowiem wypadku propozycja ta nie będzie się wiele różnić od propozycji 2. Najgorsze wydają się być propozycje 2 i 4, choćby z tego powodu, że znacznie zwiększają ilość iteracji potrzebnych do uzyskania klasyfikacji najlepszej, co przy dużej ilości obiektów i zmiennych nie jest błahostką, nawet przy zastosowaniu komputera. Natomiast wartość propozycji 5 zależy od sposobu wyboru punktów skupienia. Problemem tym zajmiemy się w dalszej części pracy.
Idea opisanych powyżej metod jest prosta. Posiadają one jednak pewne wady. Jedną z nich jest pewna sztywna formalizacja jakości klasyfikacji za pomocą funkcji. Zdaniem autora, najpoważniejszą wadą, która jednak tyczy tylko metod o stałej ilości klas, jest fakt, iż dana musi być ilość klas, na które

należy podzielić zbiór obiektów. Zadanie w sposób arbitralny /co się zdarza najczęściej/ tej liczby może spowodować, że otrzymana klasyfikacja będzie sztuczna.

Sytuację tę ilustruje poniższy rysunek:



Zbiór 12 obiektów należy sklasyfikować ze względu na 2 zmienne. Rzut oka na rysunek sugeruje, że zbiór ten w sposób "naturalny" dzieli się na 3 klasy. Jednak zwykle zbioru obiektów nie można przedstawić graficznie /ze względu na dużą ilość zmiennych/ i dlatego z reguły nie jest znana liczba klas, na które można w sposób mniej lub bardziej "naturalny" podzielić ten zbiór. Liczba, która zostanie obrana w sposób arbitralny może doprowadzić do klasyfikacji sztucznej /w przypadku zilustrowanym na rysunku sztuczna byłaby klasyfikacja składająca się z 2, 4 czy 5 klas/.

Wady tej oczywiście nie posiadają w tak dużym stopniu metody o zmiennej ilości klas, choć i w tym przypadku można otrzymać rozwiązanie dalekie od optymalnego.

Z innych wad powyższych metod wymienić należy tę, że metody te nie gwarantują otrzymania minimum globalnego funkcji jakości klasyfikacji, a jedynie minimum lokalne, którego wartość w pewnym stopniu zależy od wyboru klasyfikacji początkowej.

Metody iteracyjne bazujące na punktach skupienia

Algorytm:

1. Dane są punkty skupienia: $G_1, G_2, \ldots, G_k$.
2. W i-tej iteracji obiekty $P_1, P_2, \ldots, P_n$ /a w przypadku, gdy niektóre obiekty są punktami skupienia - odpowiednio mniejsza ich ilość/ są przydzielane do odpowiedniej klasy wg zasady najbliższej odległości, tzn.

$$t/P_i/ = \min_j d/P_i, G_j/$$

gdzie $t/P_i/$ jest numerem punktu skupienia /czyli numerem klasy/, do którego przydziela się obiekt P_i.

W zależności od stosowanych metod można wyróżnić dwa sposoby przydzielania obiektów:

a/ w danej iteracji przydzielane są wszystkie obiekty,

b/ w danej iteracji przydzielany jest tylko jeden, i-ty obiekt.

W wyniku otrzymuje się klasyfikację: $S^i = \left\{ S_1^i, S_2^i, \ldots, S_k^i \right\}$, gdzie:

$$S_l^i = \left\{ P_j \quad t/P_j/ = l \right\} \quad l = 1, \ldots, k$$

Oczywiście w przypadku sposobu pierwszego w skład klasyfikacji wchodzą wszystkie obiekty, natomiast w przypadku sposobu drugiego tylko ich część /ponieważ w każdej iteracji przydzielany jest tylko jeden obiekt/ - jest tak aż do zakończenia iteracji, po której przydzieli się ostatni obiekt.

W przypadku metod o zmiennej ilości klas po utworzeniu klasyfikacji S^i w i-tej iteracji sprawdza się pewne parametry i w zależności od ich wartości klasyfikacja ulega zmianie, przez co zmianie może ulec również ilość klas. Otrzymuje się zatem klasyfikację:

$$S^1 = \left\{S_1^1, S_2^1, \ldots, S_{k_0}^1\right\}$$

O parametrach traktuje jeden z dalszych wywodów.

3. Dla klasyfikacji S^1 oblicza się centroidy /środki ciężkości/:

$$g_{lj} = \frac{1}{n_l} \sum_{\substack{t \\ P_t \in S_l}} x_{tj}$$

Centroidy stają się nowymi punktami skupienia.

4. Proces opisany w punktach 2 i 3 powtarza się gdy zajdzie jedna z poniższych sytuacji:
   - ilość iteracji przekroczy pewną, zadaną z góry liczbę,
   - klasyfikacje w poszczególnych iteracjach nie ulegają zmianie,
   - punkty skupienia nie zmieniają się.

Funkcjonowanie metod opisanych powyżej w dużym stopniu zależy od wyboru punktów skupienia.

Wybór punktów skupienia

Istnieje wiele sposobów wyznaczania punktów skupienia:

1. obiekty: $P_1, P_2, \ldots, P_k$
2. wyznaczone przez badacza intuicyjnie k obiektów,
3. obiekty $P_{\left[\frac{n}{k}\right]}, P_{2\cdot\left[\frac{n}{k}\right]}, \ldots, P_{k\cdot\left[\frac{n}{k}\right]}$
4. losowo wybrane k obiektów,
5. sztucznie skonstruowane k obiektów o wartościach współrzędnych losowo wybranych spośród wartości zmiennych występujących w obiektach,
6. centroidy klasyfikacji składającej się z k klas,
7. k obiektów o największej wartości tzw. gęstości przestrzeni wokół obiektu, przy warunku, że obiekty te odległe są od

siebie o co najmniej d_o, gdzie d_o jest to wartość zada-
na arbitralnie. Przez gęstość przestrzeni wokół obiektu rozumiemy:

$$m/P_i, \delta/ = \text{card} \{ P_j \quad d/P_i, P_j/ < \delta \}$$

8. jako pierwszy punkt skupienia - centroid zbioru P, natomiast pozostałe k - 1 punktów skupienia wybiera się spośród kolejno rozpatrywanych obiektów, uwzględniając te, które są odległe od już wybranych punktów skupienia o co najmniej d_o, gdzie d_o jest to wartość zadana arbitralnie.

Spośród powyższych propozycji godne uwagi wydają się być propozycje 6, 7 i 8. Uwaga ta dotyczy również propozycji 2, pod warunkiem, że badacz dokonujący wyboru tych punktów posiada dostatecznie dużą wiedzę merytoryczną o badanym zjawisku. Nie polecamy natomiast pozostałych propozycji ze względu na dużą przypadkowość wyboru punktów skupienia.
W przypadku metod o zmiennej ilości skupień stosowane mogą być następujące parametry, które są sprawdzane w każdej iteracji:

1. odległość centroidów.
W przypadku, gdy $d/G_j, G_k/ < d_1$, przy czym d_1 jest to pewna zadana z góry wielkość, następuje połączenie tych dwóch klas, których centroidy spełniają ten warunek tzn. S_j^i i S_k^i. Warunek ten zapewnia małe podobieństwo obiektów, należących do różnych klas i jest sprawdzany dla wszystkich par klas.

2. odległość obiektu od centroidu.
W przypadku gdy $d/P_j, G_j/ > d_2$, przy czym d_2 jest to pewna zadana z góry wielkość, następuje oddzielenie obiektu od tej klasy. Warunek ten zapewnia duże podobieństwo obiektów należących do tej samej klasy.

3. ilość klas.

W przypadku, gdy $k > k_2$, gdzie k - ilość klas, natomiast k_2 zadana z góry /wynikająca z celu badania/ maksymalna dopuszczalna ilość klas, należy pewne klasy zgrupować. Jako sposób grupowania proponujemy przyjąć sposób stosowany przy hierarchicznych metodach grupowania. Natomiast gdy $k < k_1$, gdzie k_1 - wynikająca z celu badania minimalna dopuszczalna ilość klas, należy pewne klasy podzielić. Jako sposób podziału proponujemy przyjąć ten, który jest stosowany przy hierarchicznych metodach podziału.

4. ilość obiektów w klasie.

W przypadku gdy $n_j > n_o$, gdzie n_o - maksymalna dopuszczalna ilość obiektów w klasie, klasa S_j^i zostaje podzielona. Jako kryterium podziału proponuje się przyjąć to samo co w punkcie poprzednim.

Wydaje się, że bezwzględnie należy stosować parametry 1 i 2 natomiast stosowanie parametru 3 i 4 należy uzależnić od celu badania. W przypadku, gdy nie ma wyraźnej potrzeby ich stosowania, wynikającej z celu badania, należy je pominąć, gdyż mogą spowodować otrzymanie klasyfikacji gorszej.

3.2.4. Inne metody rozwiązywania zagadnienia klasyfikacji.

Do rozwiązania zagadnienia klasyfikacji mogą służyć również inne /oprócz metod taksonomii numerycznej/ metody. Przypomnijmy, że metody taksonomii numerycznej bazują bezpośrednio na odległościach między obiektami i polegają bądź na

grupowaniu klas, bądź na dzieleniu klas, bądź na przenoszeniu obiektów z jednej klasy do innej.
Spośród innych metod czy podejść wykorzystywanych w rozwiązywaniu zagadnienia klasyfikacji zasygnalizować należy przede wszystkim metody analizy czynnikowej. Opisane w rozdziale 2 zadanie rozwiązywane za pomocą tych metod polega na zastąpieniu zmiennych /opisanych przez n wartości zaobserwowanych w obiektach/ przez czynniki /również opisane przez n wartości/. Podejście to nazywane jest techniką R analizy czynnikowej /por. [64] ,[101] ,[107] /. Spotykane jest również inne podejście, tzw. technika Q. W podejściu tym obiekty i zmienne zamienione są rolami. W takim przypadku obiekty /opisane przez m wartości zmiennych/ zastępowane są przez " czynniki ", które można interpretować jako klasy. Podobnie jak w technice R ładunek czynnikowy jest to współczynnik korelacji między zmienną a czynnikiem, a zatem podobieństwo zmiennej do czyn tak w technice Q ładunek czynnikowy można interpretować jak stopień podobieństwa /z tym, że jest to podobieństwo kształ obiektu do klasy. Korzystając z macierzy ładunków czynnikowych można uzyskać klasyfikację, poprzez przyporządkowanie obiektowi tej klasy, dla której ładunek czynnikowy jest największy. Przykłady praktycznych rozwiązań za pomocą tego sposobu można znaleźć w [26],[31] ,[32] ,[101].
Inne ważne podejście do zagadnienia klasyfikacji wykorzystuje metody programowania matematycznego. Zaliczyć tu trzeba przede wszystkim metodę programowania całkowitoliczbowego opisaną szeroko w [120],[143]. Jako zmienne decyzyjne przyjęto przynależność obiektów do klas / ilość klas musi być dana/, zatem mogą one przyjmować wartość 0 lub 1. Jako funkcję celu przyj-

muje się np. minimum sumy wewnątrzgrupowych odległości, lub podobne, wynikające z warunków zagadnienia klasyfikacji, natomiast ograniczenia wynikają z faktu, iż obiekt może należeć tylko do jednej klasy. Sporą wadą tej metody, jakkolwiek dość intuicyjnej, jest jej olbrzymia pracochłonność, nawet w warunkach stosowania EMC.

Z innych metod programowania matematycznego stosowanych do rozwiązania zagadnienia klasyfikacji wspomnieć można o programowaniu dynamicznym /por. [72] /.

Bodaj jedną z najstarszych metod klasyfikacji /przez niektórych autorów zaliczana również do metod taksonomicznych/ jest metoda wizualna Czekanowskiego /por. [114] /. Polega ona na porządkowaniu przypadkowo otrzymanej klasyfikacji przy użyciu tylko oznaczeń graficznych odpowiadających różnym stopniom wzajemnych podobieństw między obiektami. Często stosowana w badaniach jest również tzw. metoda prostopadłościanów /por. [82] /.

3.3. Metody regionalizacji.

W niniejszym punkcie przedstawiona zostanie charakterystyka metod regionalizacji, tzn. metod realizujących sformułowane w 3.1. zagadnienie regionalizacji.

W literaturze przedmiotu nie zanotowano wielu propozycji w zakresie metod regionalizacji. Zaliczyć można do nich zaprezentowane w [21] ,[26]

Porównanie zagadnienia klasyfikacji i regionalizacji, a także analiza metod klasyfikacji pozwalają stwierdzić, iż metody klasyfikacji w swej idei bazują na wzajemnych podobieństwach obiektów /z reguły podobieństwo obrazowane jest przez odległość/. W przypadku zagadnienia regionalizacji oprócz wzajemnego podobieństwa należy przy konstruowaniu regionalizacji uwzględnić przyległość. Co więcej, warunek przyległości jest w zasadzie warunkiem dopuszczalności, ograniczającym zbiór klasyfikacji, w którym należy poszukiwać tej najlepszej.

Można zatem do rozwiązania zagadnienia regionalizacji adaptować metody klasyfikacji, oczywiście z uwzględnieniem warunku przyległości.

Poniżej zostaną przedstawione konkretne rozwiązania adaptacyjne w odniesieniu do metod klasyfikacji.

1. Hierarchiczne metody grupowania.

Jak pamiętamy, w przypadku tych metod, w każdym etapie hierarchicznego grupowania dokonuje się grupowania dwóch klas o najmniejszej odległości. Aby móc te metody zastosować do zagadnienia regionalizacji, przy grupowaniu należy rozważać jedynie klasy przylegające do siebie /przy czym dwie klasy przylegają do siebie wtedy i tylko wtedy, gdy istnieje para obiektów sąsiadujących ze sobą, przy czym każdy z tych obiektów należy do jednej z tych klas/ tzn.

- w i-tym etapie hierarchicznego grupowania zostają połączone takie dwie klasy S_j^{i-1}, S_k^{i-1}, dla których:

$$d\,/S_j^{i-1},\ S_k^{i-1}/ = \min_{\substack{S_p^{i-1},\ S_r^{i-1} \\ c\,/S_p^{i-1},\ S_r^{i-1}/ = 1}} d\,/S_p^{i-1},\ S_r^{i-1}/\ ,$$

przy czym:

$$c\,/S_p^{i-1},\ S_r^{i-1}/ = \begin{cases} 1 \text{ gdy } \exists P_{l_1} \in S_p^{i-1},\ \exists P_{l_2} \in S_r^{i-1} \\ \qquad c\,/P_{l_1},\ P_{l_2}/ = 1 \\ 0 \text{ gdy } \forall P_{l_1} \in S_p^{i-1} \forall P_{l_2} \in S_r^{i-1} \\ \qquad c\,/P_{l_1},\ P_{l_2}/ = 0 \end{cases}$$

2. Hierarchiczne metody podziału.

W odniesieniu do tej grupy metod nie zawsze jest możliwe wprowadzenie przyległości. Jest to możliwe jedynie wtedy, gdy w trakcie procesu podziału występuje zjawisko sekwencyjnego dołączania obiektów klasy dzielonej do jednej z klas nowo powstających /jest tak np. w przypadku metody odległości maksymalnej i odległości minimalnej/. Wtedy dany obiekt przydzielany jest do bliższej z dwóch nowo powstających klas pod warunkiem, że obiekt ten sąsiaduje z tą klasą /tzn. istnieje przynajmniej jeden obiekt należący do tej klasy, który przylega do danego obiektu/. W przypadku nie spełnienia tego warunku rozpatrywane są kolejne obiekty. Po zakończeniu tego procesu następuje ponowne rozpatrywanie nie rozdzielonych /z powodu nie spełniania warunku przyległości/ obiektów. Jeśli którykolwiek z tych obiektów w dalszym ciągu nie przylega do najbliższej klasy, tworzy osobną klasę. Jak łatwo zauważyć, w tym przypadku proces hierarchicznego podziału może się odbywać w trakcie mniejszej niż n-1 ilości etapów, gdyż w jednym etapie klasa może zostać podzielona na więcej niż dwie klasy.

3. Metody iteracyjne bazujące na funkcji jakości klasyfikacji. Adaptacja tych metod do zagadnienia regionalizacji doty-

czy dwóch punktów:

a/ wyboru klasyfikacji początkowej /a ściślej mówiąc regionalizacji początkowej, bowiem klasy wchodzące w skład klasyfikacji początkowej muszą być regionami/.

b/ sposobu przenoszenia obiektu z jednego regionu do innego.

Należy zwrócić uwagę, iż żadna z zaprezentowanych w punkcie 2.2.3. propozycji odnośnie wyboru klasyfikacji początkowej nie może być stosowana, gdyż nie jest w nich uwzględniony warunek przyległości.

Obecnie zaproponowana zostanie prosta procedura konstruowania regionalizacji początkowej:

- dla każdego obiektu P_i wyznacza się średnią odległość od obiektów do niego przyległych:

$$l_i = \frac{1}{s_i} \sum_{\substack{j=1 \\ c/P_i, P_j/=1}}^{s_i} d\,/P_i,\ P_j/$$

gdzie: $s_i = \overline{\text{card}} \left\{ P_k \quad c\,/P_i,\ P_k/ = 1 \right\}$

- k obiektów o najmniejszych wartościach l_i są to tzw. punkty skupienia, każdy z nich zostaje przydzielony do jednego regionu.

- pozostałe n-k obiektów zostaje kolejno rozdzielone do regionów wg. zasady: każdy obiekt zostaje przydzielony do najbliższego przyległego mu regionu /przy czym obiekt przylega do regionu, gdy przylega do co najmniej jednego obiektu należącego do regionu/

$$t\,/P_i/ = S_j^o \Leftrightarrow d\,/P_i,\ S_j^o/ = \min_{\substack{S_l^o \in S^o \\ c\,/P_i,\ S_l^o/ = 1}} d\,/P_i,\ S_l^o/$$

- procedurę rozdziału kontynuuje się aż do chwili przydzielenia wszystkich obiektów do określonych regionów /należy zaznaczyć, że odbywać się to może w kilku etapach, gdyż przy niewielkiej liczbie k, duża część obiektów w pierwszych etapach rozdzielania nie sąsiaduje z żadnym z utworzonych regionów/.

Procedura powyższa, opierająca się zarówno na podobieństwie, jak i przyległości, powinna gwarantować otrzymanie regionalizacji początkowej niezbyt odległej od optymalnej.

Natomiast w przypadku przenoszenia obiektów z jednej klasy do innej w celu zmniejszenia wartości funkcji jakości klasyfikacji należy wprowadzić następującą modyfikację, dotyczącą punktu 2 algorytmu metody:

W i-tej iteracji bada się kolejno obiekty P_1, P_2...itd., aż do znalezienia obiektu $P_j \in S_r^{i-1}$, dla którego istnieje klasa S_l^{i-1}, taka że:

1. $c\,/P_j,\ S_l^{i-1}/ = 1$

2. a/ dla metod o stałej ilości klas:

$$h_k\,/\{S_1^{i-1},\ S_2^{i-1},\dots,\ S_l^{i-1} \cup P_j,\dots,\ S_r^{i-1} - P_j,\dots,\ S_k^{i-1}\}/ <$$
$$h_k\,/\{S_1^{i-1},\ S_2^{i-1},\dots,\ S_k^{i-1}\}/$$

b/ dla metod o zmiennej ilości klas:

b1/ $h_k\,/\{S_1^{i-1},\ S_2^{i-1},\dots,\ S_l^{i-1} \cup P_j,\dots,\ S_r^{i-1} - P_j,\dots,$
$S_k^{i-1}\}\,/ < h_k\,/\{S_1^{i-1},\ S_2^{i-1},\dots,\ S_k^{i-1}\}\,/$

lub:

b2/ $h_{k+1}\,/\{S_1^{i-1},\ S_2^{i-1},\dots,\ S_r^{i-1} - P_j,\dots,\ S_k^{i-1},\ S_l^{i-1}\}/$
$< h_k\,/\{S_1^{i-1},\ S_2^{i-1},\dots,\ S_k^{i-1}\}\,/$

gdzie: $S_1^{i-1} = P_j$.

W przypadku, gdy zachodzi warunek a, obiekt P_j zostaje przydzielony do tej klasy S_s^{i-1}, która spełnia warunek:

$$h_k / \{S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots, S_s^{i-1} \cup P_j, \ldots, S_k^{i-1}\} / =$$

$$= \min_{\substack{l \\ c/S_l^{i-1}, P_j/ = 1}} h_k / \{S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots, S_l^{i-1} \cup P_j, \ldots, S_k^{i-1}\} /$$

i tworzy się nową regionalizację $S^i = \{S_1^i, S_2^i, \ldots, S_k^i\}$,

gdzie:

$S_r^i = S_r^{i-1} - P_j$; $S_s^{i-1} = S_s^{i-1} \cup P_j$; $S_l^i = S_l^{i-1}$ /$l \neq r$, $l \neq s$/

Natomiast, gdy zachodzi warunek b1 lub b2, obiekt P_j zostaje przydzielony do klasy S_s^{i-1}, jeśli:

$$h_k / \{S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots, S_s^{i-1} \cup P_j, \ldots, S_k^{i-1} /$$

$$= \min_{\substack{l \\ c/S_l^{i-1}, P_j/ = 1}} h_k / S_1^{i-1}, S_2^{i-1}, \ldots, S_r^{i-1} - P_j, \ldots,$$

$$\ldots, S_l^{i-1} \cup P_j, \ldots, S_k^{i-1} / \leqslant h_{k+1} / \{S_1^{i-1}, S_2^{i-1}, \ldots$$

$$\ldots, S_r^{i-1} - P_j, \ldots, S_k^{i-1}, S_l^{i-1}\} /$$

a do klasy S_l^{i-1} /gdzie $S_l^{i-1} = P_j$/, gdy zachodzi nierówność o przeciwnym zwrocie.

Tworzy się następnie nową regionalizację

$S^i = \{S_1^i, S_2^i, \ldots, S_k^i\}$ lub $S^i = \{S_1^i, S_2^i, \ldots, S_{k+1}^i\}$

/w zależności od zwrotu nierówności/, gdzie:

$S_r^i = S_r^{i-1} - P_j$; $S_s^i = S_s^{i-1} \cup P_j$, $S_l^i = S_l^{i-1} = S_l^{i-1}$

$/l \neq r,\ l \neq s/,\ S^{i}_{k+1} = P_j$

4. Metody iteracyjne bazujące na punktach skupienia.

W odniesieniu do tej grupy metod wprowadzenie warunku przyległości jest możliwe tylko tam, gdzie rozdział obiektów do punktów skupienia odbywa się sekwencyjnie. Stosuje się wtedy procedurę zbliżoną do omówionego wyżej przypadku dla hierarchicznych metod podziału, tzn. obiekt przydzielany jest do klasy tylko wtedy, gdy znajduje się w niej obiekt do niego przyległy. W przeciwnym przypadku /po ponownym rozpatrzeniu/ tworzy osobną klasę.

Powyżej przedstawione rozważania zawierały propozycje adaptacji metod klasyfikacji do zagadnienia regionalizacji. Poniżej zaproponowana zostanie metoda regionalizacji. Wykorzystuje ona również pojęcie punktów skupienia.

Metoda ta przebiega w kilku etapach:

1. Dla każdego obiektu P_i wyznacza się średnią odległość od obiektów przyległych:

$$\bar{s}_i = \frac{1}{\bar{l}_i} \sum_{\substack{P_l \\ \bar{c}\,/P_i,\, P_l/ = 1}} d\,/P_i,\, P_l/$$

gdzie: $\bar{l}_i = \mathrm{card}\left\{ P_l : c\,/P_i,\, P_l/ = 1 \right\}$

2. Jako punkty skupienia, czyli tzw. "jądra regionu" /por. [26] / wybiera się obiekty o najmniejszych wartościach $\bar{s}_i$. Ilość tych obiektów może być ustalona z góry lub na podstawie następującego warunku:
   - obiekty, dla których $s_i < \bar{s}$ /lub $s_i < \bar{s} - s_s$/,

   gdzie:

$$\bar{s} = \frac{1}{n} \sum_{i=1}^{n} s_i$$

$$s_s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} /s_i - \bar{s}/^2}$$

rozpatruje się w kolejności rosnących wartości s_i. Obiekt staje się punktem skupienia, jeśli jego odległość od już utworzonych punktów skupienia jest nie mniejsza niż wielkość krytyczna d_o /gdzie d_o może być np. średnią odległością między obiektami zbioru P/.

Wybrane jako punkty skupienia obiekty są, jako najbardziej podobne do swoich sąsiadów, wyróżnikami tworzonych regionów. Każdy z tych obiektów tworzy jeden region.

3. Pozostałe obiekty zostają kolejno rozdzielone do regionów wg zasady:

obiekt zostaje przydzielony do najbliższego regionu, z którym sąsiaduje, pod warunkiem, że po przydzieleniu średnica powstałego regionu jest mniejsza od wielkości krytycznej, /tzw. warunek dostatecznego wewnętrznego podobieństwa/:

$$\max_{P_t \in S_j} d\, /P_i, P_t/ \leqslant r_o$$

gdzie r_o jest wielkością zadaną, np.

$$r_o = \frac{2}{n\,/n-1/} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d\, /P_i, P_j/$$

4. Proces rozdzielania kontynuuje się aż do rozpatrzenia wszystkich obiektów. Obiekty, które z powodu nie spełnienia warunku dostatecznego wewnętrznego podobieństwa nie mogą zostać dołączone do istniejących regionów, tworzą nowe regiony.

4. TEORIA ZBIORÓW ROZMYTYCH W ANALIZIE WIELOWYMIAROWEJ.

4.1. Podstawowe pojęcia teorii zbiorów rozmytych.

Jak pamiętamy, zagadnienie klasyfikacji z reguły rozwiązuje się przy pomocy metod taksonomii numerycznej /cluster analysis/. Metody te bazują na podobieństwie wzajemnym obiektów, mierzonym z reguły przez funkcję odległości.

Załóżmy, że na zbiorze P x P określona została relacja S, którą dla ustalenia uwagi nazwiemy również relacją podobieństwa. Dana para obiektów $/P_i, P_j/$ należy do tej relacji, gdy obiekt P_i jest podobny do obiektu P_j /relację tę można określić np. w sposób następujący:

$$/P_i, P_j/ \in S \Leftrightarrow d/P_i, P_j/ \leqslant d_0/.$$

Zauważmy, że relacja ta jest zwrotna i symetryczna. Niestety, nie posiada ona własności przechodniości. Z faktu, że obiekt P_j jest podobny do obiektu P_k, a obiekt P_k do obiektu P_l nie wynika, że obiekt P_j jest podobny do obiektu P_l. Natomiast przy stosowaniu części metod taksonomii numerycznej korzysta się z tej własności /np. w metodzie najbliższego sąsiada/.
Inną niedogodnością jest to, iż relacja podobieństwa jest binarna, tzn. obiekty mogą być podobne lub niepodobne. Wydaje się, że relacja podobieństwa powinna być bardziej "uszczegółowiona", np. przez wprowadzenie poziomów: "mało podobny", „średnio podobny", itp.
Do wyeliminowania zasygnalizowanych wad zaproponowane zostanie poniżej inne podejście, będące pewnym uogólnieniem zagadnienia klasyfikacji. Podejście to zostało oparte na teorii zbiorów

rozmytych. Twórcą tej teorii jest L. A. Zadeh. Jej zarys nakreślił w [151]. Teoria ta była stopniowo rozwijana, a znalazła zastosowanie zwłaszcza w elektronice, automatyce, lingwistyce. Najciekawsze prace teoretyczne z dziedziny tej teorii to [23], [62][63], [75], [104], [153].

Poniżej przedstawione zostanie kilka najważniejszych pojęć teorii zbiorów rozmytych. Zaznaczyć należy, iż pojęcia zbioru rozmytego stanowi pewne uogólnienie pojęcia zbioru.

Przez $\mathfrak{X}$ oznacza się pewien zbiór, tzw. universum /jest to np. zbiór wszystkich obiektów, zbiór poza zakres którego nie wychodzi się w rozpatrywaniu danego problemu/. Na tym universum określa się zbiory rozmyte.

Df. Zbiór rozmyty /inaczej podzbiór rozmyty/ A na universum $\mathfrak{X}$ jest to zbiór par uporządkowanych:

$\{/X, f_A/X//\}$, gdzie:

$X \in \mathfrak{X}$

$f_A : \mathfrak{X} \to \langle 0; 1\rangle$

Odwzorowanie f_A nazywa się funkcją stopnia przynależności, lub krótko: funkcją przynależności /ang. membership function/ do zbioru rozmytego A. I tak, f_A /X/ jest to stopień przynależności elementu X do zbioru rozmytego A.

Często zbiór rozmyty zapisuje się następująco:

$$A = \int_{X \in \mathfrak{X}} (X \mid f_A /X/)$$

przy czym znak $\int$ należy rozumieć jako sumę mnogościową.

Jak widać z powyższej definicji, dla każdego elementu należącego do universum określona jest funkcja, przyjmująca wartości przedziału $\langle 0; 1\rangle$ Gdy np. $f_A /X_1/ = 0,8$, wtedy mówi się, że element X_1 należy do zbioru rozmytego A w stopniu 0,8.

Koncepcja zbioru rozmytego jest zatem w pewnym sensie uogólnieniem czy rozszerzeniem koncepcji zbioru /z teorii mnogości/. W przypadku zbioru, element należy do niego lub nie, natomiast do zbioru rozmytego element może należeć w większym lub mniejszym stopniu.
Stosując tę konwencję, każdy zbiór /w zwykłym sensie/ jest również zbiorem rozmytym. Jest bowiem:

$$A = \int_{X \in \mathcal{X}} /X \mid f_A /X//$$

gdzie:

$$f_A /X/ = \begin{cases} 1 & \text{gdy } X \in A \\ 0 & \text{gdy } X \notin A \end{cases}$$

W przypadku zbioru w zwykłym sensie zawartość informacyjna pojęcia "należenie" jest mniejsza /gdyż przyjmuje tylko 2 wartości/ niż w przypadku zbioru rozmytego. Sugeruje to, iż koncepcja rozmytości może być użyteczna przy formalizacji pewnych prawidłowości występujących np. w naukach społecznych.
Obecnie wprowadzone zostaną podstawowe działania na zbiorach rozmytych. Można je definiować wtedy, gdy zbiory rozmyte są określone na tym samym universum /oczywiście wtedy zbiór rozmyty będący wynikiem działania jest również określony na tym samym universum, natomiast działanie definiowane jest przez postać funkcji przynależności do tego zbioru/.

Df. Dopełnieniem zbioru rozmytego A na X jest zbiór rozmyty A' na $\mathcal{X}$, którego funkcja przynależności jest równa:

$$f_{A'} /X/ = 1 - f_A /X/$$

W poniższej tablicy zebrane są definicje innych działań z podaniem jednocześnie odpowiednika dla zwykłych zbiorów.

Nazwa działania	Symbol	Wartość f. przynależności	Odpowiednik dla zbiorów w zwykłym sensie
Suma	$C=A \cup B$	$f_C/X/=\max\{f_A/X/, f_B/X/\}$	$A \cup B$
iloczyn	$C=A \cap B$	$f_C/X/=\min\{f_A/X/, f_B/X/\}$	$A \cap B$
suma ograniczona	$C=A \oplus B$	$f_C/X/=\min\{1, f_A/X/+f_B/X/\}$	$A \cup B$
suma prosta	$C=A \dot{+} B$	$f_C/X/=f_A/X/+f_B/X/-f_A/X/\cdot f_B/X/$	$A \cup B$
iloczyn algebr.	$C=AB$	$f_C/X/=f_A/X/\cdot f_B/X/$	$A \cap B$
różnica ogranicz.	$C=A \ominus B$	$f_C/X/=\max\{0, f_A/X/-f_B/X/\}$	$A - B$
różnica absolutna	$C=\|A-B\|$	$f_C/X/=\|f_A/X/-f_B/X/\|$	$/A \cap B/'$

Na podstawie powyższych definicji można określić jeszcze dwa inne działania:

1. różnica $C = A - B$

$$C = A \cap B'$$

2. Suma rozłączna $C = A \not\Sigma B$

$$C = /A \cap B'/ \cup /B \cap A'/$$

Podobnie jak dla zbiorów w zwykłym sensie, tak i dla zbiorów rozmytych słuszne są: prawa przemienności sumy i iloczynu, prawa łączności sumy i iloczynu, prawa rozdzielności sumy względem iloczynu oraz iloczynu względem sumy oraz prawa de Morgana.

4.2. Klasyfikacja rozmyta i sposoby jej wyznaczania.

W niniejszym paragrafie zostanie przedstawione uogólnienie zagadnienia klasyfikacji przy zastosowaniu niektórych pojęć teorii zbiorów rozmytych. Pewne idee tego podejścia zostały zaczerpnięte z [152]

Załóżmy, że dany jest zbiór obiektów P, liczący n elementów. Każdy z tych obiektów charakteryzowany jest przez wartości m zmiennych: $X_1, X_2, \ldots, X_m$. Dana jest również liczba naturalna k /$1 < k < n$/, oznaczająca ilość klas.

Zagadnienie klasyfikacji rozmytej bezwzględnej.

Na danym zbiorze P należy określić rodzinę zbiorów rozmytych /klas/: $S_1, S_2, \ldots, S_k$, tak aby spełnione były warunki:

1. $S_1 \oplus S_2 \oplus \ldots \oplus S_k = \bar{P}$,

   gdzie $\bar{P}$ jest tzw. pełnym zbiorem rozmytym, tzn.

   $\bar{P} = \{/P_i \mid f_{\bar{P}}/P_i/ : \quad P_i \in P, \; f_{\bar{P}}/P_i/ = 1 \quad \forall P_i \in P$

2. Obiekty, dla których wartości funkcji przynależności do tego samego zbioru rozmytego są duże - są jak najbardziej podobne, natomiast wartości funkcji przynależności do różnych zbiorów rozmytych są duże - są jak najmniej podobne.

Zagadnienie klasyfikacji rozmytej względnej.

Na danym zbiorze P należy określić rodzinę zbiorów rozmytych /klas/: $S_1, S_2, \ldots, S_k$, tak aby spełnione były warunki:

1. $S_1 \oplus S_2 \oplus \ldots \oplus S_k = \bar{P}$
2. Obiekty, dla których wartości funkcji przynależności do tego samego zbioru rozmytego są duże - są jak najbardziej podobne, natomiast wartości funkcji przynależności do różnych zbiorów rozmytych są duże - są jak najmniej podobne.
3. $\forall P_i \in P \quad \sum_{j=1}^{k} f_{S_j} /P_i/ = 1$

Jak wynika z powyższych określeń, zarówno bezwzględna jak i względna klasyfikacja rozmyta charakteryzowana jest przez macierz $F = \left\{ \overline{f_{ij}} \right\}$, gdzie:

$\overline{f_{ij}} = f_{S_j} /P_i/ \qquad i = 1,2,\ldots, n \qquad j = 1,2,\ldots, k.$

W przypadku klasyfikacji rozmytej względnej, suma elementów każdego wiersza równa się 1.
Poszczególne wartości f_{S_j} /P_i/ interpretować można jako stopień przynależności obiektu P_i do klasy S_j /waha się on od 0 do 1/, przy czym w przypadku klasyfikacji rozmytej względnej ów stopień przynależności jest to udział przynależności do danej klasy w ogólnej przynależności do wszystkich klas.
Dla klasyfikacji w zwykłym sensie, stopień przynależności obiektu do klasy jest zmienną binarną, przyjmującą jedną z dwóch wartości: 0 lub 1 /w zależności od tego czy obiekt należy do klasy, czy nie/. W tym przypadku, ogólniejszym, jest zmienną ciągłą /a w praktyce zmienną dyskretną prezentowaną na skali porządkowej/.

Jak się wydaje, klasyfikacja rozmyta może być przydatna szczególnie tam, gdzie brak jest "ostrych" granic między klasami, gdzie trudno jednoznacznie przydzielić obiekt do jednej klasy, tzn. gdy brak jest klasyfikacji naturalnej.
Wyznaczanie klasyfikacji rozmytej może się odbywać za pomocą trzech podstawowych metod:

1. metoda wykorzystująca rezultaty klasyfikacji w zwykłym sensie,
2. metoda iteracyjna /por.[12],[152],
3. metoda bazująca na funkcjach obiektywnych /por.[124]/.

ad 1. Jest to najprostsze rozwiązanie, wymagające uprzedniego skonstruowania klasyfikacji w zwykłym sensie.
Jak wiadomo, stopień przynależności obiektu do klasy można wyrazić przez stopień podobieństwa obiektu do tej klasy.
Metoda ta zakłada, że klasy są znane, że klasy rozmyte są pochodną klas zwykłych.

Dla klasyfikacji rozmytej bezwzględnej wyróżnić można 2 warianty tej metody:

1. $f_{S_j}/P_i/ = s/P_i, G_j/ = \frac{1}{1 + d/P_i, G_j/}$ /4.1/

2. $f_{S_j}/P_i/ = \frac{1}{n_j} \sum_{P_k \in S_j} s/P_i, P_k/ = \frac{1}{n_j} \sum_{P_k \in S_j} \frac{1}{1 + d/P_i, P_k/}$ /4.2/

gdzie: G_j - środek ciężkości j-tej klasy

n_j - liczebność j-tej klasy

W wariancie pierwszym podobieństwo obiektu do klasy utożsamia się z podobieństwem obiektu do środka ciężkości tej klasy /jako jej najlepszego reprezentanta/, natomiast w wariancie drugim ze średnim podobieństwem do obiektów należących do tej klasy.

W podobny sposób można otrzymać rozmytą klasyfikację względną /korzystając z możliwości przekształcenia klasyfikacji bezwzględnej na względną poprzez wyznaczenie udziałów podobieństw/:

1. $f_{S_j}/P_i/ = \frac{s/P_i, G_j/}{\sum_{l=1}^{k} s/P_i, G_l/}$ /4.3/

2. $f_{S_j}/P_i/ = \frac{\frac{1}{n_j} \sum_{\substack{P_k \in S_j \\ P_k \neq P_i}} s/P_i, P_k/}{\sum_j \frac{1}{n_j} \sum_{\substack{P_k \in S_j \\ P_k \neq P_i}} s/P_i, P_k/}$ / 4.4/

3. $f_{S_j}/P_i/ = \frac{\sum_{P_k \in S_j} s/P_i, P_k/}{\sum_{P_l \in P} s/P_i, P_l/}$ /4.5/

Metoda ta, jakkolwiek prosta i intuicyjna, zawiera jedną poważną wadę. Podstawowym determinantem klasyfikacji rozmytej jest klasyfikacja w zwykłym sensie, zatem istnieje niebezpieczeństwo przeniesienia niedoskonałości tej klasyfikacji na klasyfikację rozmytą. Po prostu klasy rozmyte są pochodnymi zwykłych klas, już uprzednio wyznaczonych.

ad 2. Metoda iteracyjna - algorytm:

1. Dana jest rozmyta klasyfikacja początkowa /bezwzględna lub względna/:

   $f^0_{S_j}/P_l/ \quad l = 1,2,\ldots, n \quad j = 1,2,\ldots, k$

   Klasyfikację tę można otrzymać losowo lub za pomocą jednego ze sposobów przedstawionych w poprzednim punkcie.

2. W i-tej iteracji wyznacza się nową klasyfikację rozmytą poprzez przekształcenie poprzednio otrzymanej za pomocą jednego z następujących sposobów:

   - dla klasyfikacji bezwzględnej:

     a/ $f^i_{S_j}/P_l/ = \bar{s}/P_l, G^{i-1}_j/$

     gdzie:

     $$\bar{x}_{G^{i-1}_{l,j}} = \frac{\sum_{v=1}^{n} \bar{x}_{vj}\, f^{i-1}_{S_l}/P_v/}{\sum_{v=1}^{n} f^{i-1}_{S_l}/P_v/}$$

     b/ $$f^i_{S_j}/P_l/ = \frac{\sum_{v=1}^{n} f^{i-1}_{S_j}/P_v/\quad \bar{s}/P_v,P_l/}{\sum_{v=1}^{n} f^{i-1}_{S_j}/P_v/}$$

   - dla klasyfikacji względnej

     a/ $$f_{S_j}/P_l/ = \frac{s/P_l,\, G^{i-1}_j/}{\sum_{t=1}^{k} \bar{s}/P_l, G^{i-1}_t/}$$

b/ $$f^{i}_{S_j}/P_l/ = \frac{\sum_{v=1}^{n} f^{i-1}_{S_j}/P_v/\, s/P_v,P_l/}{\sum_{v=1}^{n} f^{i-1}_{S_j}/P_v/} : \sum_{t=1}^{k} \frac{\sum_{v=1}^{n} f^{i-1}_{S_t}/P_v/\cdot s/P_v,P_l/}{\sum_{v=1}^{n} f^{i-1}_{S_t}/P_v/}$$

3. Postępowanie kończy się w iteracji r-tej, gdy:

$$\max_{i,j} \left| f^{r+1}_{S_j}/P_i/ - f^{r}_{S_j}/P_i/ \right| \leqslant \varepsilon \qquad \varepsilon > 0$$

przy czym ε jest bliskie zeru.

Idea metody iteracyjnej jest zbliżona do idei metod iteracyjnych bazujących na punktach skupienia /przy konstruowaniu klasyfikacji w zwykłym sensie/.

W kolejnych iteracjach klasyfikacja rozmyta jest zmieniana, nowa klasyfikacja wynika z podobieństwa obiektu do klasy otrzymanej w poprzedniej iteracji, przy czym podobieństwo to /podobnie jak przy poprzedniej metodzie/ określane jest jako podobieństwo obiektu do środka ciężkości /centroidu/ klasy lub średnie podobieństwo do obiektów danej klasy. Należy zwrócić uwagę, że podobieństwa te liczone są jako średnie ważone, przy czym wagami są stopnie przynależności wyznaczone w poprzedniej iteracji. Postępowanie kończy się, gdy klasyfikacja przestanie zmieniać się w stopniu znaczącym.

ad 3. W metodzie tej zastosowane są metody programowania matematycznego. Funkcja celu jest to tzw. funkcja obiektywna /por. [124]/ mierząca wewnętrzne podobieństwo klasyfikacji rozmytej. Stosować można do tego celu następujące funkcje:

$$\sum_{j=1}^{k} \sum_{i} \sum_{l} f^{2}_{S_j}/P_i/\, f^{2}_{S_j}/P_l/ \quad d/P_i,P_l/$$

$$\sum_{j=1}^{k} \sum_{i} f^{2}_{S_j}/P_i/ \quad d/P_i,O_j/$$

Wyznaczenie klasyfikacji rozmytej /przy czym metoda pozwala na otrzymanie jedynie klasyfikacji względnej/ polega na optymalizacji programu kwadratowego:

$$\sum_j \sum_i f_{S_j} /P_i/ \quad D/P_i, S_j/ \rightarrow \min.$$

gdzie: $D/P_i,S_k/ = \sum_l f^o_{S_j} /P_l/ \cdot d/P_i,P_l/$

lub: $D/P_i,S_k/ = \sum_l d/P_i,G_l/$

przy warunkach:

$$f_{S_j} /P_i/ \geqslant 0 \quad \forall i,j$$

$$\sum_j f_{S_j} /P_i/ = 1$$

Jak widać, tu również konieczna jest znajomość klasyfikacji początkowej.

Powyższy program rozwiązać można np. za pomocą mnożników Lagrange'a, optymalizując:

$$\sum_j \sum_i f^2_{S_j} /P_i/ \quad D/P_i,S_j/ + \sum_i \lambda /P_i/ \left[\sum_j f_{S_j} /P_i/ - 1\right]$$

Licząc pochodne cząstkowe po $f_{S_k} /P_i/$ oraz $\lambda /P_i/$, których jest $nk + n$, otrzymuje się układ równań, który po rozwiązaniu daje wartości zmiennych decyzyjnych, którymi są tutaj stopnie przynależności obiektów do klas.

Powyżej przedstawione trzy metody posiadają jedną wadę czy niedogodność. Do ich stosowania musi być znana ilość klas. W pierwszej z prezentowanych metod wynika ona z klasyfikacji w zwykłym sensie jaka powinna być uprzednio utworzona.

Do wyznaczenia ilości klas w klasyfikacji rozmytej można zaproponować inny sposób, wykorzystujący pojęcie punktów skupienia, a nie wymagający tworzenia klasyfikacji w zwykłym sensie:

1. Dla każdego obiektu wyznacza się średnią odległość od pozostałych obiektów

$$\bar{d}_i = \frac{1}{n-1} \sum_{j=1}^{n} d/P_i,P_j/$$

2. Obiekty szereguje się wg rosnących wartości $\bar{d}_i$.

3. Uszeregowane obiekty rozpatruje się kolejno, pierwszym punktem skupienia staje się obiekt o najmniejszej wartości średniej odległości od pozostałych obiektów.

Następnymi punktami skupienia są kolejno rozpatrywane obiekty, pod warunkiem, że ich odległość od dotychczas wyznaczonych punktów skupienia jest równa co najmniej d_o /gdzie d_o jest to odległość krytyczna równa np. średniej odległości w zbiorze obiektów/.

Trzeci etap procedury zapewnia, że wybrane punkty skupienia będą się w dostateczny sposób różnić od siebie.

Następnie wyznacza się klasyfikację rozmytą wg wzoru:

$$f_{S_j}/P_i/ = \bar{\bar{s}}/P_i,G_j/ \qquad \text{dla klasyfikacji bezwzględnej}$$

$$f_{S_j}/P_i/ = \frac{\bar{\bar{s}}/P_i,G_j/}{\sum_l \bar{\bar{s}}/P_i,G_l/} \qquad \text{dla klasyfikacji względnej}$$

gdzie: G_l oznacza l-ty punkt skupienia.

Tak wyznaczona klasyfikacja rozmyta może służyć jako klasyfikacja początkowa w metodach iteracyjnych.

4.3. Zastosowanie pojęcia rozmytości do analizy i prognozowania zjawisk.

Podejście, które zostanie zaprezentowane w niniejszym paragrafie, wykorzystuje rozmytą klasyfikację względną i może być przydatne przy analizie i prognozowaniu zjawisk. Bazuje ono na pojęciu zmiennej rozmytej /por.[153]/, które jest uogólnieniem pojęcia zmiennej. Każda zmienna charakteryzowana jest przez zbiór wartości, jakie przyjmuje. Zmienna rozmyta natomiast charakteryzowana jest przez zbiór wartości oraz stopnie, w jakich te wartości są przyjmowane.

Df. Zmienna rozmyta jest to zbiór rozmyty określony na zbiorze wartości tej zmiennej. Załóżmy, że badane jest n obiektów ze względu na m zmiennych i wyznaczona została rozmyta klasyfikacja względna:

$$f_{Sj} /P_i/ \qquad i = 1,\dots, n \qquad j = 1,\dots, k$$

Zauważmy, że rozpatrując każdą klasę /dla przykładu j-tą/ można utworzyć m zmiennych rozmytych: $X_1, X_2,\dots, X_m$, postaci:

$$X_l = \{ (x_{il} \mid f_{Xl} /x_{il}/) \mid i = 1,\dots,n, \quad f_{Xl} /x_{il}/ = f_{Sj} \ /P_i/ \}$$

Jak widać, taka zmienna rozmyta jest zbiorem rozmytym, gdzie stopnie przynależności poszczególnych wartości do tej zmiennej /zbioru/ równe są stopniom przynależności obiektów, w których te wartości zostały zaobserwowane, do klasy rozpatrywanej.

Dla każdej zmiennej stopnie przynależności odpowiednich wartości są takie same, jest

$$f_{Xl} /x_{il}/ = f_{Xr} /x_{ir}/$$

Mamy zatem do czynienia z m zmiennymi, z których każda przyjmuje n wartości w pewnym stopniu /stopień ten można interpretować jako prawdopodobieństwo/.

Oznaczmy, przez x_{ij} wartość j-tej zmiennej dla i-tego obiektu, a przez p_i stopień przyjmowania tej wartości. Zatem:

$$p_i = f_{S_j} \; /P_i/$$

Dla tak zdefiniowanej zmiennej rozmytej wprowadzimy określenie średniej lub wartości oczekiwanej. Jest ona równa:

$$E\,/X_j/ = \frac{\sum_{i=1}^{n} x_{ij}\, p_i}{\sum_{i=1}^{n} p_i} = \bar{X}_j$$

Korzystając z klasycznej definicji wariancji i kowariancji można otrzymać wzory na wariancję i kowariancję dla zmiennych rozmytych:

$$\mathrm{var}\,/X_j/ = E\,/X_j - E\,/X_j//^2 = \frac{\sum_{i=1}^{n} /x_{ij} - \bar{X}_j/^2\, p_i}{\sum_{i=1}^{n} p_i} =$$

$$= \frac{\sum_{i=1}^{n} /x_{ij} - \bar{X}_j/^2\, p_i}{\sum_{i=1}^{n} p_i} = \frac{\sum_{i=1}^{n} x_{ij}^2\, p_i - 2\bar{X}_j \sum_{i=1}^{n} x_{ij}\, p_i + /\bar{X}_j/^2 \sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} p_i} = \frac{\sum_{i=1}^{n} x_{ij}^2\, p_i - 2\bar{X}_j\bar{X}_j \sum_{i=1}^{n} p_i + /\bar{X}_j/^2 \sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} p_i} =$$

$$= \frac{\sum_{i=1}^{n} x_{ij}^2\, p_i}{\sum_{i=1}^{n} p_i} - /\bar{X}_j/^2 = \frac{\sum_{i=1}^{n} x_{ij}^2\, p_i}{\sum_{i=1}^{n} p_i} - \frac{/\sum_{i=1}^{n} x_{ij}\, p_i/^2}{/\sum_{i=1}^{n} p_i/^2}$$

$$\mathrm{cov}\,/X_j\, X_l/ = E\;/X_j - E\,/X_j//\;\; /X_l - E\,/X_l// =$$

$$= \frac{\sum_{i=1}^{n} /x_{ij} - \bar{X}_j/ \ /x_{il} - \bar{X}_l/ \ p_i}{\sum_{i=1}^{n} p_i} =$$

$$= \frac{\sum_{i=1}^{n} x_{ij} x_{il} p_i - \bar{X}_j \sum_{i=1}^{n} x_{il} p_i - \bar{X}_l \sum_{i=1}^{n} x_{ij} p_i + \bar{X}_j \bar{X}_l \sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} p_i} =$$

$$= \frac{\sum_{i=1}^{n} x_{ij} x_{il} p_i - \bar{X}_j \bar{X}_l \sum_{i=1}^{n} p_i - \bar{X}_l \bar{X}_j \sum_{i=1}^{n} p_i + \bar{X}_j \bar{X}_l \sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} p_i} =$$

$$= \frac{\sum_{i=1}^{n} x_{ij} x_{il} p_i}{\sum_{i=1}^{n} p_i} - \bar{X}_j \bar{X}_l = \frac{\sum_{i=1}^{n} x_{ij} x_{il} p_i}{\sum_{i=1}^{n} p_i} - \frac{\sum_{i=1}^{n} x_{ij} p_i \sum_{i=1}^{n} x_{il} p_i}{/\sum_{i=1}^{n} p_i/^2}.$$

Podobnie określić można współczynnik korelacji dla zmiennych rozmytych. Jest on równy:

$$\mathrm{cor}\ /X_j X_l/ = \frac{\sum_{i=1}^{n} x_{ij} x_{lj} p_i \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} x_{ij} p_i \sum_{i=1}^{n} x_{il} p_i}{\sqrt{/\sum_{i=1}^{n} x_{ij}^2 p_i \sum_{i=1}^{n} p_i - /\sum_{i=1}^{n} x_{ij} p_i/^2 \ /\sum_{i=1}^{n} x_{il}^2 p_i \sum_{i=1}^{n} p_i - /\sum_{i=1}^{n} x_{il} p_i/^2}} \qquad /4.6/$$

Wielkości te stosowane są do odpowiednich badań ilościowych dla zmiennych rozmytych, osobno w każdej klasie.

Oczywiście współczynniki korelacji między tymi samymi zmiennymi rozmytymi, ale wyznaczone w dwóch różnych klasach, różnią się.

Pokażemy obecnie sposób budowy dla zmiennych rozmytych liniowego modelu regresji.

Opiera się ona na warunku:

$$\sum_{i=1}^{n} p_i^2 e_i^2 \rightarrow \min.$$

gdzie e_i - reszta modelu.

Warunek ten można zapisać w postaci macierzowej.

Model ma postać $Y = X a$

gdzie:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \qquad X = \begin{bmatrix} x_{11} & x_{1r} \\ \cdot & \\ \cdot & \\ \cdot & \\ \cdot & \\ \cdot & \\ x_{n1} & x_{nr} \end{bmatrix}$$

r - ilość zmiennych objaśniających.

Warunek najmniejszych kwadratów ma postać:

$$S = /p\ e/'/p\ e/ \rightarrow \min.$$

gdzie:

$$p = \begin{bmatrix} p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 \ldots & 0 \\ \cdot & & & \\ 0 & 0 & 0 \ldots & p_n \end{bmatrix} \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

Mamy zatem:

$$S = /p\ e/'/p\ e/ = [p\ /y - X\ a/]'[p\ /y - X\ a/] =$$

$$= /p\ y - p\ X\ a/'/p\ y - p\ X\ a/ = /y'p' - a'X'p'/\ \ /p\ y - p\ X\ a/ =$$

$$= y'p'p\ y - y'p'p\ X\ a - a'X'p'p\ y + a'X'p'p\ X\ a$$

Po wyznaczeniu pochodnej:

$$\frac{\partial S}{\partial a} = - X'p'p\ y - X'p'p\ y + 2\ X'p'p\ X\ a = 0$$

$$- 2\ X'p'p\ y = - 2\ X'p'p\ X\ a$$

zatem:

$$a = /X'p'p\ X/^{-1}\ X'p'p\ y$$

Wzór ten będący uogólnieniem wzoru dla klasycznej regresji liniowej służy do znalezienia wartości współczynników modelu:

$$y = a_1\ X_1 + a_2\ X_2 + \ldots + a_r\ X_r + e$$

przy czym zmienne tego modelu są zmiennymi rozmytymi.

W najprostszym przypadku, tzn. dla modelu:

$$Y = a X + b + e$$

wzory na wartości współczynników przyjmują postać:

$$a = \frac{\sum_{i=1}^{n} x_i y_i p_i^2 \sum_{i=1}^{n} p_i^2 - /\sum_{i=1}^{n} x_i p_i^2/\ /\sum_{i=1}^{n} y_i p_i/^2}{\sum_{i=1}^{n} x_i^2 p_i^2 \sum_{i=1}^{n} p_i^2 - /\sum_{i=1}^{n} x_i p_i^2/^2}$$

$$b = \frac{\sum_{i=1}^{n} x_i^2 p_i^2 \sum_{i=1}^{n} y_i p_i^2 - \sum_{i=1}^{n} x_i y_i p_i^2 \sum_{i=1}^{n} x_i p_i^2}{\sum_{i=1}^{n} x_i^2 p_i^2 \sum_{i=1}^{n} p_i^2 - /\sum_{i=1}^{n} x_i p_i^2/^2}$$

Interpretacja modeli regresji dla zmiennych rozmytych jest podobna jak klasycznego modelu regresji. Współczynniki wyrażają wpływ zmiennych objaśniających na objaśnianą.

Powyższe podejście może być stosowane również wtedy, gdy zmienną objaśniającą nie jest żadna ze zmiennych z rozpatrywanego zbioru $X_1, X_2, \ldots, X_m$, lecz jakaś inna /lub inne, w przypadku większej ilości/ zmienna, wartości której nie miały wpływu na powstanie klasyfikacji rozmytej.

Budując podobne modele regresji rozmytej dla wszystkich klas uzyskamy z pewnością różne wyniki. Ponieważ klasy można interpretować jako zbiorowości podobne, zatem przez porównanie modeli regresji można porównać zależności jakie zachodzą między zmiennymi w tych zbiorowościach.

Tak jak i klasyczny model model regresji liniowej, tak i model otrzymany za pomocą powyżej opisanej procedury powinien być weryfikowany. Z powodu podejścia przyjętego w pracy /por. 2.1/ nie można tu stosować żadnych testów statystycznych. Do zbadania stopnia dopasowania modelu do rzeczywistości /do danych/ można natomiast stosować współczynnik zbieżności i współczynnik zmienności losowej /wyrazistości/. Poniżej przedstawione są wzory na te współczynniki w przypadku modeli dla klas rozmytych:

$$\varphi^2 = \frac{var\ /e/}{var\ /Y/} = \left[\frac{\sum_{i=1}^{n} e_i^2 p_i}{\sum_{i=1}^{n} p_i} - \frac{/\sum_{i=1}^{n} e_i p_i/^2}{/\sum_{i=1}^{n} p_i/^2}\right] : \left[\frac{\sum_{i=1}^{n} y_i^2 p_i}{\sum_{i=1}^{n} p_i} - \frac{/\sum_{i=1}^{n} y_i p_i/^2}{/\sum_{i=1}^{n} p_i/^2}\right] \quad /4.7/$$

$$W = \frac{\sqrt{var\ /e/}}{\bar{y}} = \sqrt{\left[\frac{\sum_{i=1}^{n} e_i^2 p_i}{\sum_{i=1}^{n} p_i} - \frac{/\sum_{i=1}^{n} e_i p_i/^2}{/\sum_{i=1}^{n} p_i/^2}\right]} : \frac{\sum_{i=1}^{n} y_i p_i}{\sum_{i=1}^{n} p_i} \quad /4.8/$$

Łatwo zauważyć, że oba te współczynniki przyjmują wartości nieujemne. Ponadto z reguły przyjmują wartości mniejsze od 1, zatem mogą posłużyć jako mierniki dobroci modelu, przy czym ich interpretacja jest podobna jak dla klasycznej regresji liniowej. Modele te mogą posłużyć również do prognozowania. Przydatna do tego może być następująca procedura:

Załóżmy, że zbudowane zostały modele regresji pewnej zmiennej Y dla wszystkich klas rozmytej klasyfikacji względnej.

$$Y^1 = a_1^1 X_1 + a_2^1 X_2 + \ldots + a_r^1 X_r + e^1$$

$$Y^2 = a_1^2 X_1 + a_2^2 X_2 + \ldots + a_r^2 X_r + e^2$$

- -

$$Y^k = a_1^k X_1 + a_2^k X_2 + \ldots + a_r^k X_r + e^k$$

W modelach tych górny indeks wskazuje numer klasy, dla której dany model został zbudowany.

Prognoza wartości zmiennej Y dla dowolnego obiektu jest zatem określona w sposób następujący:

$$\hat{y}_{pi}^{l} = a_1^l x_{pi1} + a_2^l x_{pi2} + \ldots + a_r^l x_{pir}$$

gdzie:

$\hat{y}_{pi}^{l}$ - wartość prognozowana zmiennej Y dla i-tego obiektu wyznaczona w oparciu o model dla l-tej klasy,

x_{pij} - wartość prognozowana zmiennej X_j dla i-tego obiektu.

Prognoza wartości zmiennej Y jest ważoną średnią prognoz wyznaczonych w oparciu o modele dla poszczególnych klas:

$$y_{pi} = p_1 y^1_{pi} + p_2 y^2_{pi} + \ldots + p_k y^k_{pi} = f_{S_1}/P_i/\ y^1_{pi} + f_{S_2}/P_i/\ y^2_{pi} +$$

$$+ f_{S_k}/P_i/\ y^k_{pi} = f_{S_1}/P_i/\ /\ a^1_1 x_{pi1} + a^1_2 x_{pi2} + \quad + a^1_r x_{pir}\ / +$$

$$f_{S_2}/P_i/\ /\ a^2_1 x_{pi1} + a^2_2 x_{pi2} + \quad + a^2_r x_{pir}\ / + \quad + f_{S_k}/P_i/\ /\ a^k_1 x_{pi1}$$

$$+ a^k_2 x_{pi2} + \ldots + a^k_r x_{pir}\ / = /\ f_{S_1}/P_i/\ a^1_1 + f_{S_2}/P_i/\ a^2_1 + \quad +$$

$$f_{S_k}/P_i/\ a^k_1\ /\ x_{pi1} + /\ f_{S_1}/P_i/\ a^1_2 + f_{S_2}/P_i/\ a^2_2 + \ldots + f_{S_k}/P_i/\ a^k_2\ /$$

$$x_{pi2} + \ldots + /\ f_{S_1}/P_i/\ a^1_r + f_{S_2}/P_i/\ a^2_r + \quad + f_{S_k}/P_i/\ a^k_r\ /\ x_{pir}$$

Przy konstruowaniu prognozy ostatecznej, prognozy wyznaczone w oparciu o modele dla poszczególnych klas ważą w ostatecznej prognozie w takim stopniu, w jakim obiekt, dla którego prognoza jest wyznaczana, należy do tej klasy.

5. ZASTOSOWANIE NIEKTÓRYCH METOD ANALIZY WIELOWYMIAROWEJ DO PRZESTRZENNYCH BADAŃ USŁUG W POLSCE.

5.1. Usługi - analiza pojęcia.

W rozdziale tym usługi zostaną zbadane od strony ich poziomu, a zatem głównie od strony podażowej. W Polsce sfera usług /w przeciwieństwie do szeroko rozumianej sfery produkcji/ była traktowana gorzej, zaniedbywana. Słuszne jest twierdzenie, iż popyt na wszelkiego rodzaju usługi nie jest w pełni zaspokojony /poza tym trudne jest jego bezpośrednie zbadanie, zwłaszcza w niektórych grupach usług/. Dlatego w pracy poziom czy jakość usług będzie w pewnym sensie utożsamiana z ich podażą.
Pojęcie usług jest wieloznaczne. Klasyczna jest definicja O. Langego, który przez usługi rozumie "wszelkie czynności związane bezpośrednio lub pośrednio /np. przy podziale produktów/ z zaspokojeniem potrzeb ludzkich, ale nie służące bezpośrednio wytwarzania przedmiotów" /por. [89], s. 24/. Na bazie powyższej definicji utworzono inne, różniące się sformułowaniem, wzbogaceniem o pewne mniej istotne elementy. Przegląd tych definicji można znaleźć w [35],[140]. Wynika z nich, iż podstawową cechą usługi jest to, iż jej wykonywanie nie powoduje powstania nowego dobra a co najwyżej przekształcenie dobra już istniejącego lub /częściej/ bezpośrednie zaspokojenie ludzkich potrzeb.

Oprócz rozumienia usług jako czynności można je traktować jako sektor gospodarki narodowej. W tym rozumieniu usługi obejmują wszystkie działy gospodarki narodowej oprócz rolnictwa i leśnictwa /sektor I/ oraz przemysłu i budownictwa /sektor II/. Do sektora usług należą zatem m.in. transport i łączność, obrót towarowy, gospodarka komunalna i mieszkaniowa, ochrona zdrowia, oświata, nauka, itp.

Oba podejścia do pojęcia usług nie pokrywają się, gdyż w działach gospodarki narodowej zaliczanych do sektorów I i II również wykonywane są czynności usługowe /te, które są związane z obsługą procesu produkcyjnego, a więc np. remonty/.

W pracy usługi są rozważane jako czynności.

Obecnie pojęcie usług zostanie sprecyzowane bliżej poprzez podanie czynności wchodzących w skład usług. Dokonamy tego przez klasyfikację usług /o klasyfikacji usług traktują m.in. prace [35],[46],[68],[80],[106],[131],[140]/.

W pracy zajmować się będziemy jedynie <u>usługami dla ludności</u>.

Są to usługi świadczone ludności /a zatem i gospodarstwom domowym/ oraz jednostkom gospodarki nieuspołecznionej.

Przy klasyfikowaniu usług /a następnie przy ich badaniach/ nie jest uwzględniany element odpłatności. Przez usługi dla ludności rozumieć będziemy zarówno usługi rynkowe, tzn. te, które są świadczone odpłatnie /zaliczane do tzw. spożycia odpłatnego/, jak i pozarynkowe, czyli świadczone nieodpłatnie /zaliczane do tzw. spożycia zbiorowego i spożycia społecznego/. Niektóre z usług dzielą się na część usługi świadczoną nieodpłatnie oraz na część świadczoną odpłatnie /sprzedawane są poniżej kosztów/.

Powyżej przedstawione podejście różni się od przyjętego powszechnie /por. np.[68],[80],[131]/, gdzie do usług dla ludności zalicza się jedynie usługi rynkowe. Wychodzimy z założenia, że podział usług na rynkowe i pozarynkowe jest podziałem sztucznym, zwłaszcza obecnie, gdy system cen na usługi pozbawiony jest racjonalnych podstaw.

Dla potrzeb badań stosować będziemy klasyfikację częściowo opartą na klasyfikacji stosowanej przez GUS /por.[68],[80],[131]/. Ogół usług dla ludności można podzielić na usługi konsumpcyjne i usługi ogólnospołeczne /por.[46]/.

Do usług ogólnospołecznych zalicza się: usługi w zakresie bezpieczeństwa i obrony narodowej, usługi w zakresie finansów i ubezpieczeń, usługi w zakresie administracji publicznej, usługi instytucji wymiaru sprawiedliwości, usługi adwokackie, usługi organizacji polityczno-społecznych, usługi religijne.

W pracy nie będziemy się zajmować usługami ogólnospołecznymi, po pierwsze dlatego, że niezwykle trudna jest ich kwantyfikacja, po drugie dlatego, gdyż ich poziom w poszczególnych województwach jest w przybliżeniu jednakowy, gdyż w przypadku większości z nich wynika z podziału administracyjnego kraju.

Zatem dla potrzeb pracy pojęcie usług zostało zawężone do usług konsumpcyjnych dla ludności. W ich skład wchodzą następujące rodzaje usług:

- usługi bytowe,
- usługi w zakresie transportu i łączności,
- usługi komunalne i mieszkaniowe,
- usługi w zakresie ochrony zdrowia, opieki społecznej, kultury fizycznej, turystyki i wypoczynku,
- usługi kulturalno-rozrywkowe,
- usługi w zakresie oświaty i wychowania.

Usługi bytowe /por.[80]/ są to usługi naprawcze, budowlane i instalacyjne, usługi związane z prowadzeniem gospodarstwa domowego oraz usługi polegające na zaspokajaniu niektórych potrzeb osobistych. Zalicza się tu zatem usługi związane ze sprzętem radiowo-telewizyjnym oraz zmechanizowanym sprzętem gospodarstwa domowego, usługi prania, maglowania, chemicznego czyszczenia i farbowania, usługi krawieckie i kuśnierskie, usługi fryzjerskie i kosmetyczne,

usługi skórzane, usługi stolarsko-tapicerskie itp.
Usługi w zakresie transportu i łączności obejmują przewóz osób i ładunków różnymi rodzajami transportu oraz usługi poczty, telefonii, telegrafii i radiofonii przewodowej, itp.
Usługi komunalne i mieszkaniowe obejmują dostawę energii elektrycznej, energii cieplnej i gazu, usługi wodno-kanalizacyjne, oczyszczanie i wywóz nieczystości, komunikację miejską itp.
Usługi kulturalno-rozrywkowe obejmują usługi kin, teatrów, instytucji muzycznych, radia, telewizji i przedsiębiorstw rozrywkowych, itp.
Usługi w zakresie ochrony zdrowia itd. obejmują usługi zakładów leczniczych, opieki społecznej, turystyczne, wczasów, itp.
Usługi w zakresie oświaty i wychowania obejmują usługi ośrodków nauczania i szkolenia zawodowego oraz usługi ośrodków opieki nad dziećmi i młodzieżą.
Bardziej szczegółowe informacje odnośnie poszczególnych grup usług znaleźć można w [80],[155].

5.2. Obiekty i zmienne badania. Uwagi na temat doboru zmiennych.

Jak poprzednio zasygnalizowano, celem badań empirycznych była analiza przestrzennego zróżnicowania poziomu usług konsumpcyjnych dla ludności w Polsce. Jako zbiór obiektów przyjęto województwa /zatem n = 49/. Okresem badania był rok 1979 /ostatni rok, z którego były dostępne obserwacje/.

Badaniami objęto 5 grup konsumpcyjnych dla ludności:
1. usługi kulturalno-rozrywkowe,
2. usługi w zakresie ochrony zdrowia i opieki społecznej,
3. usługi komunalno-mieszkaniowe,
4. usługi w zakresie oświaty i wychowania,

5. usługi bytowe.

Jak widać, spośród wyodrębnionych w 5.1 grup usług konsumpcyjnych dla ludności pominięte zostały w badaniach usługi w zakresie transportu i łączności. Główną przyczyną był brak danych charakteryzujących te usługi w przekroju wojewódzkim. Wzięto pod uwagę tylko jedną zmienną charakteryzującą usługi w zakresie łączności: ilość prywatnych abonentów telefonicznych przypadających na 1000 mieszkańców. Zmienna ta została włączona do zmiennych z grupy usług komunalno-mieszkaniowych. Jak się wydaje, usługa związana z funkcjonowaniem telefonu ma podobny charakter, jak niektóre usługi mieszkaniowe; telefon został potraktowany jako wyposażenie mieszkania.

Ponadto, pominięto również usługi w zakresie turystyki i wypoczynku, gdyż poziom tych usług uzależniony jest od atrakcyjności turystycznej województwa. Dane obrazujące poziom tych usług, dotyczą usług zaspokajanych w województwie, a nie usług zaspokajanych przez ludność tego województwa.

Na początku sporządzona została wstępna lista zmiennych, charakteryzujących poszczególne grupy usług. Wybór tych zmiennych, oprócz ich zawartości merytorycznej, warunkowany był dostępnością danych.

Wstępna lista zmiennych jest następująca /przez $X_{i,j}$ oznaczono zmienną o numerze j z grupy o numerze i, przy czym numeracja grup jest taka, jak wyżej podano/:

1. Usługi kulturalno-rozrywkowe.

$X_{1,1}$ - ilość woluminów w bibliotekach przypadająca na 1000 mieszkańców,

$X_{1,2}$ - ilość abonentów radiowych przypadająca na 1000 mieszkańców,

$X_{1,3}$ - ilość abonentów telewizyjnych przypadająca na 1000 mieszkańców,

$X_{1,4}$ - ilość miejsc na widowni w kinach stałych przypadająca na 1000 mieszkańców,

$X_{1,5}$ - ilość widzów w teatrach przypadająca na 1000 mieszkańców,

$X_{1,6}$ - ilość widzów w instytucjach muzycznych przypadająca na 1000 mieszkańców,

$X_{1,7}$ - ilość wypożyczeń przypadająca na 1 czytelnika bibliotek,

$X_{1,8}$ - ilość sprzedanych gazet i czasopism przypadająca na 1 mieszkańca,

$X_{1,9}$ - wartość sprzedaży usług kultury i sztuki w złotówkach przypadająca na 1 mieszkańca.

2. Usługi w zakresie ochrony zdrowia i opieki społecznej.

$X_{2,1}$ - ilość lekarzy przypadająca na 10000 mieszkańców,

$X_{2,2}$ - ilość lekarzy dentystów przypadająca na 10000 mieszkańców,

$X_{2,3}$ - ilość pielęgniarek przypadająca na 10000 mieszkańców,

$X_{2,4}$ - ilość łóżek w szpitalach ogólnych przypadająca na 10000 mieszkańców,

$X_{2,5}$ - ilość porad udzielonych przez lekarzy i lekarzy dentystów przypadająca na 1 mieszkańca,

$X_{2,6}$ - ilość zgonów niemowląt przypadająca na 1000 urodzeń żywych,

$X_{2,7}$ - ilość farmaceutów przypadająca na 10000 mieszkańców,

$X_{2,8}$ - udział pielęgniarek o pełnych kwalifikacjach w ogólnej liczbie pielęgniarek w %,

$X_{2,9}$ - ilość położnych przypadająca na 10000 mieszkańców,

$X_{2,10}$ - ilość karetek sanitarnych przypadająca na 100 000 mieszkańców,

$X_{2,11}$ - ilość aptek przypadająca na miasto i gminę,

$X_{2,12}$ - ilość mieszkańców wsi w tys. przypadająca na ośrodek zdrowia,

$X_{2,13}$ - ilość dzieci w wieku 0-2 lat przypadająca na 1 miejsce w żłobku.

.3. Usługi komunalno-mieszkaniowe.

$X_{3,1}$ - procent ludności miast korzystający z sieci wodociągowej,

$X_{3,2}$ - procent ludności miast korzystający z sieci kanalizacyjnej,

$X_{3,3}$ - procent długości jezdni w miastach o nawierzchni twardej ulepszonej,

$X_{3,4}$ - powierzchnia parków i zieleńców w m^2 przypadająca na 1 mieszkańca miast,

$X_{3,5}$ - przeciętna powierzchnia użytkowa mieszkań w m^2 przypadająca na 1 mieszkańca,

$X_{3,6}$ - ilość prywatnych abonentów telefonicznych przypadająca na 1000 mieszkańców,

$X_{3,7}$ - zużycie wody z wodociągów w gospodarstwach domowych w m^3 przypadająca na 1 mieszkańca,

$X_{3,8}$ - zużycie energii elektrycznej w gospodarstwach domowych w kWh przypadająca na 1 mieszkańca,

$X_{3,9}$ - wartość sprzedaży usług komunikacji miejskiej w złotówkach przypadająca na 1 mieszkańca miast,

$X_{3,10}$ - wartość sprzedaży usług komunalnych w złotówkach przypadająca na 1 mieszkańca.

4. Usługi w zakresie oświaty i wychowania.

$X_{4,1}$ - ilość dzieci w przedszkolach przypadająca na 1000 dzieci w wieku 3-6 lat,

$X_{4,2}$ - ilość dzieci w przedszkolach przypadająca na 1 nauczyciela,

$X_{4,3}$ - ilość dzieci w przedszkolach przypadająca na 100 miejsc,

$X_{4,4}$ - ilość korzystających z dziecińców wiejskich przypadająca na 1 nauczyciela,

$X_{4,5}$ - procent pracowni przedmiotowych w ogólnej ilości izb lekcyjnych w szkołach podstawowych,

$X_{4,6}$ - procent dzieci w wieku 7-15 lat uczących się w szkołach podstawowych,

$X_{4,7}$ - procent młodzieży w wieku 16-20 lat uczącej się w szkołach średnich,

$X_{4,8}$ - ilość uczniów w szkołach podstawowych przypadająca na 1 nauczyciela pełnozatrudnionego,

$X_{4,9}$ - ilość uczniów w szkołach podstawowych przypadająca na 1 pomieszczenie,

$X_{4,10}$ - ilość uczniów w liceach ogólnokształcących przypadająca na 1 oddział.

5. Usługi bytowe.

$X_{5,1}$ - wartość sprzedaży usług motoryzacyjnych w złotówkach przypadająca na 1 prywatny samochód osobowy,

$X_{5,2}$ - wartość sprzedaży usług naprawczych w złotówkach przypadająca na 1 mieszkańca,

$X_{5,3}$ - zatrudnienie w usługach naprawczych przypadające na 10000 mieszkańców,

$X_{5,4}$ - wartość sprzedaży usług dziewiarskich, odzieżowych, bieliźniarskich i szewskich w złotówkach przypadająca na 1 mieszkańca,

$X_{5,5}$ - zatrudnienie w usługach dziewiarskich, odzieżowych, bieliźniarskich i szewskich przypadające na 10000 mieszkańców,

$X_{5,6}$ - wartość sprzedaży usług pralniczych w złotówkach przypadająca na 1 mieszkańca,

$X_{5,7}$ - zatrudnienie w usługach pralniczych przypadające na 10000 mieszkańców,

$X_{5,8}$ - wartość sprzedaży usług kowalskich, ślusarskich, szklarskich, stolarskich, meblarskich i tapicerskich w złotówkach przypadająca na 1 mieszkańca,

$X_{5,9}$ - zatrudnienie w usługach kowalskich, ślusarskich, szklarskich, stolarskich, meblarskich i tapicerskich przypadające na 10000 mieszkańców,

$X_{5,10}$ - wartość sprzedaży usług osobistych w złotówkach przypadająca na 1 mieszkańca,

$X_{5,11}$ - zatrudnienie w usługach osobistych przypadające na 10000 mieszkańców.

Przy badaniu poziomu usług, najlepszymi zmiennymi charakteryzującymi go, są zmienne odzwierciedlające efekty działalności usługowej. Toteż tam, gdzie było to możliwe, takie zmienne zostały uwzględnione /w ujęciu ilościowym lub wartościowym/. Zmienne te zostały uzupełnione o zmienne obrazujące nakłady w sferze usług /należy zwrócić uwagę, iż w przypadku niektórych zmiennych nie można jednoznacznie określić, czy odzwierciedlają efekty czy nakłady - dotyczy to np. zmiennych $X_{3,1}$ i $X_{3,2}$/. We wstępnej liście zmiennych znalazły się zmienne niosące informacje o poziomie usług należących do danej grupy. Oczywiście niektóre ze zmiennych mogą budzić wątpliwości, dotyczy to przykładowo części zmiennych z czwartej grupy - $X_{4,6}$, $X_{4,7}$, $X_{4,10}$, ale w przypadku usług w zakresie oświaty i wychowania /a zwłaszcza oświaty/ były to jedyne dostępne dane.

Jak nietrudno zauważyć, wiele zmiennych znajdujących się na wstępnej liście zmiennych niesie bardzo zbliżone informacje, a zatem informacje te powiela. Badacze, stosujący metody klasyfikacji, wiele uwagi poświęcają problematyce powielania informacji i związanemu z nią doborowi zmiennych. Formułuje się różne postulaty odnośnie zmiennych wykorzystywanych w badaniach za pomocą tych metod /nazywane są te zmienne diagnostycznymi/. I tak np. W. Pluta stwierdza /por.[114], s. 45/, że należy dążyć "do uzyskania zmiennych, które w sposób możliwie pełny charakteryzowałyby badane jednostki, a przy tym tworzyły zespół jak najmniej liczny. Podane wymagania są spełnione wtedy, gdy zmienne diagnostyczne posiadają następujące własności:

- są nieskorelowane lub co najwyżej słabo skorelowane między sobą,
- są silnie skorelowane ze zmiennymi nie wchodzącymi do zespołu diagnostycznego,
- posiadają zdolność dyskryminacji badanych jednostek, tj. charakteryzują się wysoką zmiennością wśród wszystkich jednostek zbioru, a niską wśród jednostek wydzielonych grup,
- nie ulegają wpływom zewnętrznym".

Spośród najczęściej spotykanych postulatów odnośnie zmiennych /oprócz wymogu merytorycznego, traktującego o tym, że zmienne powinny możliwie dokładnie reprezentować zjawiska wchodzące w zakres badania/ wymienić należy dużą zmienność i małe wzajemne skorelowanie. W literaturze przedmiotu, zwłaszcza krajowej, dość mocno podkreśla się ten drugi wymóg, motywując to przede wszystkim tym, że silne skorelowanie zmiennych oznacza powielanie informacji niesionych przez zmienne. Wydaje się jednak, skorelowanie zmiennych nie zawsze oznacza powielanie informacji przez te zmienne. Zmienne mogą być silnie skorelowane, a z punktu widzenia badanego problemu nieść inne informacje.

Większość formalnych metod doboru zmiennych jako podstawę przyjmuje słabe skorelowanie zmiennych diagnostycznych. Ich stosowanie może spowodować pominięcie zmiennej istotnej z punktu widzenia celu badania tylko dlatego, że jest silnie skorelowana z innymi. W celu uniknięcia tego, w pracy stosowana będzie następująca heurystyka wyboru zmiennych:

1. na podstawie analizy merytorycznej problemu zostaje sporządzona wstępna lista zmiennych,
2. ze wstępnej listy zmiennych wyeliminowane zostają zmienne o relatywnie niskiej zmienności /może ona być mierzona za pomocą współczynnika zmienności/,
3. na podstawie analizy merytorycznej pozostałych zmiennych wyodrębnione zostają grupy zmiennych, co do których istnieje przypuszczenie, że mogą powielać informacje,
4. analizowane są współczynniki korelacji zmiennych występujących w tych grupach w porównaniu ze wszystkimi współczynnikami korelacji /ściślej mówiąc, analizowane są wartości bezwzględne współczynników korelacji/. W przypadku, gdy wartości te potwierdzają przypuszczenie o powielaniu informacji, z grupy zmiennych zostaje wybrana ich część /z reguły jedna zmienna/. O wyborze w tym wypadku powinny decydować również względy merytoryczne, a w przypadku ich braku, np. mniejsze skorelowanie z innymi zmiennymi diagnostycznymi, większa zmienność lub większe skorelowanie z innymi zmiennymi tej grupy.

Opisana powyżej heurystyka pozwala na otrzymanie zmiennych diagnostycznych obejmujących pewne zakresy informacji o badanym problemie.

Ten sposób postępowania został również zastosowany do zmiennych wchodzących w skład wstępnej listy zmiennych. Na wstępie dla każdej zmiennej wyznaczono współczynnik zmienności. Rezultaty prezentowane są w tablicy 5.1. Wyznaczono również średnie arytmetyczne współczynników zmienności dla wszystkich grup. Wyniosły one odpowiednio:

0,3289, 0,3192, 0,3037, 0,0983, 0,3876

Jak widać zmienne wszystkich grup oprócz czwartej /tzn. zmiennych charakteryzujących usługi oświaty i wychowania/ charakteryzują się zbliżoną zmiennością, średnio w granicach 0,3 - 0,4. Współczynnik zmienności dla tych zmiennych waha się w granicach: dla grupy pierwszej 0,1040-0,7841, dla grupy drugiej 0,0337 - 0,7301, dla grupy trzeciej 0,0563 - 0,5892, dla grupy piątej 0,2122 - 0,6250. Biorąc pod uwagę średni współczynnik zmienności i przyjmując dostateczną zmienność 0,1 zauważamy, że pominąć należy zmienne $X_{2,8}$ oraz $X_{3,5}$, które charakteryzują się wyjątkowo niskimi wartościami współczynnika zmienności.

Nieco inna sytuacja jest w przypadku zmiennych grupy czwartej. Większość zmiennych charakteryzuje się zaledwie kilkuprocentową zmiennością. Jedynie dwie zmienne mają współczynnik zmienności przekraczający 0,1. Jednak pozostawienie tylko tych dwu zmiennych zawęziłoby w dużym stopniu obszar, z którego pochodzą informacje. Oprócz tego, stwierdzić należy, że zmienne charakteryzujące usługi oświaty i wychowania /a zwłaszcza usługi oświaty/ powinny z merytorycznego punktu widzenia mieć małą zmienność, gdyż ich poziom jest w dużym stopniu równomierny w układzie

Tablica 5.

Wartość współczynników zmienności zmiennych ze wstępnej listy

Zmienna	Wartość współczynnika zmienności
$X_{1,1}$	0,1382
$X_{1,2}$	0,1405
$X_{1,3}$	0,1611
$X_{1,4}$	0,3118
$X_{1,5}$	0,7841
$X_{1,6}$	0,7350
$X_{1,7}$	0,1040
$X_{1,8}$	0,2646
$X_{1,9}$	0,3208
$X_{2,1}$	0,4738
$X_{2,2}$	0,3845
$X_{2,3}$	0,2152
$X_{2,4}$	0,2412
$X_{2,5}$	0,2446
$X_{2,6}$	0,1029
$X_{2,7}$	0,5533
$X_{2,8}$	0,0337
$X_{2,9}$	0,2310
$X_{2,10}$	0,1882
$X_{2,11}$	0,7301
$X_{2,12}$	0,1272
$X_{2,13}$	0,6245
$X_{3,1}$	0,1445
$X_{3,2}$	0,1646
$X_{3,3}$	0,1581
$X_{3,4}$	0,3213
$X_{3,5}$	0,0563
$X_{3,6}$	0,5818
$X_{3,7}$	0,5892
$X_{3,8}$	0,3446
$X_{3,9}$	0,3084
$X_{3,10}$	0,3770
$X_{4,1}$	0,0992
$X_{4,2}$	0,0671
$X_{4,3}$	0,0656
$X_{4,4}$	0,1811
$X_{4,5}$	0,2571
$X_{4,6}$	0,0072
$X_{4,7}$	0,0901
$X_{4,8}$	0,0608
$X_{4,9}$	0,0714
$X_{4,10}$	0,0833
$X_{5,1}$	0,2122
$X_{5,2}$	0,2388
$X_{5,3}$	0,2192
$X_{5,4}$	0,4315
$X_{5,5}$	0,3479
$X_{5,6}$	0,6250
$X_{5,7}$	0,5848
$X_{5,8}$	0,3545
$X_{5,9}$	0,3928
$X_{5,10}$	0,4427
$X_{5,11}$	0,4146

Źródło: obliczenia własne.

przestrzennym. Dlatego zdecydowano się pominąć jedynie zmienną $X_{4,6}$, mającą zdecydowanie najniższy współczynnik zmienności, natomiast pozostałe zmienne, których współczynnik zmienności oscyluje wokół średniego, pozostawiono do dalszych badań.
W sumie zatem, analiza zmienności wykluczyła trzy zmienne: $X_{2,8}$, $X_{3,5}$, $X_{4,6}$.
Następnie przeprowadzona została dalsza część opisanego wyżej sposobu postępowania.
W grupie pierwszej zmienna $X_{1,9}$ niesie część informacji przekazywanych przez większość pozostałych zmiennych. Współczynniki korelacji tej zmiennej z pozostałymi zmiennymi wynoszą:

$$\mathrm{cor}\ /X_{1,1},\ X_{1,9}/ = -0{,}219$$
$$\mathrm{cor}\ /X_{1,2},\ X_{1,9}/ = 0{,}833$$
$$\mathrm{cor}\ /X_{1,3},\ X_{1,9}/ = 0{,}794$$
$$\mathrm{cor}\ /X_{1,4},\ X_{1,9}/ = 0{,}476$$
$$\mathrm{cor}\ /X_{1,5},\ X_{1,9}/ = 0{,}783$$
$$\mathrm{cor}\ /X_{1,6},\ X_{1,9}/ = 0{,}757$$
$$\mathrm{cor}\ /X_{1,7},\ X_{1,9}/ = 0{,}539$$
$$\mathrm{cor}\ /X_{1,8},\ X_{1,9}/ = 0{,}928$$

Średnia bezwzględnych wartości współczynnika korelacji dla wszystkich zmiennych tej grupy wynosi 0,506. Wielkość ta obrazuje przeciętną siłę zależności między zmiennymi należącymi do tej grupy. Ponieważ aż dla sześciu spośród ośmiu zmiennych siła zależności z $X_{1,9}$ jest większa, zatem zmienna $X_{1,9}$ zostaje pominięta.

W grupie drugiej zmienna $X_{2,5}$ niesie zbliżone informacje co zmienna $X_{2,1}$ oraz $X_{2,2}$. Współczynniki korelacji wynoszą:

$$\mathrm{cor}\ /X_{2,1},\ X_{2,5}/ = 0{,}879$$
$$\mathrm{cor}\ /X_{2,2},\ X_{2,5}/ = 0{,}849$$

Średni bezwzględny współczynnik korelacji w tej grupie wynosi 0,410. Zatem zmienna $X_{2,5}$ zostaje pominięta.

W grupie trzeciej istnieje przypuszczenie co do powielania informacji przez zmienne $X_{3,1}$ i $X_{3,7}$. Współczynnik korelacji cor $/X_{3,1}, X_{3,7}/$ = 0,730 /wobec średniego równego 0,406/, zatem pomijamy jedną z tych dwóch zmiennych. Jest to zmienna $X_{3,1}$, obrazująca nakłady, podczas gdy zmienna $X_{3,7}$ obrazuje raczej efekty działalności usługowej, zatem wydaje się lepiej odzwierciedlać poziom usług.

Natomiast zmienna $X_{3,10}$ powiela część informacji niesionych przez pozostałe zmienne. Współczynniki korelacji są następujące:

cor $/X_{3,2}, X_{3,10}/$ = 0,680
cor $/X_{3,3}, X_{3,10}/$ = 0,039
cor $/X_{3,4}, X_{3,10}/$ = 0,127
cor $/X_{3,6}, X_{3,10}/$ = 0,877
cor $/X_{3,7}, X_{3,10}/$ = 0,879
cor $/X_{3,8}, X_{3,10}/$ = 0,873
cor $/X_{3,9}, X_{3,10}/$ = 0,445

Średni bezwzględny współczynnik korelacji wynosi 0,406. I w tym przypadku duże skorelowanie powoduje wykluczenie zmiennej $X_{3,10}$.

Wreszcie, w grupie piątej dla pięciu rodzajów usług występują po 2 zmienne: jedna dotycząca wartości usług sprzedanych, a druga zatrudnienia. Ponieważ zatrudnienie obrazuje nakłady, a wartość sprzedaży efekty, a współczynniki korelacji między odpowiednimi zmiennymi są wysokie, rzędu 0,7 - 0,9, zatem pominięte zostają zmienne: $X_{5,3}, X_{5,5}, X_{5,7}, X_{5,9}, X_{5,11}$.

W efekcie w wyniku zastosowania opisanego wyżej sposobu postępowania w skład zbioru zmiennych diagnostycznych weszły następujące zmienne:

dla usług kulturalno-rozrywkowych: $X_{1,1}$, $X_{1,2}$, $X_{1,3}$, $X_{1,4}$, $X_{1,5}$, $X_{1,6}$, $X_{1,7}$, $X_{1,8}$;

dla usług w zakresie ochrony zdrowia i opieki społecznej: $X_{2,1}$, $X_{2,2}$, $X_{2,3}$, $X_{2,4}$, $X_{2,6}$, $X_{2,7}$, $X_{2,9}$, $X_{2,10}$, $X_{2,11}$, $X_{2,12}$, $X_{2,13}$;

dla usług komunalno-mieszkaniowych: $X_{3,2}$, $X_{3,3}$, $X_{3,4}$, $X_{3,6}$, $X_{3,7}$, $X_{3,8}$, $X_{3,9}$;

dla usług w zakresie oświaty i wychowania: $X_{4,1}$, $X_{4,2}$, $X_{4,3}$, $X_{4,4}$, $X_{4,5}$, $X_{4,7}$, $X_{4,8}$, $X_{4,9}$, $X_{4,10}$;

dla usług bytowych: $X_{5,1}$, $X_{5,2}$, $X_{5,4}$, $X_{5,6}$, $X_{5,8}$, $X_{5,10}$.

5.3. Klasyfikacja województw Polski ze względu na poziom poszczególnych rodzajów usług.

5.3.1. Uwagi ogólne.

W oparciu o wyodrębniony w poprzednim paragrafie zbiór zmiennych diagnostycznych dokonano klasyfikacji województw. Przeprowadzono ją osobno dla każdej grupy usług. Jako metodę klasyfikacji przyjęto metodę średniej międzygrupowej należącą do grupy hierarchicznych metod grupowania. Zalety tej metody w porównaniu z innymi zostały zasygnalizowane w rozdziale 3. Po każdym etapie grupowania zostały wyznaczone wartości średniej odległości wewnątrzgrupowej W'p oraz maksymalnej odległości wewnątrzgrupowej Wp. Wielkości te charakteryzujące wewnątrzklasowe podobieństwo dla otrzymanej po danym etapie grupowania klasyfikacji, posłużą do wyboru ostatecznej klasyfikacji spośród ich ciągu /por. 3.2/. Przyjęto, że ilość klas powinna /ze względu na zawartość informacyjną klasyfikacji/ zawierać się

w przedziale ⟨5; 15⟩ . Nie czyni się natomiast żadnych założeń co do ilości obiektów w klasie.
Dla wybranej za pomocą kryterium najwyższego spadku podobieństwa /a zatem maksymalnej straty informacji/ klasyfikacji wyznaczone zostały średnie arytmetyczne każdej ze zmiennych osobno dla każdej klasy. Będą one pomocne przy analizie klasyfikacji.
Ponadto wyznaczone zostały również uporządkowania liniowe województw za pomocą metody sum standaryzowanych /por.[28]/. Jest to stosunkowo najprostsza metoda porządkowania liniowego zbioru obiektów. Otrzymane uporządkowania posłużą do porównań z klasyfikacjami, a także do analizy tych klasyfikacji. W kolejnych punktach przedstawione zostaną rezultaty klasyfikacji dla poszczególnych grup usług. Zaznaczyć należy, iż dla usług oświaty i wychowania rezultaty te /podobnie jak rezultaty badań przedstawione w dalszej części/ należy traktować z pewną rezerwą. Wynika to z małej zmienności zmiennych, a także /o czym wspomniano w 5.2/ z faktu, iż przyjęcie pewnych zmiennych, zwłaszcza dotyczących usług w zakresie oświaty, wydaje się być dyskusyjne.

5.3.2. Klasyfikacja dla usług kulturalno-rozrywkowych.

Rezultaty grupowania dla tej grupy usług przedstawione są w tablicy 5.2 /w tablicy tej, podobnie jak następnych, prezentujących rezultaty grupowania, przedstawione są tylko rezultaty od 34-go do 45-go etapu, ze względu na wymóg ilości klas/. Wskazują one, że najwyższe spadki podobieństwa /obrazowane przez wzrost Wp oraz W'p/, a zatem straty informacji obserwuje się po 35-tym oraz 43-cim etapie grupowania. Dlatego wybrano dwie klasyfikacje o różnym stopniu szczegółowości.

Tablica 5.2

Rezultaty grupowania dla usług kulturalno-rozrywkowych

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,1298	0,6334
35	1,1873	0,6509
36	1,5300	0,7511
37	1,5300	0,7735
38	1,5300	0,7863
39	1,5300	0,7883
40	1,5300	0,7889
41	1,5300	0,7962
42	1,5300	0,8042
43	1,6882	0,8439
44	1,9789	0,9229
45	1,9789	0,9406

Źródło: obliczenia własne.

Klasyfikacja pierwsza składa się z 14 klas.

W skład poszczególnych obszarów typu regionalnego wchodzą następujące województwa /będziemy je oznaczać przez ich stolice/:

I - Warszawa.

II - Biała Podlaska, Ciechanów, Konin, Krosno, Łomża, Nowy Sącz, Ostrołęka, Przemyśl, Radom, Siedlce, Tarnobrzeg, Tarnów, Zamość.

III - Białystok, Rzeszów.

IV - Bielsko-Biała, Chełm, Częstochowa, Kalisz, Kielce, Lublin, Piotrków Tryb., Płock, Sieradz, Skierniewice.

V - Bydgoszcz, Gdańsk, Poznań, Wrocław.

VI - Elbląg, Gorzów Wlkp., Olsztyn, Opole, Piła, Słupsk, Toruń, Zielona Góra.

VII - Jelenia Góra, Koszalin, Wałbrzych.

VIII - Katowice, Legnica.

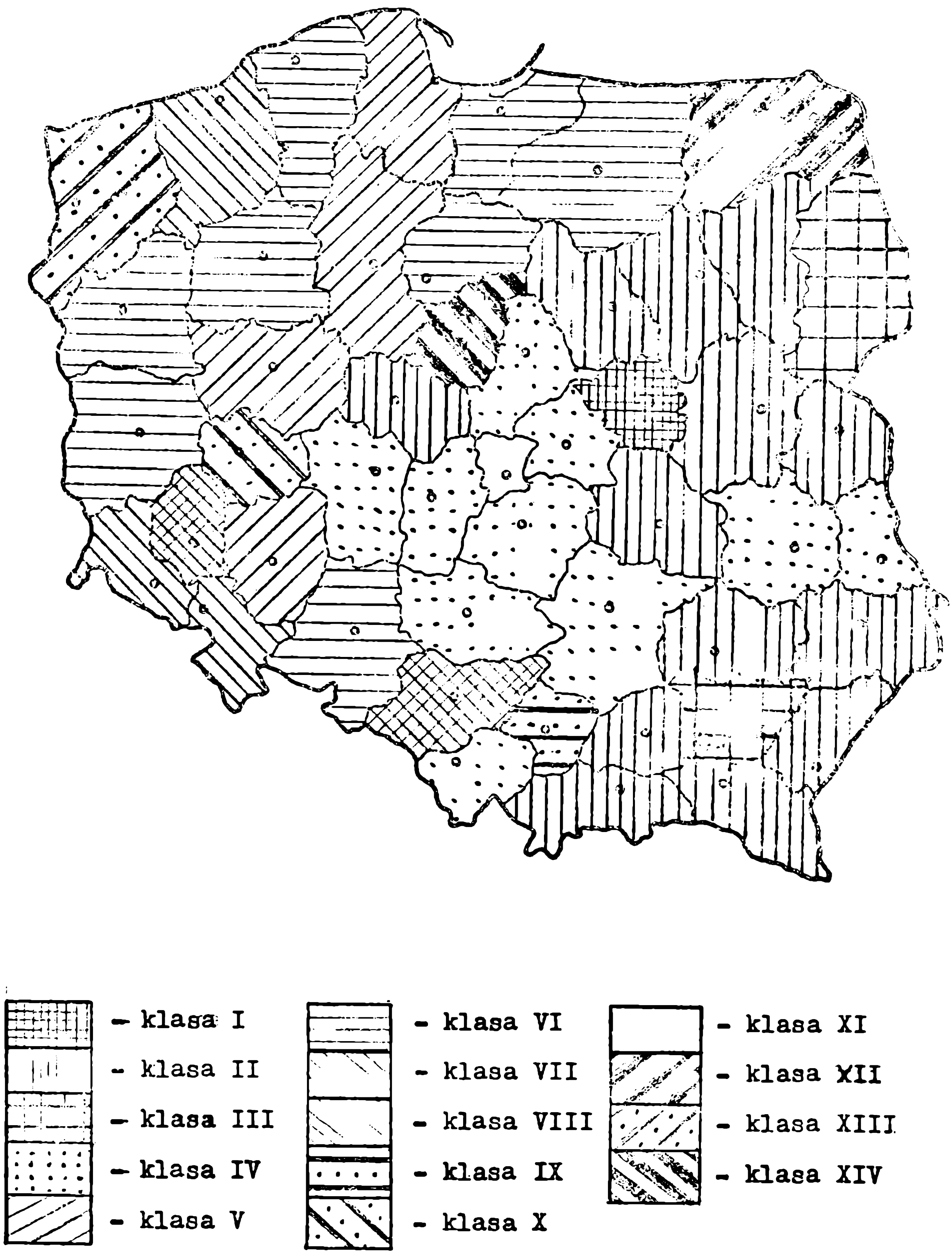
IX - Kraków.
X - Leszno.
XI - Łódź.
XII - Suwałki.
XIII - Szczecin.
XIV - Włocławek.

Rysunek 5.1 przedstawia mapę układu obszarów typu regionalnego powyższej klasyfikacji, natomiast tablica 5.3 średnie arytmetyczne dla wszystkich zmiennych wyznaczone osobno dla każdego obszaru.

Analiza tej klasyfikacji wskazuje, że usługi kulturalno-rozrywkowe charakteryzują się zróżnicowaniem terytorialnym. Poziom tych usług jest zależny od położenia województwa. Wskazuje na to układ obszarów typu regionalnego.

Spośród wyodrębnionych klas jednoelementowych zwracają uwagę trzy województwa: warszawskie, łódzkie i szczecińskie. Wyróżniają się one wysokimi wartościami zmiennych. Wyjątkiem jest jedynie zmienna $X_{1,1}$ w przypadku woj. warszawskiego oraz $X_{1,1}$ i $X_{1,7}$ woj. łódzkiego. Wskazywałoby to na relatywny niedorozwój w dziedzinie bibliotek w tych dwóch województwach. Oprócz tych trzech klas wysokie wartości zmiennych posiadają klasa V, zawierająca woj. bydgoskie, gdańskie, poznańskie i wrocławskie, oraz klasa IX zawierająca woj. krakowskie. Widać zatem /czego należało oczekiwać, że najwyższy poziom usług kulturalno-rozrywkowych obserwuje się w województwach, których stolicami jest osiem największych /oprócz Katowic/ miast Polski. Najniższe wartości zmiennych obserwuje się w klasie II, zawierającej województwa wschodnie i południowo-wschodnie oraz w klasie IV, zawierającej województwa Polski środkowej oraz województwo bielskobialskie.

Rys. 5.1. Klasyfikacja województw Polski ze względu na poziom usług kulturalno-rozrywkowych - wariant pierwszy.



Tablica 5.3

Średnie arytmetyczne zmiennych charakteryzujących usługi kulturalno-rozrywkowe dla poszczególnych klas - - klasyfikacja pierwsza

Nr klasy	Zmienna							
	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$
1	2043	302	271	15,3	931	462	23,8	163
2	2757	193	166	11,0	81	98	17,8	65
3	2967	198	166	12,5	542	256	19,8	83
4	2436	221	200	11,4	138	129	19,4	77
5	2436	265	239	13,1	247	410	21,5	123
6	3048	244	225	17,2	234	228	18,3	96
7	2744	251	231	24,5	283	254	19,4	96
8	2062	252	238	17,1	189	111	21,5	103
9	1964	246	217	9,3	508	342	22,8	115
10	3087	239	224	9,5	70	188	21,2	94
11	2361	322	288	13,3	511	716	16,9	127
12	3287	207	189	18,4	119	116	22,4	75
13	3411	270	245	24,5	330	424	23,0	123
14	3045	236	209	15,0	536	159	20,3	76

Źródło: obliczenia własne.

Pewnym wyjątkiem jest woj. suwalskie, tworzące odrębną klasę ze względu na stosunkowo wysokie wartości zmiennych $X_{1,1}$, $X_{1,4}$ i $X_{1,7}$ /kina i biblioteki/.

Zauważyć należy również, iż mniejsze pod względem liczby ludności województwa zaliczane do tzw. Ziem Zachodnich i Północnych utworzyły dwie klasy: VI i VII, mają one wyższe wartości zmiennych niż województwa środkowe i wschodnie.

Klasyfikacja druga składa się z 6 klas, Są to:

I - Warszawa,

II - Biała Podlaska, Białystok, Bielsko-Biała, Chełm, Ciechanów, Częstochowa, Kalisz, Kielce, Konin, Krosno, Lublin, Łomża, Nowy Sącz, Ostrołęka, Piotrków Tryb., Płock, Przemyśl, Radom, Rzeszów, Siedlce, Sieradz, Skierniewice, Suwałki, Tarnobrzeg, Tarnów, Zamość,

III - Bydgoszcz, Gdańsk, Katowice, Kraków, Legnica, Poznań, Wrocław,

IV - Elbląg, Gorzów Wlkp., Jelenia Góra, Koszalin, Olsztyn, Opole, Piła, Słupsk, Toruń, Wałbrzych, Włocławek, Zielona Góra,

V - Łódź,

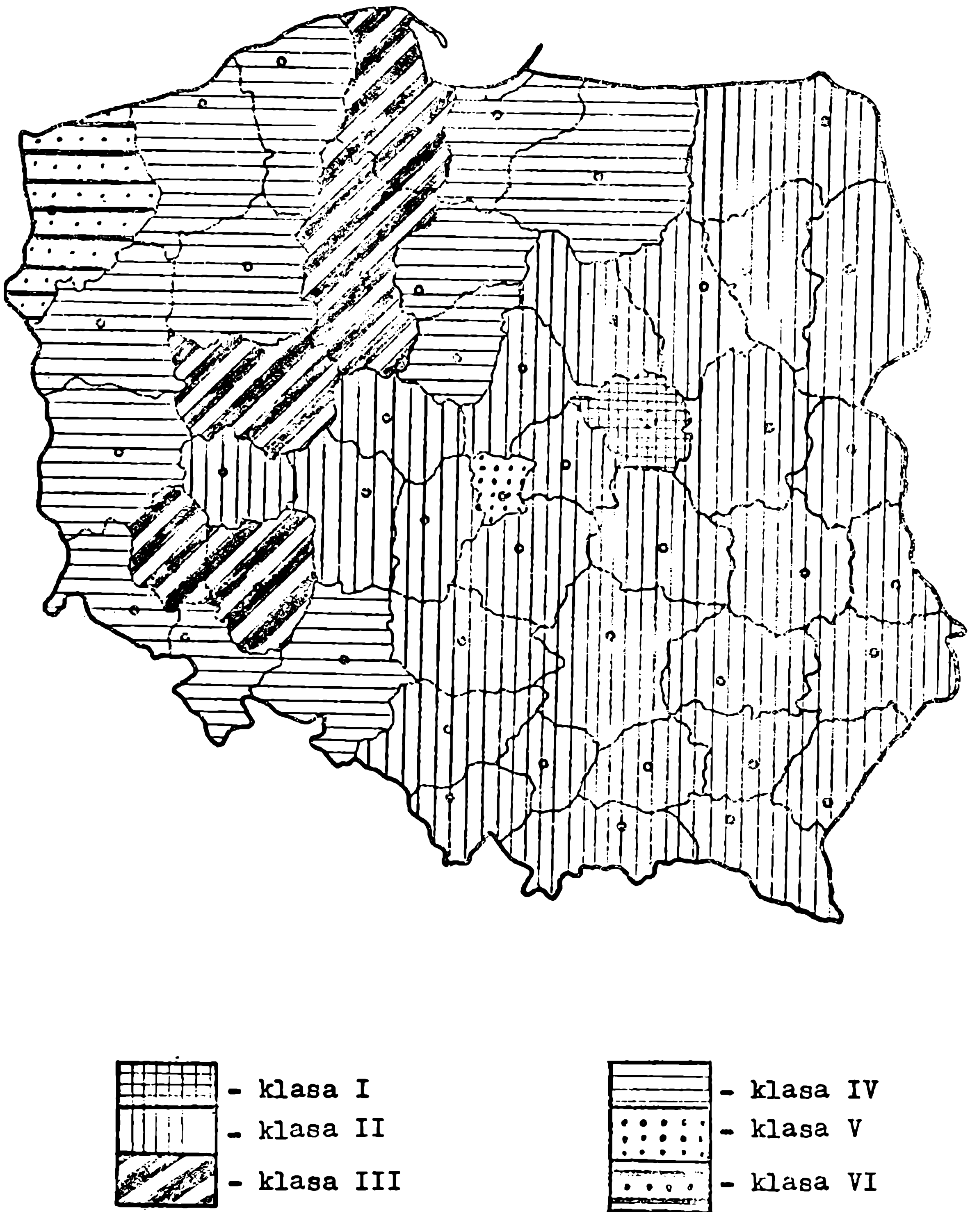
VI - Szczecin.

Rysunek 5.2 przedstawia mapę układu obszarów typu regionalnego powyższej klasyfikacji, natomiast tablica 5.4 średnie arytmetyczne dla wszystkich zmiennych wyznaczone osobno dla każdej klasy.

Rzut oka na rys. 5.2 w sposób ewidentny potwierdza dużą różnicę między obszarami należącymi do Ziem Zachodnich i Północnych i dużymi aglomeracjami a resztą kraju. Województwa Polski środkowej i wschodniej /wyjąwszy woj. warszawskie i łódzkie, które utworzyły, jak poprzednio, odrębne klasy, oraz woj. krakowskie i katowickie/ utworzyły jedną, bardzo liczną klasę, przy czym średnie wartości wszystkich zmiennych /oprócz zmiennych obrazujących usługi biblioteczne/ są zdecydowanie najniższe spośród wszystkich klas.

Województwa Ziem Zachodnich i Północnych utworzyły trzy klasy: woj. szczecińskie, województwa, których stolicami są duże miasta /w tej klasie znajdują się również woj. krakowskie, katowickie i legnickie/ oraz klasa zawierająca województwa mniejsze pod względem liczby ludności, dla której to klasy obserwujemy niższe wartości zmiennych niż dla dwóch wyżej wymienionych.

Rys.5.2. Klasyfikacja województw Polski ze względu na poziom usług kulturalno-rozrywkowych - wariant drugi.



Tablica 5.4

Średnie arytmetyczne zmiennych charakteryzujących usługi kulturalno-rozrywkowe dla poszczególnych klas - - klasyfikacja druga

Nr klasy	Zmienna							
	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$
1	2043	302	271	15,3	931	462	23,8	163
2	2685	206	182	11,5	117	125	18,8	72
3	2261	259	236	13,7	268	315	21,7	116
4	2972	245	225	18,9	271	229	18,7	95
5	2361	322	288	13,3	511	716	16,9	127
6	3411	270	245	24,5	330	424	23,0	123

Źródło: obliczenia własne.

Tablica 5.5 przedstawia wartości miary rozwoju dla poszczególnych województw. Potwierdzają one zasygnalizowane powyżej fakty, tzn. wysoki poziom usług w województwie warszawskim, łódzkim i szczecińskim, wysokie pozycje województw zachodnich, a niskie środkowych i wschodnich. Potwierdzony jest również stosunkowo wysoki poziom usług w województwie suwalskim.

Należy zwrócić uwagę, iż województwa, które znajdują się w jednej klasie klasyfikacji zajmują podobne pozycje po uszeregowaniu wg wartości miary rozwoju. Wyjątkiem jest jedynie woj. legnickie, mające stosunkowo niską, jak na województwo zachodnie, pozycję.

Tablica 5.5

Wartości miary rozwoju usług kulturalno-rozrywkowych

Lp.	Nazwa województwa	Wartość miary rozwoju	Lp.	Nazwa województwa	Wartość miary rozwoju
1	Warszawa	54,2911	33	Przemyśl	36,6855
2	Szczecin	52,2017	34	Chełm	36,4826
3	Łódź	49,9495	35	Skierniewice	36,0234
4	Wałbrzych	46,6095	36	Tarnów	35,2468
5	Wrocław	45,8432	37	Sieradz	35,0884
6	Poznań	45,7563	38	Radom	34,9691
7	Koszalin	45,5571	39	Łomża	34,9334
8	Elbląg	45,4891	40	Piotrków Tryb.	34,7297
9	Gdańsk	44,9441	41	Ciechanów	34,5247
10	Bydgoszcz	44,6007	42	Konin	34,4948
11	Toruń	44,5862	43	Biała Podlaska	34,1211
12	Kraków	43,3389	44	Zamość	34,0461
13	Gorzów Wlkp.	43,1754	45	Krosno	33,2654
14	Katowice	43,0819	46	Nowy Sącz	33,0309
15	Włocławek	43,0301	47	Tarnobrzeg	32,2891
16	Jelenia Góra	42,8006	48	Siedlce	31,1578
17	Olsztyn	42,7890	49	Ostrołęka	29,8653
18	Opole	42,6746			
19	Zielona Góra	42,1128			
20	Suwałki	41,0795			
21	Piła	41,0017			
22	Leszno	40,9533			
23	Słupsk	40,3025			
24	Legnica	40,1086			
25	Białystok	39,4484			
26	Rzeszów	38,2315			
27	Częstochowa	37,7985			
28	Bielsko-Biała	37,7785			
29	Lublin	37,6802			
30	Kielce	37,4728			
31	Kalisz	37,3850			
32	Płock	37,0370			

Źródło: obliczenia własne.

5.3.3. Klasyfikacja dla usług w zakresie ochrony zdrowia i opieki społecznej.

Rezultaty grupowania dla tej grupy przedstawione są w tablicy 5.6.

Tablica 5.6

Rezultaty grupowania dla usług w zakresie ochrony i opieki społeczne

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,3099	0,7431
35	1,3099	0,7455
36	1,3099	0,7536
37	1,3325	0,7657
38	1,3325	0,7714
39	1,7026	0,8861
40	1,7026	0,8891
41	1,7026	0,8925
42	2,2623	0,9526
43	2,2623	0,9563
44	2,4197	1,0353
45	2,6337	1,0588

Źródło: obliczenia własne.

Największe spadki podobieństwa obserwuje się po 38-mym oraz 43-cim etapie grupowania.

Klasyfikacja pierwsza, o większym stopniu szczegółowości składa się z 11 klas:

I - Warszawa,

II - Biała Podlaska, Ostrołęka, Przemyśl,

III - Białystok, Gorzów Wlkp., Słupsk, Suwałki,

IV - Bielsko-Biała, Bydgoszcz, Chełm, Ciechanów, Częstochowa, Elbląg, Kalisz, Kielce, Krosno, Leszno, Łomża, Nowy Sącz, Piła, Piotrków Tryb., Płock, Radom, Rzeszów, Siedlce, Sieradz, Skierniewice, Tarnobrzeg, Tarnów, Toruń, Włocławek, Zamość,

V - Gdańsk, Lublin,

VI - Jelenia Góra, Koszalin, Legnica, Olsztyn, Opole, Wałbrzych, Zielona Góra,

VII - Katowice,

VIII - Konin,

IX - Kraków, Poznań, Wrocław,

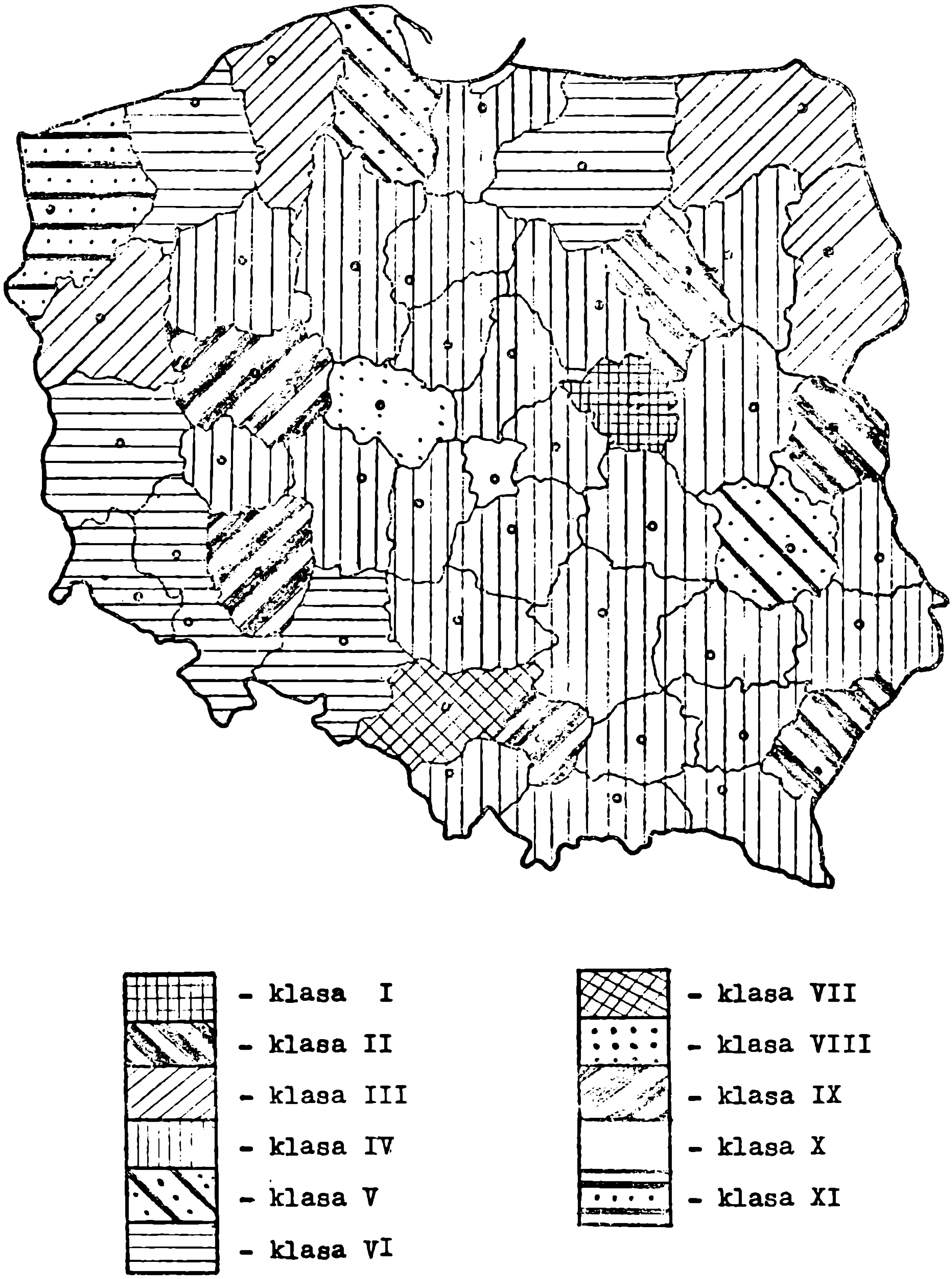
X - Łódź,

XI - Szczecin.

Rysunek 5.3 przedstawia graficzny układ obszarów typu regionalnego, a tablica 5,7 średnie wartości zmiennych. Wynika z niej, że najniższymi /a w przypadku destymulant - $X_{2,6}$, $X_{2,12}$ i $X_{2,13}$ - najwyższymi/ wartościami zmiennych dysponuje klasa jednoelementowa - woj. konińskie, a w drugiej kolejności klasa II, zawierająca trzy nisko rozwinięte województwa wschodnie: bialskopodlaskie, przemyskie i ostrołęckie, a także najliczniejsza klasa IV zawierająca większość województw Polski środkowej i wschodniej. Należy zwrócić uwagę na mniejsze zróżnicowanie w porównaniu z usługami kulturalno-rozrywkowymi między województwami zachodnimi i wschodnimi. Widać to na przykładzie klas III i V, gdzie znalazły się województwa Ziem Zachodnich i Północnych oraz województwa wschodnie.

Najwyższy poziom usług obserwuje się w dwóch klasach jednoelementowych: woj. warszawskim i łódzkim, a w dalszej kolejności klasie IX i XI, zawierającej województwa skoncentrowane wokół wielkich aglomeracji.

Rys.5.3. Klasyfikacja województw Polski ze względu na poziom usług w zakresie ochrony zdrowia i opieki społecznej - wariant pierwszy.



Tablica 5.7

Średnie arytmetyczne zmiennych charakteryzujących usługi w zakresie ochrony zdrowia i opieki społecznej dla poszczególnych klas - klasyfikacja pierwsza

Nr klasy	Zmienna										
	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,6}$	$X_{2,7}$	$X_{2,9}$	$X_{2,10}$	$X_{2,11}$	$X_{2,12}$	$X_{2,13}$
1	44,1	11,3	64,0	78,8	20,7	10,9	4,8	8,9	3,6	5,8	14
2	10,3	2,9	35,5	37,0	24,6	3,0	4,5	14,0	0,8	4,7	82
3	16,8	4,2	48,6	58,1	20,6	2,8	6,3	15,4	0,7	3,5	15
4	12,0	3,6	38,0	42,9	20,8	3,0	4,1	11,4	0,9	4,7	35
5	24,3	5,9	43,8	58,3	23,5	8,8	4,8	9,2	1,5	5,1	22
6	14,2	4,0	49,9	63,3	19,4	3,1	4,0	11,6	0,9	4,3	11
7	18,6	4,7	50,2	71,2	22,9	3,4	4,5	8,4	3,4	4,9	21
8	8,9	3,3	29,0	36,7	22,8	2,2	4,1	7,8	0,8	6,5	62
9	28,9	6,9	56,4	67,3	18,3	8,0	5,7	10,3	1,7	4,7	19
10	34,2	9,2	64,5	64,8	20,5	9,2	6,6	14,5	5,2	4,1	7
11	23,6	7,0	52,3	60,8	23,6	3,6	6,1	13,3	1,1	5,6	12

Źródło: obliczenia własne.

Klasyfikacja druga składa się z 6 klas:

I - Warszawa,

II - Biała Podlaska, Bielsko-Biała, Bydgoszcz, Chełm, Ciechanów, Częstochowa, Elbląg, Jelenia Góra, Kalisz, Kielce, Koszalin, Krosno, Legnica, Leszno, Łomża, Nowy Sącz, Olsztyn, Opole, Ostrołęka, Piła, Piotrków Tryb., Płock, Przemyśl, Radom, Rzeszów, Siedlce, Sieradz, Skierniewice, Tarnobrzeg, Tarnów, Toruń, Wałbrzych, Włocławek, Zamość, Zielona Góra,

III - Białystok, Gorzów Wlkp., Słupsk, Suwałki,

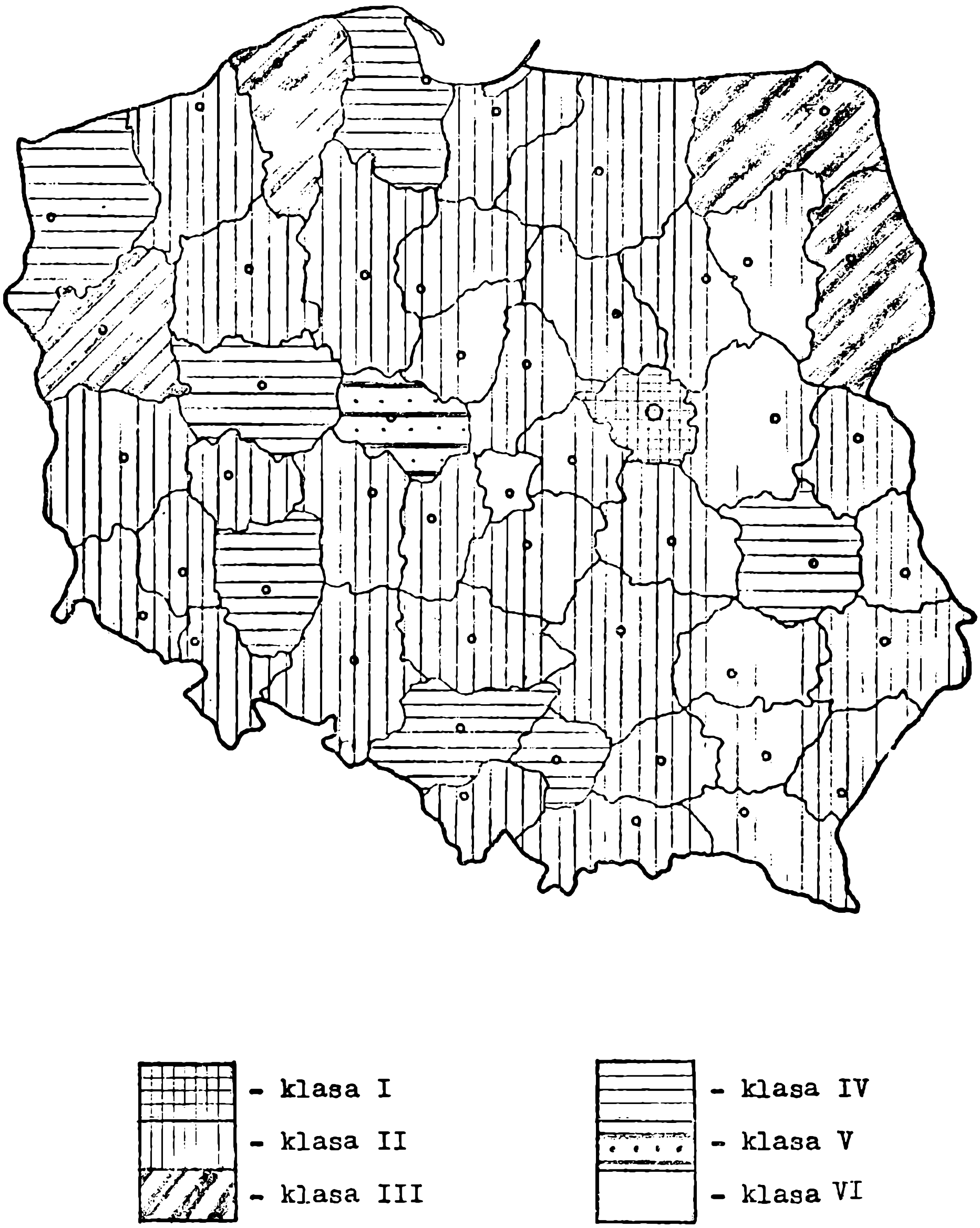
IV - Gdańsk, Katowice, Kraków, Lublin, Poznań, Szczecin, Wrocław,

V - Konin,

VI - Łódź.

Tablica 5.8 i rysunek 5.4 prezentują przeciętne wartości zmiennych i mapkę układu obszarów typu regionalnego.

RYS.5.4. Klasyfikacja województw Polski ze względu na poziom usług w zakresie ochrony zdrowia i opieki społecznej - wariant drugi.



Tablica 5.8

Średnie artymetyczne zmiennych charakteryzujących usługi w zakresie ochrony zdrowia i opieki społecznej dla poszczególnych klas - klasyfikacja druga

Nr klasy	Zmienna										
	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,6}$	$X_{2,7}$	$X_{2,9}$	$X_{2,10}$	$X_{2,11}$	$X_{2,12}$	$X_{2,13}$
1	44,1	11,3	64,0	78,8	20,7	10,9	4,8	8,9	3,6	5,8	14
2	12,3	3,7	40,1	47,5	21,4	3,1	4,2	11,7	0,9	4,7	35
3	16,8	4,2	48,6	58,1	20,6	2,8	6,3	15,4	0,7	3,5	15
4	25,3	6,3	51,3	64,7	21,2	7,0	5,3	10,1	1,8	5,0	19
5	8,9	3,3	29,0	36,7	22,8	2,2	4,1	7,8	0,8	6,5	62
6	34,2	9,2	64,5	64,8	20,5	9,2	6,6	14,5	5,2	4,1	7

Źródło: obliczenia własne.

Charakterystyczną cechą tego układu jest mniejsze zróżnicowanie między Polską zachodnią i resztą kraju. Wyodrębniły się tutaj trzy klasy jednoelementowe. Są to województwa o najwyższym poziomie usług: warszawskie i łódzkie oraz o najniższym poziomie usług - województwo konińskie.

Klasa IV zawiera województwa, których stolicami są wielkie aglomeracje, jest to w kolejności poziomu usług trzecia klasa. Pozostałe województwa /z wyjątkiem czterech/ znalazły się w klasie II. Są tam zarówno województwa zachodnie, jak i środkowe oraz wschodnie. Pewną niespodzianką jest klasa III, zawierająca województwa o wyższym poziomie usług, niż najbardziej liczna klasa II. Udział woj. białostockiego i suwalskiego można wytłumaczyć chyba tym, że w Białymstoku jest Akademia Medyczna. Siłą rzeczy powoduje to wzrost wartości pewnych zmiennych, zwłaszcza zmiennych charakteryzujących poziom kadr służby zdrowia. Tablica 5,9 przedstawia uporządkowane wartości miary rozwoju. Potwierdza ona powyższę fakty, tzn. wyraźnie wysoki poziom usług w woj. Warszawskim i łódzkim oraz wyraźnie niski w województwie konińskim.

Tablica 5.9

Wartość miary rozwoju usług w zakresie ochrony zdrowia i opieki społecznej

Lp.	Nazwa województwa	Wartość miary rozwoju	Lp.	Nazwa województwa	Wartość miary rozwoju
1	Łódź	67,7923	26	Krosno	42,7514
2	Warszawa	61,3895	27	Rzeszów	42,6478
3	Kraków	56,7209	28	Toruń	41,5354
4	Białystok	56,2432	29	Łomża	41,3952
5	Wrocław	55,8615	30	Przemyśl	40,8074
6	Poznań	52,8436	31	Bielsko-Biała	40,4602
7	Gorzów Wlkp.	51,1971	32	Leszno	40,1954
8	Słupsk	49,9209	33	Piła	40,1646
9	Szczecin	49,5845	34	Piotrków Tryb.	40,0373
10	Olsztyn	49,0835	35	Tarnów	40,0321
11	Koszalin	49,0526	36	Włocławek	39,8842
12	Wałbrzych	48,8057	37	Chełm	39,8502
13	Jelenia Góra	47,9636	38	Kalisz	39,2554
14	Gdańsk	47,7745	39	Częstochowa	39,1918
15	Katowice	47,0977	40	Płock	38,9104
16	Suwałki	47,0338	41	Skierniewice	38,8462
17	Lublin	46,8678	42	Biała Podlaska	38,3152
18	Zielona Góra	46,2103	43	Radom	38,2450
19	Legnica	46,1414	44	Ciechanów	37,8726
20	Nowy Sącz	44,7637	45	Tarnobrzeg	37,7783
21	Opole	44,2221	46	Zamość	36,1870
22	Elbląg	44,1076	47	Siedlce	36,1142
23	Kielce	43,9620	48	Ostrołęka	35,8424
24	Bydgoszcz	43,4812	49	Konin	32,9376
25	Sieradz	43,0639			

Źródło: obliczenia własne.

Potwierdza się wysoki poziom woj. białostockiego. W przypadku niektórych klas jak np. klasy II, obejmującej woj. ostrołęckie bialsko-podlaskie i przemyskie, występuje dość istotna różnica w pozycjach zajmowanych przez te województwa na skali miary rozwoju.

Różnica między klasyfikacją a uporządkowaniem wg miary rozwoju wynika z faktu, iż klasyfikacja tworzona jest na bazie wzajemnych podobieństw między obiektami, natomiast liniowe uporządkowanie bezpośrednio na podstawie wartości zmiennych.

5.3.4. Klasyfikacja dla usług komunalno-mieszkaniowych.

Rezultaty grupowania przedstawione są w tablicy 5.10.

Tablica 5.10

Rezultaty grupowania dla usług komunalno-mieszkaniowych

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,0842	0,6453
35	1,1912	0,6845
36	1,3736	0,7359
37	1,4937	0,7784
38	1,4937	0,7834
39	1,6803	0,9076
40	1,6803	0,9107
41	1,6803	0,9151
42	1,6803	0,9194
43	1,8572	0,9559
44	2,1053	1,0165
45	2,1053	1,0360

Źródło: obliczenia własne.

Podobnie jak w przypadku poprzednich grup usług, wyróżnione zostały dwie klasyfikacje, otrzymane w etapach, po których obserwuje się największe straty informacji /a zatem 38-my i 43-ci/.

Klasyfikacja o większym stopniu szczegółowości liczy 11 klas:

I - Warszawa,

II - Biała Podlaska, Łomża,

III - Białystok, Bydgoszcz, Częstochowa, Gorzów Wlkp., Jelenia Góra, Kalisz, Lublin, Olsztyn, Przemyśl, Radom, Toruń, Włocławek, Zielona Góra,

IV - Bielsko-Biała,

V - Chełm, Ciechanów, Konin, Leszno, Ostrołęka, Piła, Piotrków Tryb., Płock, Siedlce, Sieradz, Skierniewice, Suwałki, Tarnobrzeg, Zamość,

VI - Elbląg, Koszalin, Legnica, Opole, Słupsk, Wałbrzych,

VII - Gdańsk, Kraków, Łódź, Szczecin, Wrocław,

VIII - Katowice,

IX - Kielce, Nowy Sącz, Rzeszów, Tarnów,

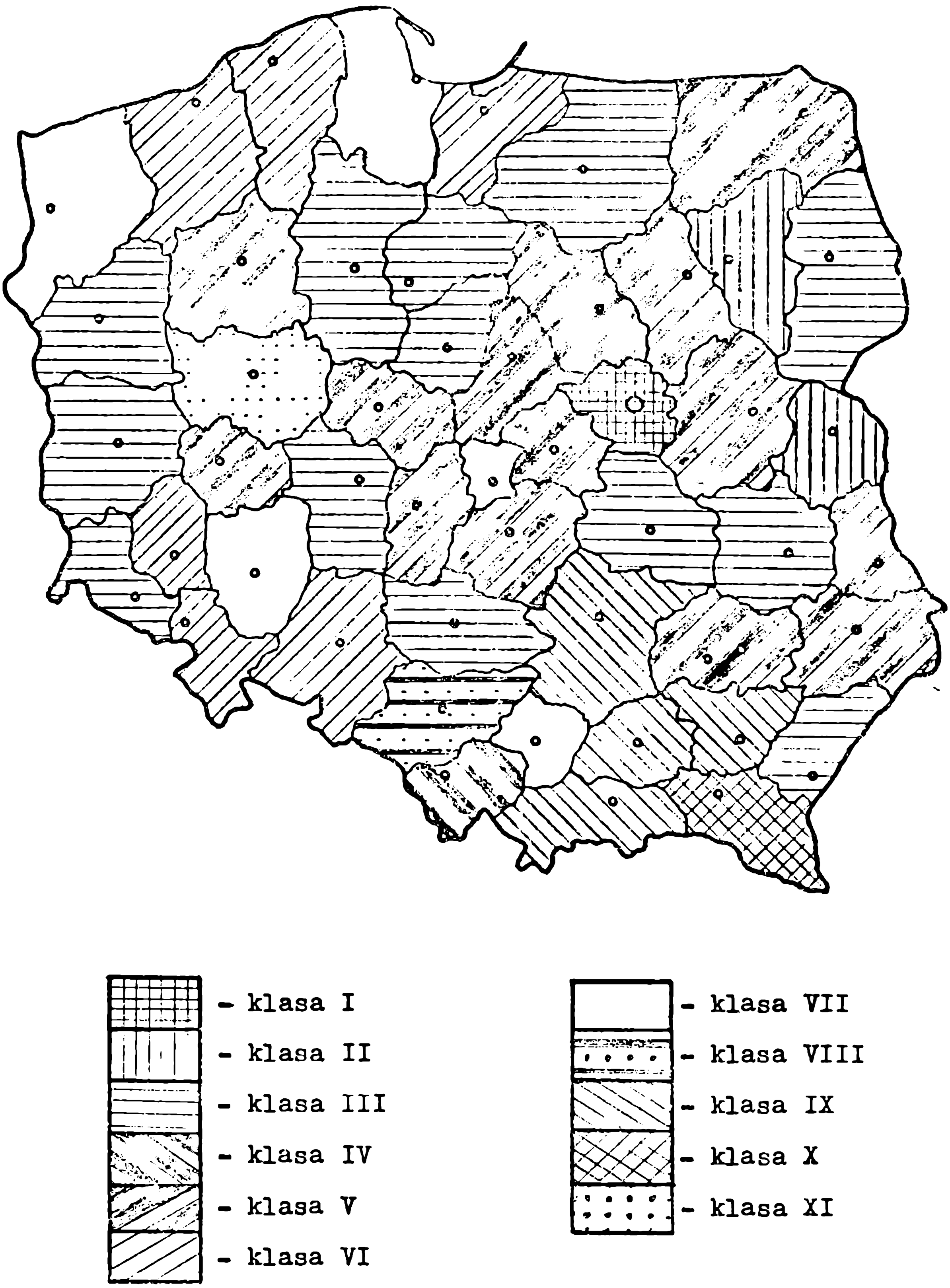
X - Krosno,

XI - Poznań.

Rysunek 5.5 przedstawia układ graficzny powyższej klasyfikacji, a tablica 5.11 średnie wartości zmiennych w poszczególnych klasach.

Wyodrębniły się klasy o wyższych przeciętnych wartościach zmiennych: klasa II - woj. warszawskie /oprócz zmiennej $X_{3,3}$ - procent długości jezdni w miastach o nawierzchni twardej ulepszonej/, klasa VI zawierająca niektóre województwa zachodnie i północne, klasa VII - województwa, których stolicami są wielkie aglomeracje, klasa VIII - województwo katowickie oraz klasa XI - woje-

Rys. 5.5. Klasyfikacja województw Polski ze względu na poziom usług komunalno-mieszkaniowych - wariant pierwszy.



Tablica 5.11

Średnie arytmetyczne zmiennych charakteryzujących usługi komunalno-mieszkaniowe dla poszczególnych klas - klasyfikacja pierwsza.

Nr klasy	Zmienna						
	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$
1	77,4	43,2	11,8	108,2	62,2	453,1	1234
2	46,1	43,9	8,0	17,1	6,9	146,6	677
3	72,6	53,1	11,6	24,6	25,1	229,9	952
4	69,5	64,4	10,0	23,8	24,1	406,7	1031
5	58,4	58,1	10,1	18,2	13,5	183,8	626
6	87,5	72,1	13,1	24,4	36,0	268,8	720
7	86,6	62,0	10,3	48,5	48,4	355,7	1063
8	81,1	66,4	12,9	28,6	72,2	395,1	646
9	69,4	59,0	10,3	20,5	17,6	172,9	1421
10	58,8	62,2	24,3	16,2	13,3	155,7	1069
11	79,2	49,2	18,1	39,4	34,9	366,7	1162

Źródło: obliczenia własne.

wództwo poznańskie. Są to zatem województwa bardziej zurbanizowane. Wyodrębnienie się w jednoelementową klasę województwa krośnieńskiego wynika w dużym stopniu z wysokiej wartości zmiennej $X_{3,4}$ - powierzchnia parków i zieleńców w m^2 przypadająca na 1 mieszkańca miast. Najniższym poziomem usług komunalno-mieszkaniowych charakteryzują się klasa II i V, zawierające większość województw środkowych i wschodnich, ale również kilka województw Polski zachodniej. Charakterystyczne jest, iż większość województw małopolskich znalazła się w jednej klasie /klasa IX/. Interesujące wyniki zawiera również klasyfikacja druga, licząca 6 klas.

I - Warszawa,

II - Biała Podlaska, Białystok, Bydgoszcz, Chełm, Ciechanów, Częstochowa, Gorzów Wlkp., Jelenia Góra, Kalisz, Konin, Leszno, Lublin, Łomża, Olsztyn, Ostrołęka, Piła, Piotrków Tryb., Płock, Przemyśl, Radom, Siedlce, Sieradz, Skierniewice, Suwałki, Tarnobrzeg, Toruń, Włocławek, Zamość, Zielona Góra,

III - Bielsko-Biała, Gdańsk, Kraków, Łódź, Poznań, Szczecin, Wrocław,

IV - Elbląg, Katowice, Koszalin, Legnica, Opole, Słupsk, Wałbrzych,

V - Kielce, Nowy Sącz, Rzeszów, Tarnów,

VI - Krosno.

Rysunek 5.6 przedstawia układ obszarów typu regionalnego, a tablica 5.12 średnie wartości zmiennych.

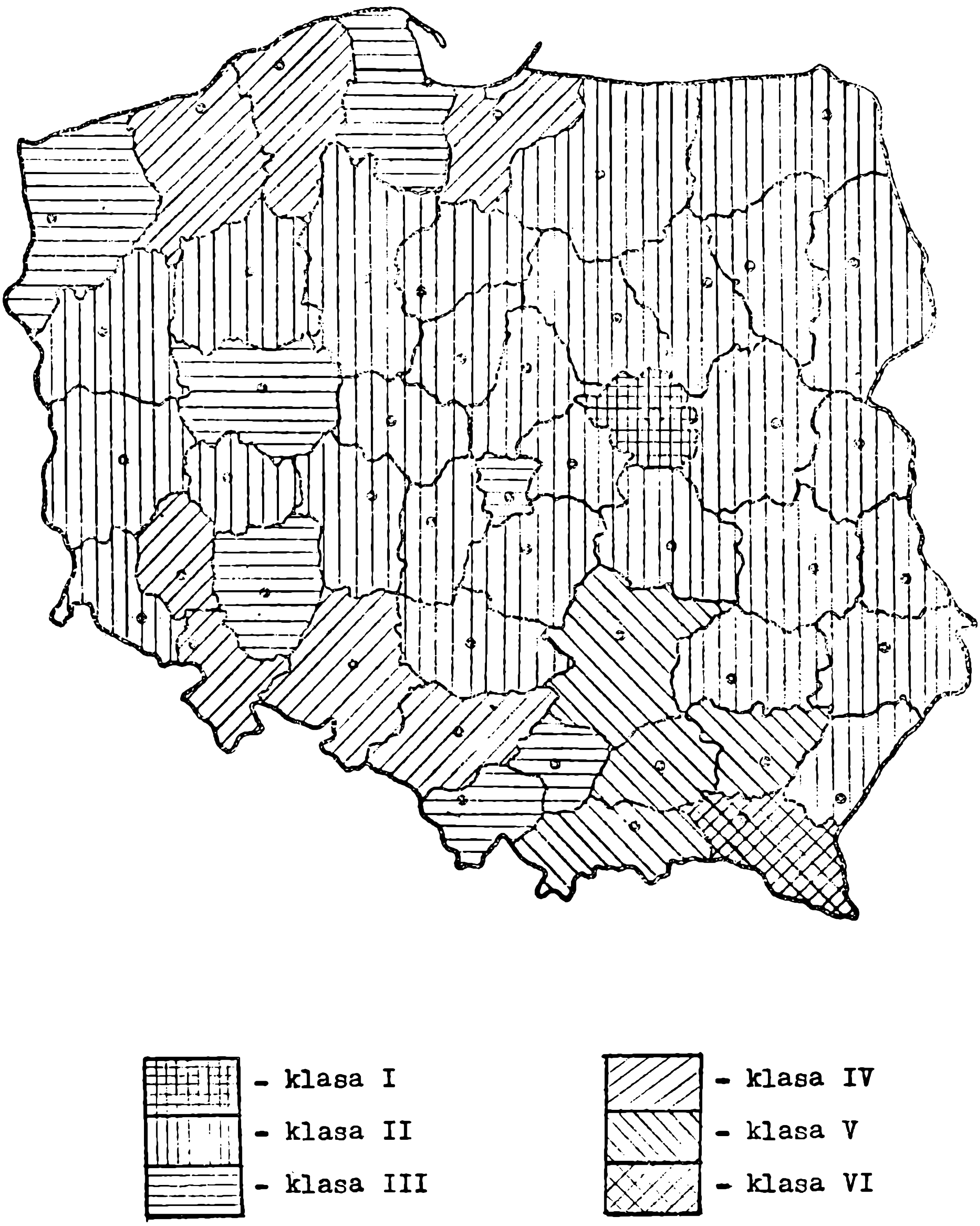
Tablica 5.12

Średnie arytmetyczne zmiennych charakteryzujących usługi komunalno-mieszkaniowe dla poszczególnych klas - klasyfikacja druga.

Nr klasy	Zmienna						
	$x_{3,2}$	$x_{3,3}$	$x_{3,4}$	$x_{3,6}$	$x_{3,7}$	$x_{3,8}$	$x_{3,9}$
1	77,4	43,2	11,8	108,2	62,2	453,1	1234
2	67,2	54,9	10,6	20,8	19,0	379,3	750
3	83,1	60,5	11,4	43,6	43,0	364,6	1073
4	86,6	71,3	13,1	25,0	41,2	286,9	709
5	69,4	59,0	10,3	20,5	17,6	172,9	1421
6	58,8	62,2	24,3	16,2	13,3	155,7	1069

Źródło: obliczenia własne.

W zasadzie klasyfikacja ta nie różni się bardzo od poprzedniej, z tym, że woj. bielsko-bialskie znajduje się wraz z województwami zawierającymi wielkie aglomeracje miejskie /klasa III - jest to klasa o najwyższym po woj. warszawskim poziomie usług/, Rys.5.6. Klasyfikacja województw Polski ze względu na poziom usług komunalno-mieszkaniowych - wariant drugi.



Tablica 5.13

Wartość miary rozwoju usług komunalno-mieszkaniowych

Lp.	Nazwa województwa	Wartość miary rozwoju	Lp.	Nazwa województwa	Wartość miary rozwoju
1	Warszawa	35,6158	26	Piła	24,7067
2	Wrocław	34,3415	27	Nowy Sącz	24,6713
3	Łódź	32,0334	28	Kielce	24,6167
4	Katowice	31,2833	29	Kalisz	23,1766
5	Szczecin	30,9211	30	Tarnów	23,0807
6	Koszalin	30,7461	31	Płock	22,6044
7	Poznań	30,5765	32	Radom	22,5644
8	Kraków	30,4660	33	Przemyśl	22,4698
9	Legnica	29,4144	34	Suwałki	22,3897
10	Gdańsk	29,3522	35	Konin	22,2854
11	Wałbrzych	29,2783	36	Leszno	22,1416
12	Słupsk	28,4392	37	Częstochowa	21,8929
13	Rzeszów	27,4793	38	Sieradz	21,1593
14	Bielsko-Biała	27,4660	39	Skierniewice	21,1446
15	Elbląg	27,3688	40	Ciechanów	20,5802
16	Toruń	27,1479	41	Białystok	20,5275
17	Gorzów Wlkp.	27,0982	42	Ostrołęka	20,4030
18	Jelenia Góra	27,0359	43	Piotrków Tryb.	20,3879
19	Opole	26,5844	44	Chełm	19,9552
20	Lublin	26,4625	45	Zamość	19,7574
21	Krosno	26,1641	46	Tarnobrzeg	19,1263
22	Bydgoszcz	26,0815	47	Siedlce	17,9058
23	Olsztyn	25,8676	48	Łomża	17,1334
24	Zielona Góra	25,6215	49	Biała Podlaska	16,2826
25	Włocławek	24,7216			

Źródło: obliczenia własne.

a województwa środkowe, wschodnie, a [illegible] kilka zachodnich znalazło się w jednej klasie /klasa II/.
Tablica 5.13 przedstawia uporządkowane wartości miary rozwoju dla tej grupy usług. Wskazuje ona na większe zróżnicowanie między zachodnią a wschodnią Polską niż to wynika z klasyfikacji. Najwyżej rozwinięte są województwa zawierające wielkie aglomeracje miejskie i wydaje się to zrozumiałe.

5.3.5. Klasyfikacja dla usług w zakresie oświaty i wychowania.

Rezultaty grupowania przedstawione są w tablicy 5.14.

Tablica 5.14

Rezultaty grupowania dla usług w zakresie oświaty i wychowania

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,2673	0,7994
35	1,2673	0,8134
36	1,4372	0,8641
37	1,4372	0,8652
38	1,4372	0,9220
39	1,4372	0,9258
40	1,7487	0,9717
41	1,7487	0,9725
42	1,7739	1,0403
43	1,8935	1,0607
44	2,0559	1,2361
45	2,0559	1,2366

Źródło: obliczenia własne.

W wyniku ich analizy wybrano, podobnie jak w poprzednich przypadkach, dwie klasyfikacje, mniej i bardziej szczegółową.

Klasyfikacja pierwsza składa się z 14 klas:

I - Warszawa, Łódź,

II - Biała Podlaska, Kielce, Krosno, Przemyśl, Zamość,

III - Białystok, Lublin, Rzeszów,

IV - Bielsko-Biała, Kraków,

V - Bydgoszcz, Gdańsk, Poznań, Wrocław,

VI - Chełm, Ciechanów, Łomża, Ostrołęka, Płock, Radom, Siedlce, Suwałki, Tarnobrzeg, Włocławek,

VII - Częstochowa, Konin, Nowy Sącz, Piotrków Tryb., Sieradz, Skierniewice, Tarnów,

VIII - Elbląg, Gorzów Wlkp., Kalisz, Koszalin, Legnica, Leszno, Piła, Słupsk, Szczecin, Zielona Góra,

IX - Jelenia Góra,

X - Katowice,

XI - Olsztyn,

XII - Opole,

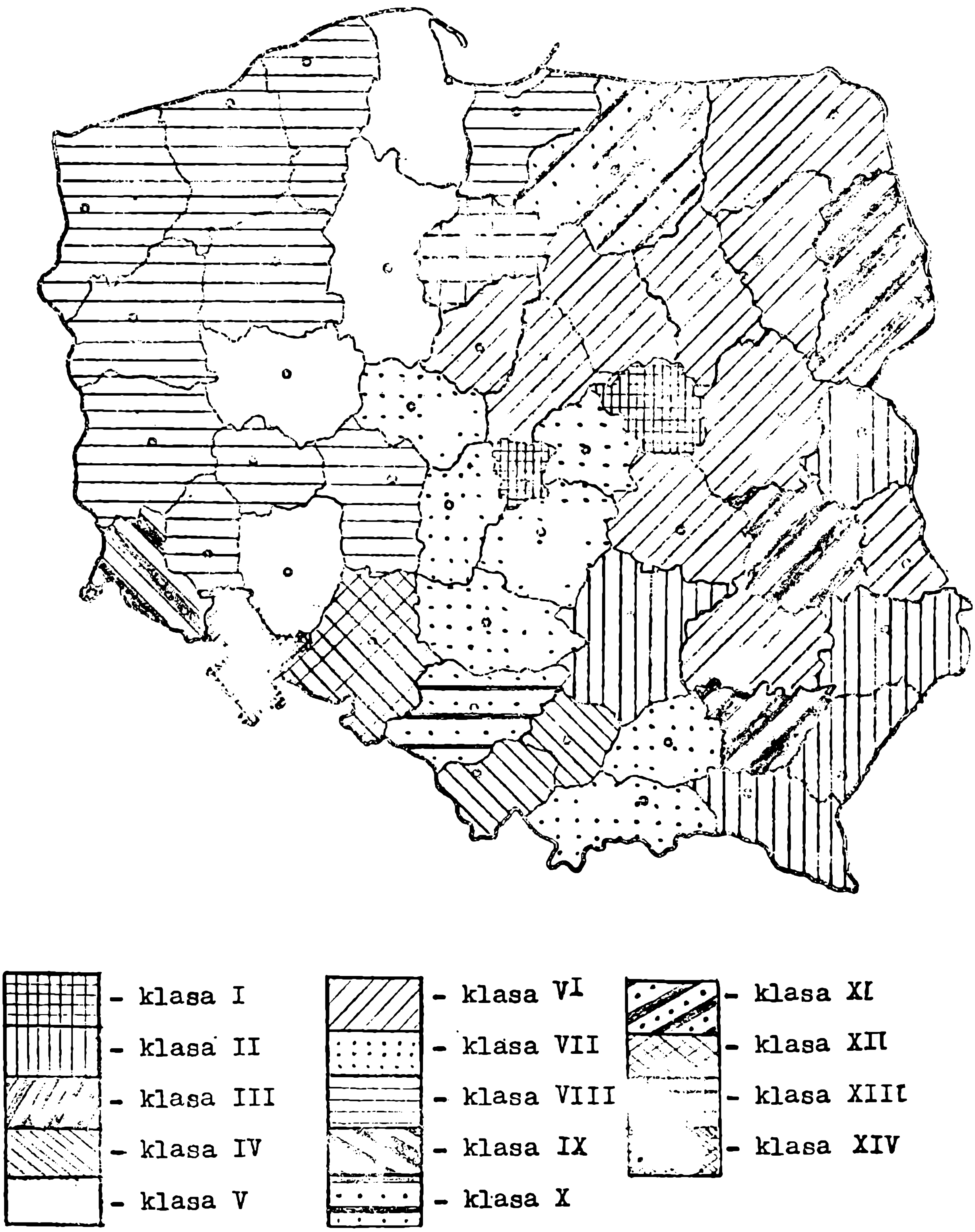
XIII - Toruń,

XIV - Wałbrzych.

Ponieważ, jak wspomnieliśmy, rezultaty te należy traktować z pewną rezerwą, ograniczymy się do krótkiego komentarza, odsyłając do rys. 5.7 i tablicy 5.15.

Zwraca uwagę fakt, iż na podstawie analizy średnich wartości zmiennych trudno wyróżnić klasy,dla których wszystkie wartości zmiennych byłyby wysokie /lub w przypadku destymulant $X_{4,2}$, $X_{4,3}$, $X_{4,4}$, $X_{4,8}$, $X_{4,9}$, $X_{4,10}$ - niskie/. Również w tej grupie usług widoczne jest terytorialne zróżnicowanie między województwami zawierającymi wielkie aglomeracje /utworzyły one klasy I oraz V/ a resztą kraju, a także województwami zachodnimi /klasy V i VIII/ a środkowymi i wschodnimi /klasy II, VI i VII/.

Rys.5.7. Klasyfikacja województw Polski ze względu na poziom usług w zakresie oświaty i wychowania - wariant pierwszy.



Tablica 5.15

Średnie arytmetyczne zmiennych charakteryzujących usługi w zakresie oświaty i wychowania dla poszczególnych klas - klasyfikacja pierwsza.

Nr klasy	Zmienna								
	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$
1	586,1	17,9	125,1	14,8	29,2	74,8	23	31	32
2	419,2	16,3	123,2	16,5	19,9	65,2	21	25	32
3	465,1	17,3	132,9	15,2	26,1	71,4	21	27	33
4	480,1	16,5	123,4	15,1	25,2	65,7	22	29	30
5	499,2	18,0	131,5	19,9	25,0	64,0	22	30	33
6	425,2	17,3	133,3	17,6	21,8	58,1	21	26	31
7	449,2	17,5	119,4	14,4	19,9	56,6	21	27	32
8	495,9	18,8	121,2	20,0	29,0	59,6	22	28	31
9	523,6	18,1	114,1	24,4	28,4	60,4	22	28	27
10	464,3	18,4	114,2	19,3	23,9	67,4	26	30	34
11	510,3	16,4	120,9	21,2	41,4	62,3	21	26	31
12	572,3	17,1	107,8	17,8	32,1	64,3	22	28	29
13	451,0	22,5	128,8	22,9	19,4	63,3	22	28	34
14	521,0	17,9	109,8	25,3	29,7	59,0	21	26	28

Źródło: obliczenia własne.

Klasyfikacja druga składa się z 6 klas:

I - Warszawa, Łódź,

II - Biała Podlska, Białystok, Bielsko-Biała, Chełm, Ciechanów, Częstochowa, Kielce, Konin, Kraków, Krosno, Lublin, Łomża, Nowy Sącz, Ostrołęka, Piotrków Tryb., Płock, Przemyśl, Radom, Rzeszów, Siedlce, Sieradz, Skierniewice, Suwałki, Tarnobrzeg, Tarnów, Włocławek, Zamość,

III - Bydgoszcz, Elbląg, Gdańsk, Gorzów Wlkp., Kalisz, Koszalin, Legnica, Leszno, Olsztyn, Opole, Piła, Poznań, Słupsk, Szczecin, Wałbrzych, Wrocław, Zielona Góra,

IV - Jelenia Góra,

V - Katowice,

VI - Toruń.

Klasyfikacja ta /dotyczą jej również rysunek 5.8 i tablica 5.16/ niesie nieco ciekawsze informacje niż poprzednia. Wyodrębniają się tu dwie klasy relatywnie wysoko rozwinięte /I - zawierająca woj. warszawskie i łódzkie, IV - woj. jeleniogórskie/. Na przeciwnym biegunie znajduje się województwo toruńskie oraz częściowo katowickie. Pozostałe województwa utworzyły dwie klasy: jedna z nich zawiera województwa należące do Ziem Zachodnich i Północnych oraz województwa wielkopolskie, druga zawiera województwa środkowe i wschodnie, przy czym nie można jednoznacznie stwierdzić, która z tych klas jest wyżej rozwinięta.

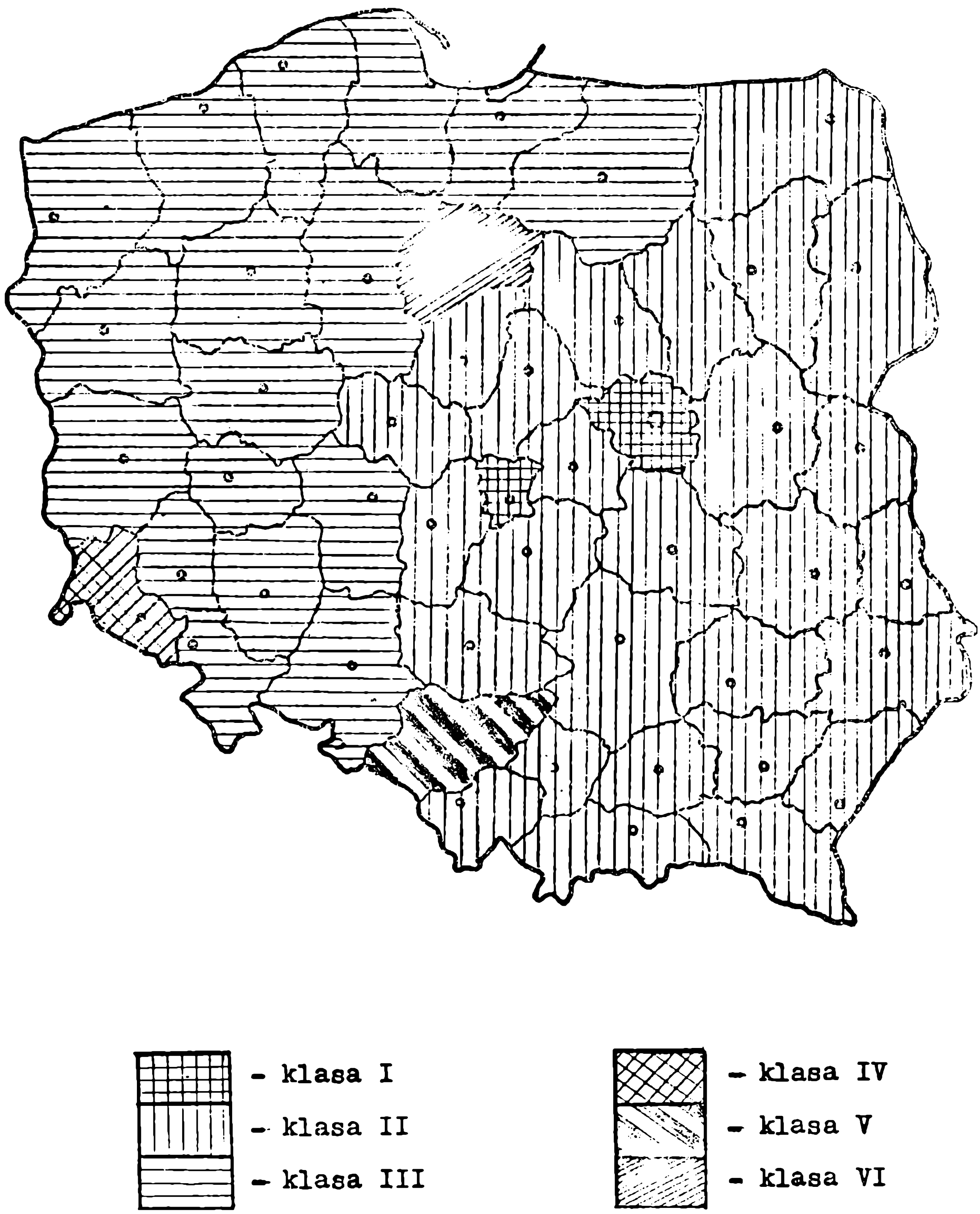
Tablica 5.16

Średnie arytmetyczne zmiennych charakteryzujących usługi w zakresie oświaty i wychowania dla poszczególnych klas - klasyfikacja druga

Nr klasy	Zmienna								
	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$
1	586,1	17,9	125,1	14,8	25,2	74,8	23	31	32
2	460,5	17,8	129,5	16,5	21,8	63,9	22	28	32
3	503,5	18,3	122,2	20,5	28,2	61,0	22	29	31
4	523,6	18,1	114,1	24,4	28,4	60,4	22	28	27
5	464,3	18,4	114,2	19,3	23,9	67,4	26	30	34
6	451,0	22,5	128,8	22,9	19,4	63,3	22	28	34

Tablica 5.17 przedstawia uporządkowane wartości miary rozwoju dla tej grupy usług. Potwierdza się wysoka pozycja woj. jeleniogórskiego i niska woj. toruńskiego. Występują stosunkowo niewielkie różnice między wartościami miary rozwoju /wynika to m.in. z małych zmienności zmiennych tej grupy/.

Rys. 5.8. Klasyfikacja województw Polski ze względu na poziom usług w zakresie oświaty i wychowania - wariant drugi.



Tablica 5.17

Wartość miary rozwoju usług w zakresie oświaty i wychowania

Lp.	Nazwa województwa	Wartość miary rozwoju	Lp.	Nazwa województwa	Wartość miary rozwoju
1	Jelenia Góra	107,0492	26	Łomża	100,3546
2	Opole	106,8286	27	Rzeszów	100,3125
3	Olsztyn	106,1842	28	Radom	100,2001
4	Biała Podlaska	105,9999	29	Szczecin	99,5861
5	Białystok	104,5836	30	Nowy Sącz	99,4309
6	Łódź	104,5266	31	Siedlce	99,4099
7	Wałbrzych	103,9436	32	Elbląg	99,2763
8	Kielce	103,7864	33	Tarnobrzeg	98,9340
9	Sieradz	103,2020	34	Ciechanów	98,7509
10	Lublin	103,1221	35	Kalisz	98,5838
11	Krosno	103,0689	36	Tarnów	98,5645
12	Piotrków Tryb.	102,6377	37	Piła	98,3866
13	Skierniewice	102,4934	38	Chełm	97,9328
14	Przemyśl	102,3323	39	Poznań	97,7507
15	Zamość	102,2817	40	Płock	97,5310
16	Kraków	102,2463	41	Włocławek	97,2976
17	Warszawa	101,8662	42	Legnica	96,8391
18	Bielsko-Biała	101,5944	43	Katowice	96,4020
19	Zielona Góra	101,2738	44	Leszno	95,7483
20	Suwałki	101,1992	45	Ostrołęka	95,5699
21	Słupsk	100,8040	46	Bydgoszcz	95,4153
22	Wrocław	100,7820	47	Konin	95,1674
23	Gorzów Wlkp.	100,7034	48	Gdańsk	94,3427
24	Częstochowa	100,7012	49	Toruń	92,5955
25	Koszalin	100,5206			

Źródło: obliczenia własne.

5.3.6. Klasyfikacja dla usług bytowych.

Rezultaty grupowania zamieszczone są w tablicy 5.18.

Tablica 5.18

Rezultaty grupowania usług bytowych

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,1836	0,6102
35	1,2230	0,6449
36	1,2230	0,6515
37	1,2230	0,6591
38	1,2230	0,6619
39	1,2230	0,6669
40	1,7404	0,8012
41	1,7404	0,8051
42	1,7404	0,8351
43	1,7404	0,8419
44	2,0930	0,9269
45	2,1378	1,0786

Źródło: obliczenia własne.

W efekcie wybrano dwie klasyfikacje, otrzymane po 39-tym i 43-cim etapie grupowania.

Klasyfikacja pierwsza, bardziej szczegółowa składa się z 10 klas:

I - Warszawa,

II - Biała Podlaska, Chełm, Konin, Krosno, Łomża, Ostrołęka, Piła, Przemyśl, Suwałki, Tarnobrzeg, Tarnów, Zamość,

III - Białystok, Elbląg, Kielce, Olsztyn, Opole, Piotrków Tryb., Radom, Rzeszów, Sieradz, Toruń, Włocławek,

IV - Bielsko-Biała, Częstochowa, Kalisz, Leszno,

V - Bydgoszcz, Gdańsk, Kraków, Poznań,

VI - Ciechanów, Skierniewice, Słupsk, Szczecin, Zielona Góra, VII - Gorzów Wlkp., Jelenia Góra, Katowice, Koszalin, Legnica Wałbrzych,

VIII - Lublin,

IX - Łódź, Wrocław,

X - Nowy Sącz, Płock, Siedlce.

Klasyfikacji tej dotyczy również rysunek 5.9 oraz tablica 5.19

Tablica 5.19

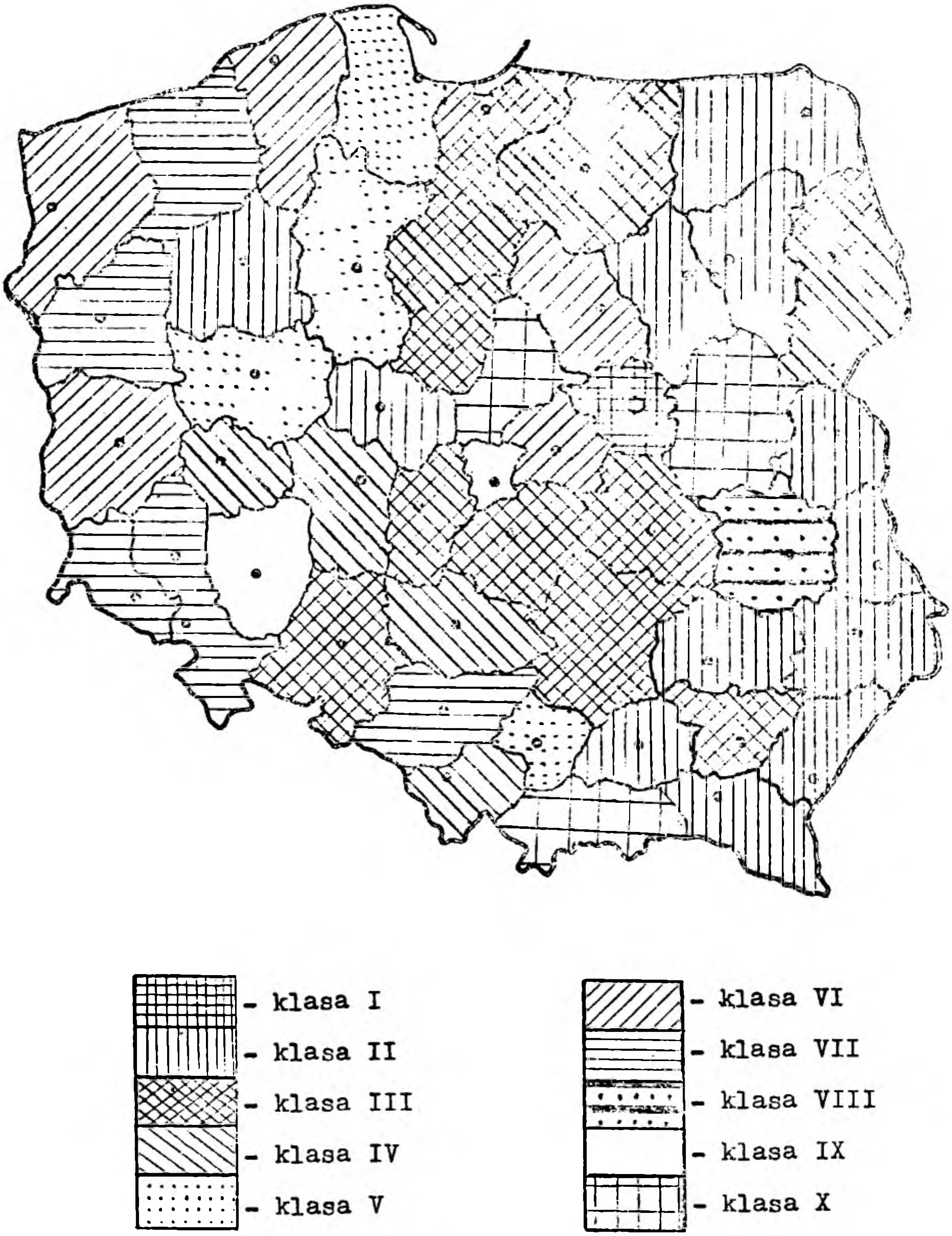
Średnie arytmetyczne zmiennych charakteryzujących usługi bytowe dla poszczególnych klas - klasyfikacja pierwsza.

Nr klasy	Zmienna					
	$X_{5,1}$	$X_{5,2}$	$X_{5,4}$	$X_{5,6}$	$X_{5,8}$	$X_{5,10}$
1	5071,4	114,8	627,7	169,1	324,6	408,6
2	3536,7	59,6	210,8	24,7	235,3	89,3
3	4945,2	69,2	243,2	38,7	340,3	138,4
4	5263,0	64,5	257,2	48,9	568,1	129,6
5	4949,3	77,5	355,2	98,1	386,6	272,9
6	6761,6	84,4	277,4	52,3	301,9	159,4
7	4751,1	86,6	300,4	63,5	217,4	165,5
8	5408,9	77,0	543,5	69,6	188,2	167,5
9	4689,8	98,3	566,6	114,5	247,9	193,5
10	6663,5	42,4	144,5	24,0	223,0	117,3

Źródło: obliczenia własne.

Widać, iż najwyższymi wartościami zmiennych charakteryzuje się klasa, zawierająca woj. warszawskie, a w dalszej kolejności klasy V i IX zawierające województwa, których stolicami są aglomeracje miejskie. Wyjątkiem jest jedynie zmienna $X_{5,1}$, odpowiadająca usługom motoryzacyjnym /w wielkich miastach nasyconych bardzo mocno w samochody osobowe, baza usługową nie nadąża za rozwojem motoryzacji/.

Natomiast najniższymi wartościami zmiennych dysponują klasa II, Rys.5.9. Klasyfikacja województw Polski ze względu na poziom usług bytowych - wariant pierwszy.



zawierająca większość województw Polski wschodniej oraz województwo pilskie oraz klasa X zawierająca województwa nowosądeckie, płockie i siedleckie, przy czym wyjątkiem jest znów zmienna $X_{5,1}$.

Uwagę zwraca również wysoki poziom zmiennej $X_{5,8}$ /usługi stolarskie, kowalskie, meblarskie, itp./ w klasie IV.

Natomiast klasyfikacja druga składa się z 6 klas:

I - Warszawa,

II - Biała Podlaska, Białystok, Chełm, Elbląg, Kielce, Konin, Krosno, Łomża, Olsztyn, Opole, Ostrołęka, Piła, Piotrków Tryb., Przemyśl, Radom, Rzeszów, Sieradz, Suwałki, Tarnobrzeg, Tarnów, Toruń, Włocławek, Zamość,

III- Bielsko Biała, Częstochowa, Kalisz, Leszno,

IV - Bydgoszcz, Gdańsk, Kraków, Łódź, Poznań, Wrocław,

V - Ciechanów, Gorzów Wlkp., Jelenia Góra, Katowice, Koszalin, Legnica, Lublin, Skierniewice, Słupsk, Szczecin, Wałbrzych, Zielona Góra,

VI - Nowy Sącz, Płock, Siedlce.

Najniższy poziom usług obserwowany jest w klasie II, obejmującej większość województw Polski środkowej i wschodniej, a także kilka województw północnych, oraz w klasie VI, obejmującej jak poprzednio województwo nowosądeckie, płockie i siedleckie /por. rysunek 5.10 i tablica 5.20/.

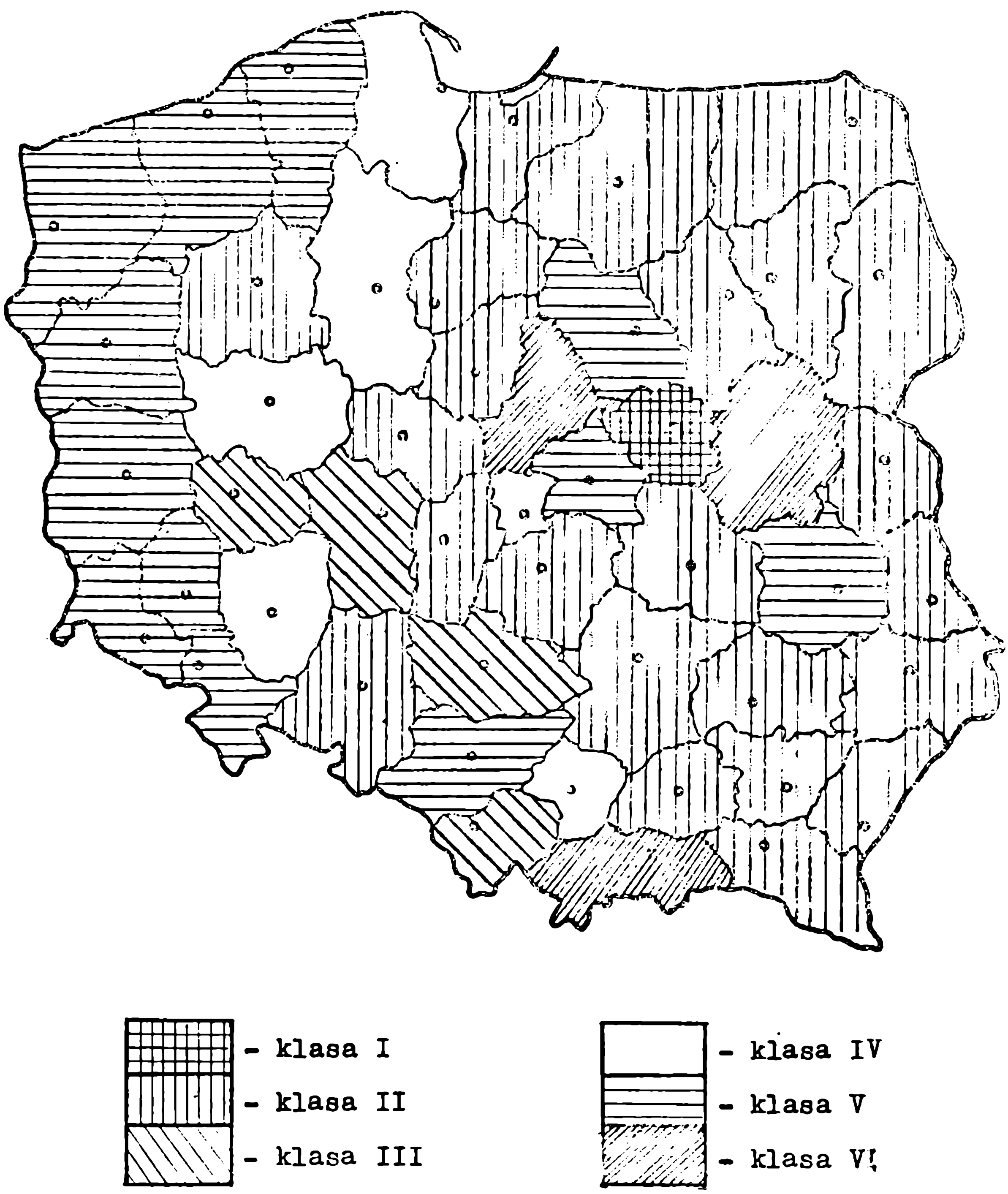
Tablica 5.20

Średnie arytmetyczne zmiennych charakteryzujących usługi bytowe dla poszczególnych klas - klasyfikacja druga.

Nr klasy	Zmienna					
	$X_{5.1}$	$X_{5.2}$	$X_{5.4}$	$X_{5.6}$	$X_{5.8}$	$X_{5.10}$
1	5071,4	114,8	627,7	169,1	324,6	408,6
2	4428,0	64,1	218,0	34,1	366,5	112,8
3	5263,5	64,5	257,2	48,9	568,1	129,6
4	4862,9	84,4	425,7	103,6	360,1	267,6
5	5643,6	84,9	311,0	54,0	250,2	170,8
6	6663,5	42,4	144,5	24,0	223,0	117,3

Źródło: obliczenia własne.

Rys.5.10. Klasyfikacja województw Polski ze względu na poziom usług bytowych - wariant drugi.



Natomiast najwyższe wartości zmiennych ma woj. warszawskie. Klasyfikacja ta nie różni się wiele od klasyfikacji pierwszej. Tablica 5.21 przedstawia uporządkowane wartości miary rozwoju dla tej grupy usług. Najwyższe pozycje zajmują województwa, których stolicami są wielkie aglomeracje, natomiast najniższe województwa Polski wschodniej.

Tablica 5.21

Wartość miary rozwoju usług bytowych

Lp.	Nazwa województwa	Wartość miary rozwoju	Lp.	Nazwa województwa	Wartość miary rozwoju
1	Warszawa	31,0121	26	Gorzów Wlkp.	17,5055
2	Łódź	26,9738	27	Białystok	17,3031
3	Kraków	24,8541	28	Sieradz	16,9512
4	Wrocław	23,2331	29	Radom	16,8308
5	Poznań	23,2085	30	Elbląg	16,8049
6	Gdańsk	22,7773	31	Nowy Sącz	16,7544
7	Szczecin	21,7908	32	Olsztyn	16,6928
8	Słupsk	21,7413	33	Kielce	16,4281
9	Opole	21,0231	34	Rzeszów	16,3516
10	Lublin	20,5514	35	Piotrków Tryb.	15,1028
11	Bielsko-Biała	20,5067	36	Przemyśl	14,3080
12	Zielona Góra	20,3072	37	Piła	14,1953
13	Bydgoszcz	20,2142	38	Siedlce	14,0958
14	Jelenia Góra	20,1619	39	Suwałki	14,0651
15	Częstochowa	20,0781	40	Tarnów	13,8688
16	Skierniewice	19,9316	41	Tarnobrzeg	13,2817
17	Toruń	19,4575	42	Łomża	12,8914
18	Koszalin	19,1584	43	Biała Podlaska	12,7793
19	Kalisz	18,7369	44	Płock	12,5755
20	Leszno	18,6152	45	Zamość	12,4870
21	Katowice	18,5803	46	Ostrołęka	12,4678
22	Legnica	18,3859	47	Konin	12,3944
23	Wałbrzych	18,3613	48	Krosno	11,7941
24	Ciechanów	18,0880	49	Chełm	11,4632
25	Włocławek	18,0338			

Źródło: obliczenia własne.

Zatem i w tej grupie usług uwidoczniło się zróżnicowanie terytorialne między województwami zachodnimi oraz województwami o dużej liczbie ludności a województwami wschodnimi.

5.4. Regionalizacja województw Polski ze względu na poziom wybranych grup usług konsumpcyjnych dla ludności.

Podobnie jak w poprzednim paragrafie przeprowadzona została klasyfikacja, tak w tym dokonana zostanie regionalizacja województw ze względu na poziom usług. Różni się ona tym, że zawiera obszary przestrzennie ciągłe. Jest to niewątpliwa zaleta z punktu widzenia potrzeb planistycznych.

Regionalizacji dokonano metodą średniej międzygrupowej należądą do grupy hierarchicznych metod grupowania, przy czym do metody tej został wprowadzony warunek przyległości zgodnie ze sposobem opisanym w 3.3. Również i tutaj narzucono warunek dotyczący ilości klas.

Rezultaty grupowania regionalnego przedstawiają tablice 5.22, 5.23, 5.24, 5.25, 5.26. Ponieważ w tym przypadku dla każdej z grup trudno jest wyodrębnić dwa etapy, po których obserwuje się stosunkowo duże spadki wewnętrznego podobieństwa /a więc straty informacji/, poprzestano na jednej regionalizacji /dla kążdej grupy/. Dla poszczególnych grup usług największe spadki wewnętrznego podobieństwa obserwuje się po następujących etapach:

dla usług kulturalno-rozrywkowych - po 41-szym etapie,

dla usług w zakresie ochrony zdrowia i opieki społecznej - po 41-szym etapie,

dla usług komunalno-mieszkaniowych - po 39-tym etapie,

dla usług w zakresie oświaty i wychowania - po 42-gim etapie,

dla usług bytowych - po 43-cim etapie.

Tablica 5.22

Rezultaty grupowania regionalnego dla usług kulturalno-rozrywkowych

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,4164	0,7704
35	1,5300	0,7906
36	1,5300	0,7912
37	1,5300	0,7996
38	1,5300	0,8010
39	1,5300	0,8313
40	1,5300	0,8320
41	1,6695	0,8531
42	2,3465	1,0862
43	2,4635	1,1095
44	2,4635	1,1520
45	2,5574	1,1733

Źródło: obliczenia własne.

Tablica 5.23

Rezultaty grupowania regionalnego dla usług w zakresie ochrony zdrowia i opieki społecznej

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,9046	0,8974
35	1,9046	0,9169
36	1,9046	0,9742
37	1,9135	0,9967
38	1,9135	1,0175
39	1,9146	1,0368
40	1,9146	1,0568
41	1,9146	1,0752
42	2,1508	1,0967
43	2,3430	1,1177
44	2,5292	1,1375
45	2,6331	1,1587

Źródło: obliczenia własne.

Tablica 5.24

Rezultaty grupowania regionalnego dla usług komunalno-mieszkaniowych

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,6768	0,8231
35	1,6768	0,8254
36	1,6768	0,8241
37	1,6768	0,8313
38	1,6768	0,8491
39	1,6956	0,8621
40	2,1053	0,9514
41	2,1053	0,9617
42	2,1053	0,9731
43	2,1053	0,9745
44	2,1053	0,9901
45	2,7008	1,2145

Źródło: obliczenia własne

Tablica 5.25

Rezultaty grupowania regionalnego dla usług w zakresie oświaty i wychowania

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,4372	0,8979
35	1,4372	0,9015
36	1,5900	0,9561
37	1,5900	0,9975
38	1,6368	1,0109
39	1,8935	1,0267
40	1,8935	1,0394
41	1,8935	1,0524
42	1,8935	1,0609
43	2,0559	1,2364
44	2,2820	1,2565
45	2,3247	1,2800

Źródło: obliczenia własne.

Tablica 5.26

Rezultaty grupowania regionalnego
dla usług bytowych

Nr etapu grupowania	Wartość Wp	Wartość W'p
34	1,8086	0,8375
35	1,8086	0,8648
36	1,8086	0,8743
37	1,8176	0,9063
38	1,8176	0,9156
39	2,0256	0,9320
40	2,0280	0,9493
41	2,0280	0,9691
42	2,0930	0,9891
43	2,4690	1,1780
44	2,4690	1,1708
45	2,4690	1,1820

Źródło: obliczenia własne.

Jak widać, otrzymane regionalizacje charakteryzować się będą dość dużym stopniem ogólności.

Otrzymane regionalizacje są następujące:

dla usług kulturalno-rozrywkowych - 8 regionów:

I - Warszawa,

II - Biała Podlaska, Białystok, Bielsko Biała, Chełm, Ciechanów, Częstochowa, Kalisz, Kielce, Konin, Krosno, Lublin, Łomża, Nowy Sącz, Ostrołęka, Piotrków Tryb., Płock, Przemyśl, Radom, Rzeszów, Siedlce, Sieradz, Skierniewice, Suwałki, Tarnobrzeg, Tarnów, Zamość,

III - Bydgoszcz, Elbląg, Gorzów Wlkp., Jelenia Góra, Koszalin, Legnica, Leszno, Olsztyn, Opole, Piła, Słupsk, Toruń, Wałbrzych, Włocławek, Wrocław, Zielona Góra,

IV - Gdańsk,

V - Katowice, Kraków,

VI - Łódź,

VII - Poznań,

VIII - Szczecin;

dla usług w zakresie ochrony zdrowia i opieki społecznej - 8 regionów:

I - Warszawa,

II - Biała Podlaska, Bielsko-Biała, Bydgoszcz, Chełm, Ciechanów, Częstochowa, Elbląg, Gdańsk, Gorzów Wlkp., Jelenia Góra, Kalisz, Katowice, Kielce, Koszalin, Krosno, Legnica, Leszno, Lublin, Łomża, Nowy Sącz, Olsztyn, Opole, Piła, Piotrków Tryb., Płock, Poznań, Przemyśl, Radom, Rzeszów, Siedlce, Sieradz, Skierniewice, Słupsk, Suwałki, Szczecin, Tarnobrzeg, Tarnów, Toruń, Wałbrzych, Włocławek, Zamość, Zielona Góra,

III - Białystok,

IV - Konin,

V - Kraków,

VI - Łódź,

VII - Ostrołęka,

VIII - Wrocław;

dla usług komunalno-mieszkaniowych - 10 regionów:

I - Warszawa,

II - Biała Podlaska, Białystok, Chełm, Ciechanów, Częstochowa, Kalisz, Konin, Leszno, Łomża, Ostrołęka, Piotrków Tryb., Płock, Siedlce, Sieradz, Skierniewice, Suwałki, Tarnobrzeg, Zamość,

III - Bielsko-Biała, Kraków,

IV - Bydgoszcz, Elbląg, Gdańsk, Gorzów Wlkp., Jelenia Góra, Koszalin, Legnica, Olsztyn, Opole, Piła, Słupsk, Szczecin, Toruń, Wałbrzych, Zielona Góra,

V - Katowice,

VI - Kielce, Lublin, Nowy Sącz, Przemyśl, Radom, Rzeszów, Tarnów,

VII - Krosno,

VIII - Łódź,

IX - Poznań,

X - Wrocław;

dla usług w zakresie oświaty i wychowania - 7 regionów:

I - Warszawa,

II - Biała Podlaska, Białystok, Bielsko-Biała, Chełm, Ciechanów, Częstochowa, Kielce, Konin, Kraków, Krosno, Lublin, Łomża, Nowy Sącz, Ostrołęka, Piotrków Tryb., Płock, Przemyśl, Radom, Rzeszów, Siedlce, Sieradz, Skierniewice, Suwałki, Tarnobrzeg, Tarnów, Włocławek, Zamość,

III - Bydgoszcz, Elbląg, Gdańsk, Gorzów Wlkp., Kalisz, Koszalin, Legnica, Leszno, Olsztyn, Opole, Piła, Poznań, Słupsk, Szczecin, Wałbrzych, Wrocław, Zielona Góra,

IV - Jelenia Góra,

V - Katowice,

VI - Łódź,

VII - Toruń;

dla usług bytowych - 7 regionów:

I - Warszawa,

II - Biała Podlaska, Białystok, Bielsko-Biała, Chełm, Ciechanów, Częstochowa, Elbląg, Kalisz, Kielce, Konin, Krosno, Leszno, Łomża, Nowy Sącz, Olsztyn, Ostrołęka, Piotrków Tryb., Przemyśl, Radom, Rzeszów, Siedlce, Sieradz, Skierniewice, Suwałki, Tarnobrzeg, Tarnów, Toruń, Włocławek, Zamość, III - Bydgoszcz, Gdańsk, Gorzów Wlkp., Jelenia Góra, Katowice, Koszalin, Kraków, Legnica, Opole, Poznań, Słupsk, Szczecin, Wałbrzych, Wrocław, Zielona Góra,

IV - Lublin,

V - Łódź,

VI - Piła,

VII - Płock.

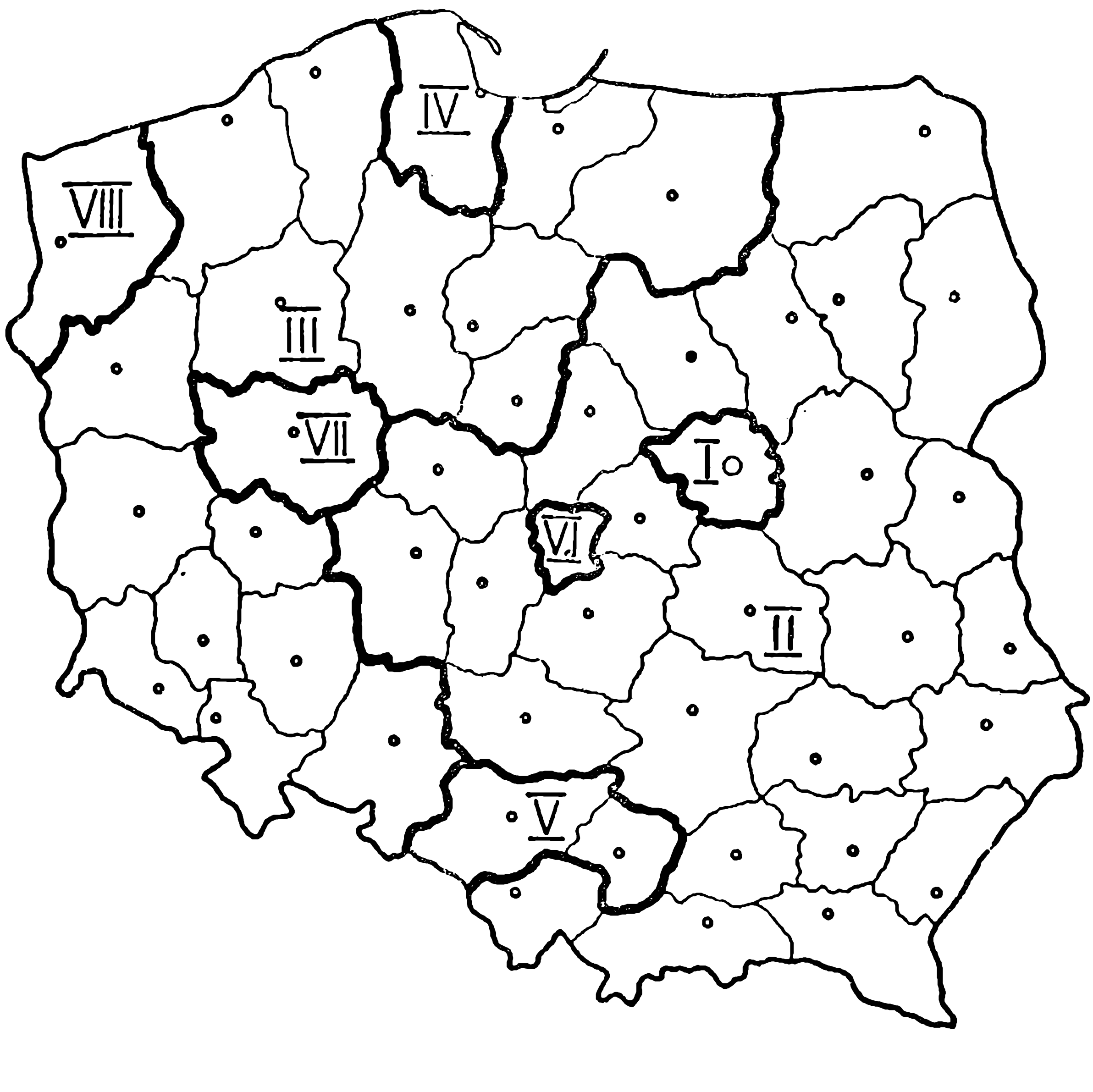
Rysunki 5.11, 5.12, 5.13, 5.14, 5.15 przedstawiają układy regionalne dla poszczególnych grup usług, natomiast tablice 5.27, 5.28, 5.29, 5.30, 5.31 przeciętne wartości zmiennych dla poszczególnych regionów.

Łatwo zauważyć, że dla trzech grup usług /kulturalno-rozrywkowe, oświaty i wychowania, bytowe/ Polska rozbija się na 2 podstawowe regiony - umownie można je nazwać zachodnim i wschodnim.

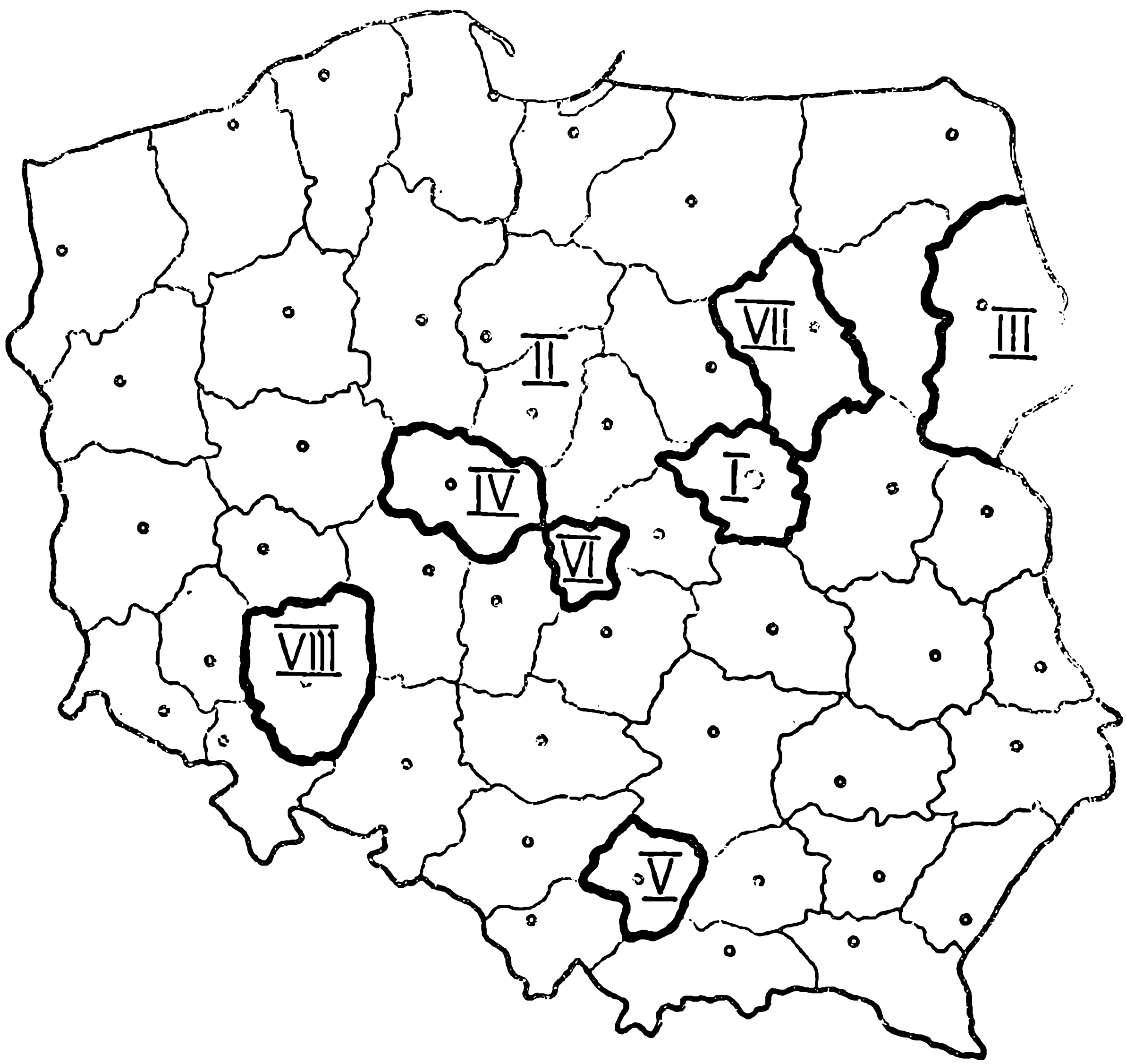
Dość charakterystyczne jest, iż w przypadku usług kulturalno-rozrywkowych oraz oświaty i wychowania oba te regiony w zasadzie pokrywają się /z nielicznymi wyjątkami, wynikającymi z istnienia regionów jednoelementowych/, przy czym region zachodni obejmuje województwa dolnośląskie, wielkopolskie, lubuskie, nadbałtyckie.

Natomiast w przypadku usług bytowych region ten jest nieco mniejszy / nie obejmuje np. województw elbląskiego, olsztyńskiego i wielkopolskich/. Jest również tak, że region zachodni posiada średnie wartości zmiennych wyższe niż region wschodni. Świadczy to o tym, iż dysproporcje rozwojowe mające swe podstawy na przełomie wieku, w lwiej części funkcjonują obecnie.

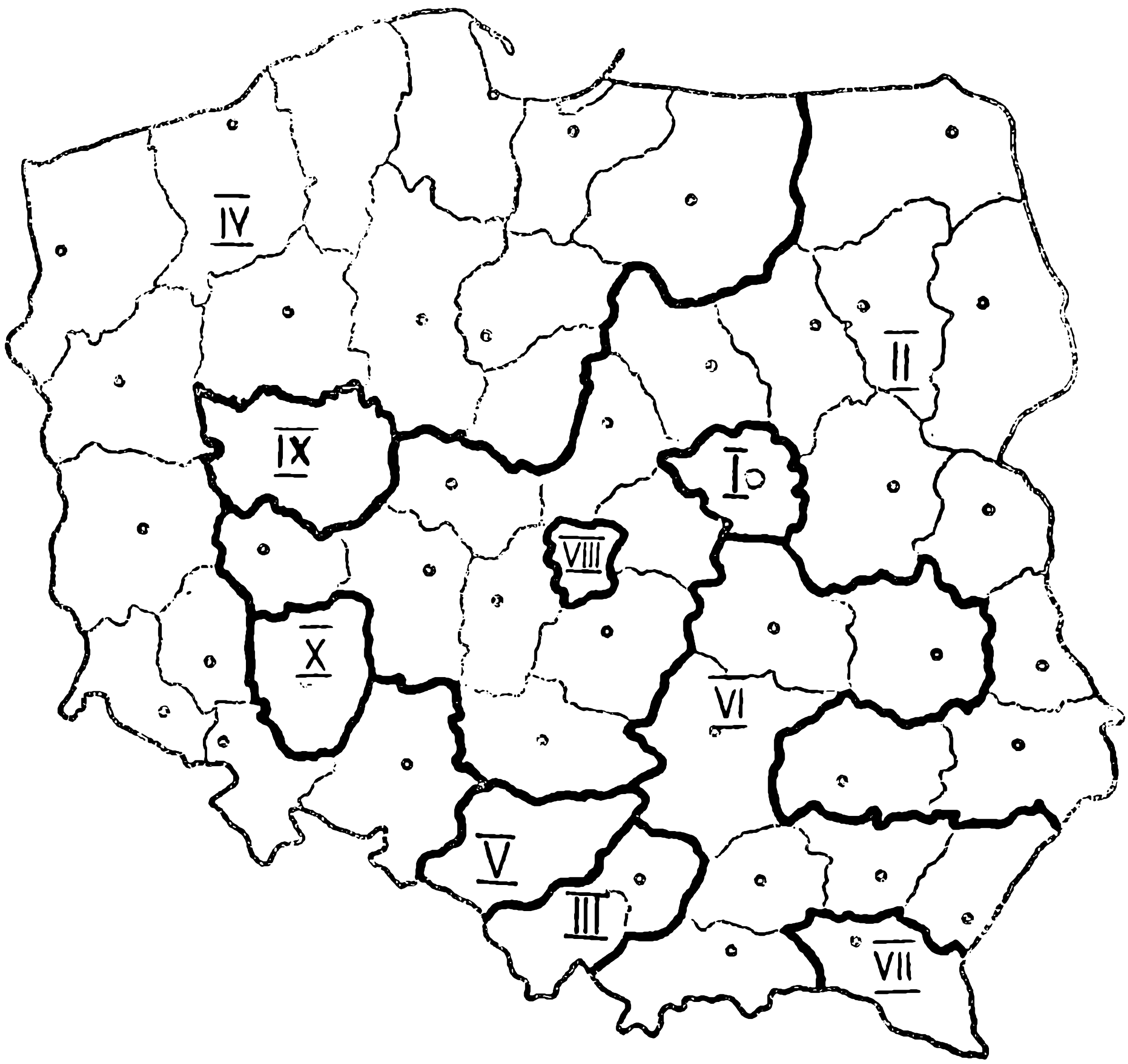
Zbliżona sytuacja jest również w przypadku usług komunalno-mieszkaniowych, przy czym region wschodni rozpada się na 2 mniejsze regiony; jeden obejmujący województwa północno-wschodnie i środkowe, drugi województwa małopolskie, dodajmy przy tym, Rys.5.11. REgionalizacja województw Polski ze względu na poziom usług kulturalno-rozrywkowych.



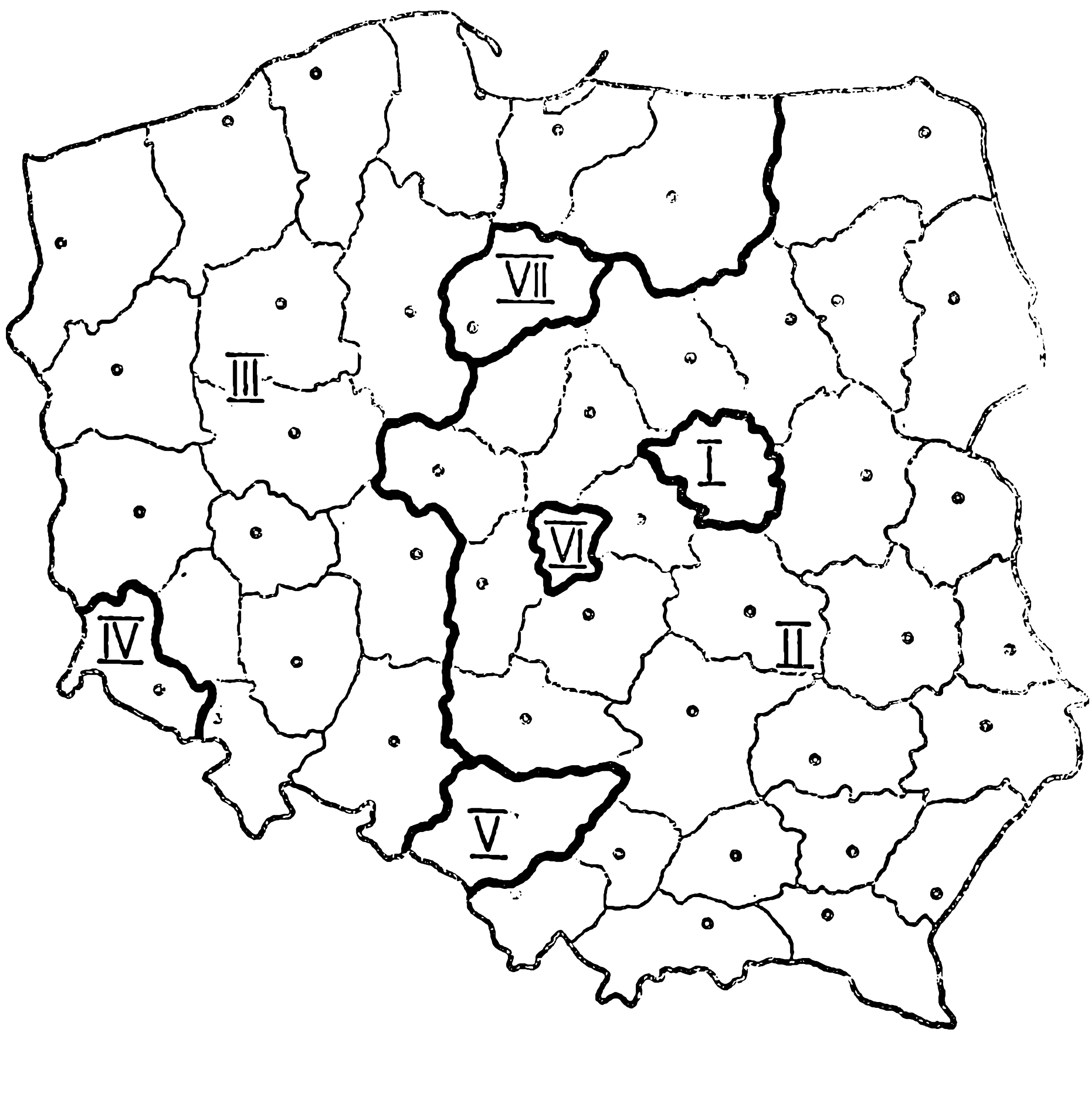
Rys.5.12. Regionalizacja województw Polski ze względu na poziom usług w zakresie ochrony zdrowia i opieki społecznej.



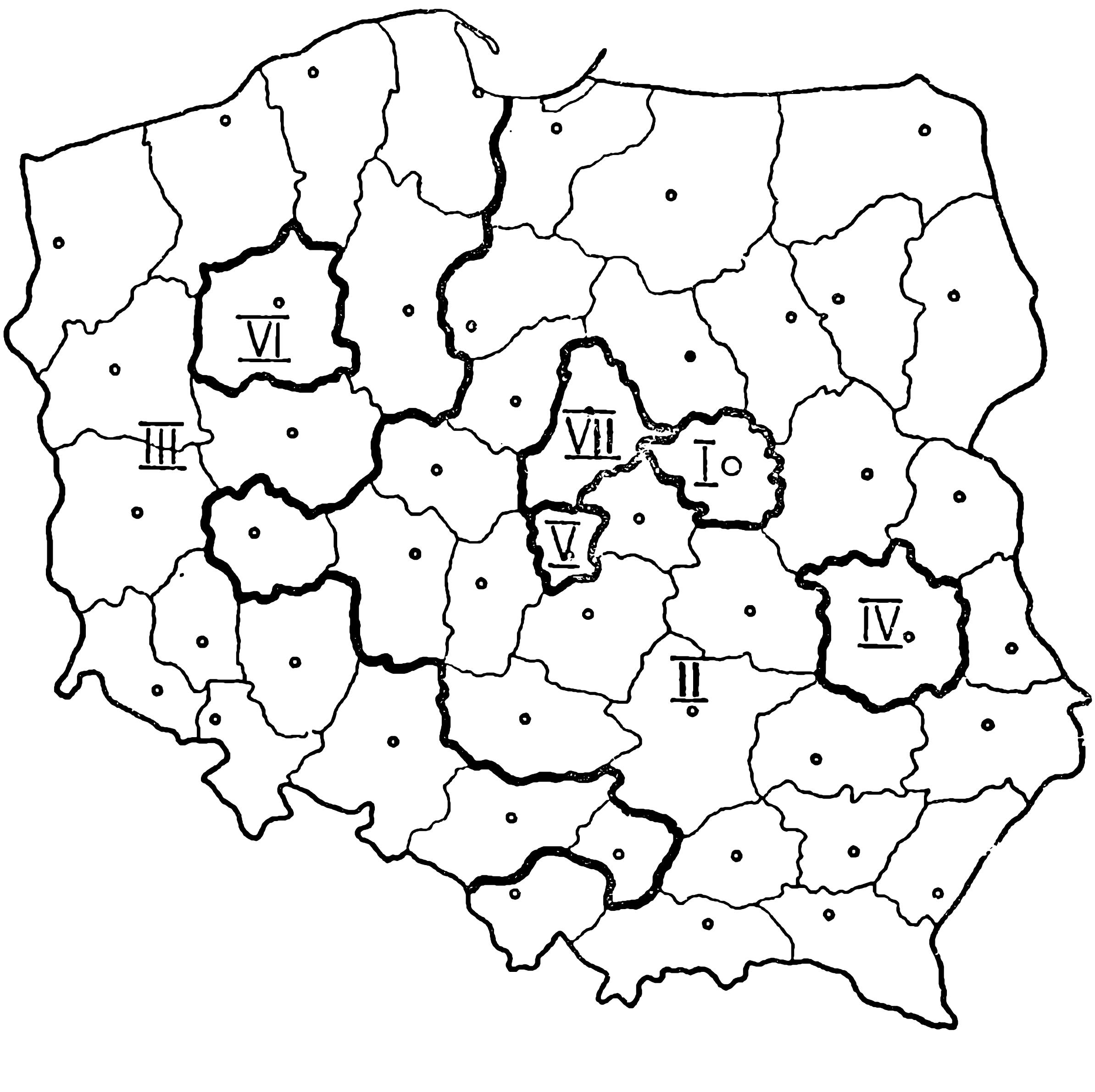
Rys.5.13. Regionalizacja województw Polski ze względu na poziom usług komunalno-mieszkaniowych.



Rys.5.14. Regionalizacja województw Polski ze względu na poziom usług w zakresie oświaty i wychowania.



Rys.5.15. Regionalizacja województw Polski ze względu na poziom usług bytowych.



Tablica 5.27

Średnie arytmetyczne zmiennych charakteryzujących usługi kulturalno-rozrywkowe dla poszczególnych regionów

Nr. regionu	Zmienna							
	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$
1	2043	302	271	15,3	931	462	23,8	163
2	2670	205	180	11,6	119	123	18,7	72
3	2894	233	226	17,7	247	216	19,3	98
4	2107	268	239	12,0	334	380	22,0	126
5	1919	207	234	13,2	342	263	22,5	113
6	2361	322	288	13,3	511	716	16,9	127
7	2337	270	245	10,1	232	553	21,8	125
8	3411	270	245	24,5	330	424	23,0	123

Źródło: obliczenia własne.

Tablica 5.28

Średnie arytmetyczne zmiennych charakteryzujących usługi w zakresie ochrony zdrowia i opieki społecznej dla poszczególnych regionów

Nr regionu	Zmienna										
	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,6}$	$X_{2,7}$	$X_{2,9}$	$X_{2,10}$	$X_{2,11}$	$X_{2,12}$	$X_{2,13}$
1	44,1	11,3	64,0	78,8	20,7	10,9	4,8	8,9	3,6	5,8	14
2	13,7	4,0	41,9	50,3	21,4	3,5	4,5	11,7	1,0	4,6	32
3	27,6	4,7	55,3	65,0	19,9	3,4	7,3	14,9	0,8	3,4	12
4	8,9	3,3	29,0	36,7	22,8	2,2	4,1	7,8	0,8	6,5	62
5	32,2	6,5	59,2	60,3	17,3	10,6	6,9	8,2	2,1	4,9	27
6	34,2	9,2	64,5	64,8	20,5	9,2	6,6	14,5	5,2	4,1	7
7	9,7	2,9	28,3	33,1	24,7	2,7	3,7	13,8	0,7	4,8	28
8	28,8	7,1	58,0	76,8	19,9	6,2	5,4	10,8	1,8	4,6	10

Źródło: obliczenia własne.

Tablica 5.29

Średnie arytmetyczne zmiennych charakteryzujących usługi komunalno-mieszkaniowe dla poszczególnych regionów

Nr regionu	Zmienna						
	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,6}$	$X_{3,7}$	$X_{3,8}$	$X_{3,9}$
1	77,4	43,2	11,8	108,2	62,2	453,1	1234
2	61,6	54,0	9,4	18,2	13,7	181,7	690
3	78,0	64,2	10,0	31,0	31,9	373,9	1080
4	81,9	63,7	12,9	27,7	34,1	272,9	783
5	81,1	66,4	12,9	28,6	72,2	395,1	646
6	71,6	55,8	10,2	22,9	19,6	176,3	1259
7	58,8	62,2	24,3	16,2	13,3	155,7	1069
8	86,6	49,8	8,6	51,2	63,8	443,9	1110
9	79,2	49,2	18,1	39,4	34,9	366,7	1162
10	90,4	75,0	14,9	55,9	46,1	323,8	1062

Źródło: obliczenia własne.

Tablica 5.30

Średnie arytmetyczne zmiennych charakteryzujących usługi w zakresie oświaty i wychowania dla poszczególnych regionów

Nr regionu	Zmienna								
	$X_{4,1}$	$X_{4,2}$	$X_{4,3}$	$X_{4,4}$	$X_{4,5}$	$X_{4,7}$	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$
1	577,1	17,9	129,7	18,1	29,8	78,6	24	31	32
2	460,5	17,8	129,5	16,5	21,8	63,9	22	28	32
3	503,5	18,3	122,2	20,5	28,2	61,0	22	29	31
4	523,6	18,1	114,1	24,4	28,3	60,4	22	28	27
5	464,3	18,4	114,2	19,3	23,9	67,4	26	30	34
6	595,2	17,9	120,6	11,6	28,7	71,0	23	32	33
7	451,0	22,5	128,8	22,9	19,4	63,3	22	28	34

Źródło: obliczenia własne.

Tablica 5.31

Średnie arytmetyczne zmiennych charakteryzujących usługi bytowe dla poszczególnych regionów

Nr regionu	Zmienna					
	$X_{5,1}$	$X_{5,2}$	$X_{5,4}$	$X_{5,6}$	$X_{5,8}$	$X_{5,10}$
1	5071,4	114,8	627,7	169,1	324,6	408,6
2	4828,4	63,1	220,8	33,0	332,3	112,9
3	5178,9	85,9	326,9	74,9	295,5	208,9
4	5408,9	77,0	543,5	69,6	188,2	167,5
5	5033,3	101,8	647,3	117,8	307,0	295,7
6	4148,5	55,8	140,3	36,6	326,2	118,5
7	6680,2	31,9	126,3	17,5	96,9	130,6

Źródło: obliczenia własne.

że region małopolski charakteryzuje się wyższymi średnimi wartościami zmiennych.

Dla usług w zakresie ochrony zdrowia i opieki społecznej brak jest podziału na region wschodni i zachodni. Występuje tutaj siedem regionów jednoelementowych, przy czym pięć z nich zawiera województwa o wysokim poziomie usług: warszawskie, łódzkie, krakowskie, wrocławskie i białostockie, natomiast dwa województwa o bardzo niskim poziomie usług: ostrołęckie i konińskie. Pozostałe województwa znalazły się w jednym regionie. Sugerowałoby to, iż w dziedzinie usług w zakresie ochrony zdrowia i opieki społecznej występują mniejsze dysproporcje między wschodem i zachodem Polski.

Charakterystyczną cechą dla wszystkich grup usług jest występowanie jedno- lub kilkuelementowych regionów zawierających województwa, których stolicami jest osiem największych miast Polski. Regiony te charakteryzują się oczywiście wyższym poziomem usług.

5.5. Klasyfikacja rozmyta województw Polski ze względu na poziom usług konsumpcyjnych dla ludności.

Przy wyznaczaniu rozmytej klasyfikacji bezwzględnej przyjęto najprostszy sposób, tzn. za pomocą wzoru /4.1/, wyrażającego podobieństwo obiektu do klasy /podzbioru rozmytego/ przez odległość obiektu od środka ciężkości /centroidu/ danej klasy. Konieczne jest zatem uprzednie skonstruowanie klasyfikacji w zwykłym sensie. Wykorzystane tutaj zostaną klasyfikacje o większym stopniu ogólności, otrzymane w 5.3. Dla każdej grupy usług klasyfikacje te liczą po 6 klas. Mamy zatem w każdym przypadku do czynienia z sześcioma podzbiorami rozmytymi określonymi na zbiorze obiektów. Po wyznaczeniu odległości obiektu od środka ciężkości odległość ta jest przekształcana na podobieństwo wg wzorów:

$$s^{*}/P_i, G_k/ = \frac{1}{1 + d/P_i, G_k/}$$

$$f_{S_k}/P_i/ = s/P_I, G_k/ = 1.25\ s^{*}/P_i, G_k/ - 0.25$$

Stosowanie drugiego z powyższych wzorów wynika z tego, iż wartości funkcji odległości liczone wg formuły euklidesowej zawierały się /w badaniach prowadzonych w pracy/ w przedziale $\langle 0; 4 \rangle$, a więc stosowanie tylko pierwszego z powyższych wzorów spowoduje, iż podobieństwo zawierać się będzie w przedziale $\langle \frac{1}{4}\ \ 1 \rangle$. Wzór drugi stosuje się w celu przekształcenia przedziału zmienności wartości funkcji podobieństwa z $\langle \frac{1}{4}\ \ 1 \rangle$ na $\langle 0; 1 \rangle$.

Za pomocą opisanego powyżej sposobu wyznaczono klasyfikację rozmytą bezwzględną. Rezultaty obliczeń przedstawione są w tablicach 5.32, 5.33, 5.34, 5.35, 5.36.

Tablica 5.32

Wartość funkcji przynależności do poszczególnych klas dla klasyfikacji rozmytej bezwzględnej - usługi kulturalno-rozrywkowe.

Nazwa województwa	Wartość funkcji przynależności do klasy					
	I	II	III	IV	V	VI
Warszawa	1,0000	0,0626	0,1937	0,1176	0,1993	0,1479
Biała Podlaska	0,0471	0,6981	0,2412	0,3019	0,0933	0,1281
Białystok	0,0938	0,5039	0,3198	0,3898	0,1280	0,2041
Bielsko Biała	0,0950	0,5610	0,3424	0,4304	0,1289	0,1763
Bydgoszcz	0,1542	0,2503	0,5050	0,4847	0,2539	0,3068
Chełm	0,0749	0,6663	0,3254	0,3811	0,1106	0,1668
Ciechanów	0,0443	0,6244	0,2257	0,2882	0,0803	0,1268
Częstochowa	0,1036	0,4579	0,4206	0,3404	0,1339	0,1676
Elbląg	0,1341	0,3122	0,3889	0,5620	0,1841	0,3970
Gdańsk	0,2343	0,2219	0,6773	0,2860	0,2478	0,2276
Gorzów Wlkp.	0,1005	0,3648	0,3382	0,6714	0,1461	0,3183
Jelenia Góra	0,1044	0,2848	0,2832	0,4888	0,1286	0,2569
Kalisz	0,0883	0,6085	0,3669	0,3741	0,1311	0,1698
Katowice	0,1637	0,2296	0,5150	0,2888	0,1714	0,2087
Kielce	0,0798	0,5780	0,2921	0,4221	0,1374	0,1717
Konin	0,0535	0,6535	0,2509	0,3116	0,1046	0,1273
Koszalin	0,1384	0,2591	0,3602	0,4948	0,2092	0,3561
Kraków	0,2367	0,2202	0,4582	0,2359	0,1889	0,1746
Krosno	0,0313	0,4713	0,1725	0,2697	0,0641	0,1111
Legnica	0,1184	0,3686	0,4230	0,3963	0,1307	0,2075
Leszno	0,0888	0,4049	0,3470	0,3776	0,1311	0,2243
Lublin	0,1003	0,4712	0,3915	0,3852	0,1627	0,1701
Łomża	0,0324	0,5000	0,1863	0,2628	0,0614	0,1304
Łódź	0,1993	0,1012	0,2214	0,1789	1,0000	0,1616
Nowy Sącz	0,0330	0,5236	0,1880	0,2402	0,0560	0,1102
Olsztyn	0,1195	0,3843	0,3414	0,6295	0,1713	0,2809
Opole	0,0908	0,3301	0,2762	0,6224	0,1588	0,2638
Ostrołęka	0,0126	0,4100	0,1396	0,2052	0,0493	0,0734
Piła	0,0722	0,3771	0,2651	0,5583	0,1367	0,2354
Piotrków Trybunalski	0,0670	0,4594	0,2862	0,2558	0,0829	0,1234
Płock	0,0836	0,6038	0,3486	0,3577	0,1261	0,1634
Poznań	0,1934	0,2033	0,4958	0,2611	0,2779	0,2289
Przemyśl	0,0494	0,5046	0,1293	0,3763	0,0981	0,1617
Radom	0,0594	0,8567	0,2634	0,3305	0,0958	0,1402
Rzeszów	0,0683	0,4913	0,2403	0,3431	0,0940	0,1773
Siedlce	0,0231	0,5011	0,1699	0,2366	0,0627	0,0893
Sieradz	0,0561	0,6817	0,2653	0,3162	0,0947	0,1376
Skierniewice	0,0679	0,6196	0,3103	0,3320	0,1066	0,1509
Słupsk	0,1054	0,3713	0,3680	0,4828	0,2026	0,2121
Suwałki	0,0623	0,3679	0,2397	0,3601	0,0767	0,2411
Szczecin	0,1479	0,1535	0,2578	0,3140	0,1616	1,0000
Tarnobrzeg	0,0381	0,5742	0,2019	0,2673	0,0739	0,1041
Tarnów	0,0513	0,6145	0,2292	0,2892	0,0740	0,1365
Toruń	0,1321	0,2915	0,3302	0,4903	0,1796	0,2838
Wałbrzych	0,1223	0,2205	0,3063	0,4711	0,1751	0,3726
Włocławek	0,1293	0,3326	0,2991	0,4396	0,1430	0,2443
Wrocław	0,1833	0,2544	0,6368	0,3640	0,2001	0,3293
Zamość	0,0451	0,6373	0,2285	0,2994	0,0924	0,1267
Zielona Góra	0,0890	0,3207	0,2928	0,5786	0,1799	0,2440

Źródło: obliczenia własne.

Tablica 5.33

Wartości funkcji przynależności do poszczególnych klas dla klasyfikacji rozmytej bezwzględnej - usługi w zakresie ochrony zdrowia i opieki społecznej

Nazwa województwa	Wartość funkcji przynależności do klasy					
	I	II	III	IV	V	VI
Warszawa	1,0000	0,0830	0,0747	0,2256	0,0522	0,2384
Biała Podlaska	0,0493	0,4236	0,2329	0,2020	0,2991	0,0661
Białystok	0,0951	0,2410	0,5236	0,2747	0,0941	0,1597
Bielsko Biała	0,0943	0,5317	0,2477	0,3134	0,3153	0,0923
Bydgoszcz	0,0956	0,6246	0,2816	0,3552	0,2724	0,1204
Chełm	0,0842	0,4653	0,2484	0,2992	0,2783	0,0867
Ciechanów	0,0495	0,4945	0,2350	0,2050	0,2998	0,0607
Częstochowa	0,0757	0,5688	0,2869	0,2933	0,3387	0,0883
Elbląg	0,0733	0,5160	0,4162	0,2726	0,2072	0,1054
Gdańsk	0,2153	0,3080	0,2127	0,5741	0,2141	0,1760
Gorzów Wlkp.	0,0706	0,2962	0,5941	0,2511	0,1286	0,1260
Jelenia Góra	0,1064	0,3982	0,3255	0,3361	0,1773	0,1186
Kalisz	0,0786	0,5101	0,2477	0,2815	0,4111	0,0819
Katowice	0,1614	0,2936	0,2119	0,3835	0,1885	0,1753
Kielce	0,0930	0,6011	0,4061	0,3370	0,2481	0,1248
Konin	0,0522	0,2983	0,1290	0,1797	1,0000	0,0376
Koszalin	0,0726	0,3642	0,3895	0,2437	0,1401	0,1105
Kraków	0,2515	0,1701	0,1687	0,3741	0,1046	0,2206
Krosno	0,0743	0,6155	0,3632	0,2832	0,2339	0,1011
Legnica	0,1027	0,4975	0,3612	0,3546	0,2092	0,1168
Leszno	0,0679	0,4773	0,2398	0,2386	0,2698	0,0711
Lublin	0,1712	0,2898	0,2139	0,4459	0,1836	0,1453
Łomża	0,0453	0,4396	0,3628	0,2195	0,2217	0,0756
Łódź	0,2384	0,0990	0,1305	0,2101	0,0376	1,0000
Nowy Sącz	0,1043	0,5218	0,3969	0,3709	0,2223	0,1278
Olsztyn	0,6969	0,4252	0,4279	0,3282	0,1715	0,1238
Opole	0,1010	0,4498	0,3050	0,3505	0,2109	0,1110
Ostrołęka	0,0290	0,3003	0,1670	0,1452	0,2804	0,0401
Piła	0,0788	0,6363	0,3013	0,2921	0,3224	0,0886
Piotrków Tryb.	0,0882	0,6773	0,2996	0,3262	0,3224	0,0950
Płock	0,0816	0,6302	0,2657	0,2926	0,3329	0,0840
Poznań	0,1953	0,3003	0,2920	0,5149	0,1473	0,1894
Przemyśl	0,0499	0,3652	0,2669	0,2039	0,2457	0,0747
Radom	0,0646	0,5566	0,2604	0,2471	0,2943	0,0776
Rzeszów	0,1012	0,5937	0,3214	0,3711	0,2857	0,1198
Siedlce	0,0471	0,4537	0,2030	0,1933	0,3383	0,0525
Sieradz	0,0783	0,5899	0,3577	0,2898	0,2675	0,1015
Skierniewice	0,0943	0,5722	0,2439	0,3240	0,4020	0,0858
Słupsk	0,0675	0,3413	0,6432	0,2495	0,1250	0,1167
Suwałki	0,0541	0,3959	0,5518	0,2306	0,1482	0,1012
Szczecin	0,1612	0,2983	0,2856	0,4159	0,1847	0,1615
Tarnobrzeg	0,0720	0,5822	0,2603	0,2760	0,3698	0,0804
Tarnów	0,0719	0,6554	0,2950	0,2732	0,3173	0,0909
Toruń	0,0893	0,7271	0,3167	0,3247	0,2837	0,0986
Wałbrzych	0,1149	0,3307	0,2653	0,3296	0,1683	0,1135
Włocławek	0,0712	0,5867	0,3025	0,2673	0,2814	0,0883
Wrocław	0,2437	0,2469	0,2680	0,5662	0,1239	0,2362
Zamość	0,0550	0,5141	0,2219	0,2216	0,3380	0,0641
Zielona Góra	0,0893	0,4424	0,3318	0,3120	0,2057	0,1009

Źródło: obliczenia własne.

Tablica 5.34

Wartości funkcji przynależności do poszczególnych klas dla klasyfikacji rozmytej bezwzględnej - usługi komunalno-mieszkaniowe

Nazwa województwa	Wartości funkcji przynależności do klasy					
	I	II	III	IV	V	VI
Warszawa	1,0000	0,0783	0,1851	0,0923	0,0754	0,0385
Biała Podlaska	0,0344	0,3528	0,1363	0,1500	0,2164	0,1663
Białystok	0,0887	0,4279	0,2295	0,1921	0,3370	0,1693
Bielsko Biała	0,1086	0,3457	0,4250	0,3517	0,3086	0,1759
Bydgoszcz	0,1220	0,4644	0,4798	0,3940	0,3808	0,1793
Chełm	0,0584	0,6099	0,2182	0,2592	0,3146	0,1960
Ciechanów	0,0549	0,6399	0,2215	0,2633	0,3619	0,2389
Częstochowa	0,0952	0,5100	0,2908	0,2480	0,3743	0,1674
Elbląg	0,0753	0,3453	0,3471	0,6221	0,2630	0,1646
Gdańsk	0,1843	0,2870	0,6245	0,3527	0,2605	0,1216
Gorzów Wlkp.	0,0908	0,3985	0,3413	0,4562	0,2703	0,2831
Jelenia Góra	0,1181	0,4400	0,4461	0,4460	0,3207	0,2316
Kalisz	0,0855	0,7696	0,3108	0,3278	0,4095	0,2582
Katowice	0,1210	0,2080	0,3571	0,3906	0,1545	0,1202
Kielce	0,0823	0,3817	0,2771	0,2373	0,7997	0,2160
Konin	0,0408	0,4168	0,2003	0,3169	0,2345	0,2473
Koszalin	0,0921	0,2772	0,3377	0,5029	0,2323	0,2516
Kraków	0,1535	0,2965	0,7080	0,4281	0,3028	0,1538
Krosno	0,0385	0,2367	0,1571	0,1898	0,2335	1,0000
Legnica	0,1128	0,3118	0,4402	0,5840	0,2570	0,1702
Leszno	0,0721	0,5488	0,2458	0,2951	0,2649	0,2568
Lublin	0,1160	0,4947	0,4211	0,4014	0,4421	0,2352
Łomża	0,0507	0,3850	0,1582	0,1520	0,2783	0,1627
Łódź	0,2482	0,1777	0,4271	0,2311	0,1651	0,0813
Nowy Sącz	0,0712	0,3432	0,2449	0,2159	0,6006	0,2383
Olsztyn	0,1018	0,4647	0,4076	0,5236	0,3024	0,1733
Opole	0,0651	0,3649	0,2912	0,5818	0,2327	0,2179
Ostrołęka	0,0458	0,4668	0,1970	0,2226	0,4028	0,2217
Piła	0,0601	0,4049	0,2467	0,3739	0,2350	0,2909
Piotrków Tryb.	0,0710	0,5720	0,2491	0,2680	0,3040	0,1563
Płock	0,0759	0,6709	0,2977	0,3632	0,3367	0,1955
Poznań	0,1674	0,2629	0,4135	0,2806	0,2516	0,2247
Przemyśl	0,0740	0,5047	0,2493	0,2431	0,4901	0,2608
Radom	0,0917	0,4848	0,2946	0,2641	0,4226	0,1595
Rzeszów	0,0634	0,2892	0,2527	0,2502	0,5625	0,2571
Siedlce	0,0500	0,4928	0,1793	0,1927	0,2859	0,1830
Sieradz	0,0522	0,5906	0,2242	0,3034	0,2810	0,2452
Skierniewice	0,0711	0,7105	0,2576	0,2908	0,3199	0,2110
Słupsk	0,0685	0,3003	0,3352	0,5098	0,2757	0,1609
Suwałki	0,0584	0,5324	0,2610	0,3577	0,3196	0,1834
Szczecin	0,1941	0,2747	0,6727	0,4026	0,2520	0,1417
Tarnobrzeg	0,0536	0,5640	0,2010	0,2376	0,2940	0,1878
Tarnów	0,0742	0,4266	0,2630	0,2438	0,5992	0,1889
Toruń	0,1196	0,4300	0,4091	0,4143	0,3046	0,2577
Wałbrzych	0,0821	0,2944	0,3509	0,7474	0,2124	0,1869
Włocławek	0,0679	0,4659	0,2624	0,3142	0,3620	0,4183
Wrocław	0,1600	0,2026	0,4374	0,3834	0,2036	0,1530
Zamość	0,0343	0,4338	0,1751	0,2184	0,3361	0,2210
Zielona Góra	0,0984	0,5103	0,4120	0,5163	0,3171	0,1933

Źródło: obliczenia własne.

Tablica 5.35

Wartości funkcji przynależności do poszczególnych klas dla klasyfikacji rozmytej bezwzględnej - usługi w zakresie oświaty i wychowania

Nazwa województwa	Wartości funkcji przynależności do klasy					
	I	II	III	IV	V	VI
Warszawa	0,5944	0,1796	0,2709	0,1042	0,2664	0,1488
Biała Podlaska	0,1198	0,3755	0,2402	0,0974	0,1309	0,1345
Białystok	0,2318	0,3709	0,2823	0,0887	0,1519	0,1403
Bielsko-Biała	0,2942	0,3968	0,3974	0,1470	0,3242	0,1554
Bydgoszcz	0,2538	0,3713	0,4199	0,1196	0,2948	0,2593
Chełm	0,1782	0,5516	0,3131	0,1221	0,2212	0,2106
Ciechanów	0,1541	0,5685	0,3578	0,1674	0,1951	0,2154
Częstochowa	0,1416	0,4226	0,3069	0,1445	0,1915	0,1807
Elbląg	0,2484	0,3694	0,6203	0,2030	0,2558	0,2315
Gdańsk	0,2873	0,3280	0,3806	0,1012	0,3266	0,2814
Gorzów Wlkp.	0,3200	0,3310	0,6238	0,1707	0,3080	0,2146
Jelenia Góra	0,1007	0,1304	0,1840	1,0000	0,0996	0,0880
Kalisz	0,2361	0,3602	0,5079	0,2301	0,2775	0,2699
Katowice	0,2680	0,2145	0,2973	0,0996	1,0000	0,2022
Kielce	0,1670	0,4943	0,2643	0,1081	0,1413	0,1533
Konin	0,1809	0,4119	0,3224	0,1149	0,2673	0,1962
Koszalin	0,2567	0,4118	0,6629	0,1677	0,2904	0,2415
Kraków	0,3258	0,3893	0,3875	0,1533	0,2452	0,1636
Krosno	0,1562	0,5430	0,2725	0,1167	0,1714	0,1667
Legnica	0,2480	0,3093	0,4742	0,1373	0,2710	0,3794
Leszno	0,1946	0,2621	0,4709	0,1654	0,2811	0,3047
Lublin	0,2828	0,3944	0,3052	0,1150	0,2092	0,1691
Łomża	0,1294	0,5237	0,2994	0,1206	0,1589	0,1912
Łódź	0,5944	0,1926	0,2646	0,0890	0,2380	0,1292
Nowy Sącz	0,2148	0,6301	0,3987	0,1306	0,2826	0,2191
Olsztyn	0,1915	0,2592	0,3523	0,1509	0,1664	0,1260
Opole	0,2859	0,2501	0,3900	0,1978	0,2245	0,1341
Ostrołęka	0,1420	0,4047	0,3214	0,1248	0,1861	0,2030
Piła	0,2234	0,3596	0,5283	0,2349	0,2696	0,2607
Piotrków Tryb.	0,1736	0,4365	0,3214	0,1181	0,2072	0,1620
Płock	0,2145	0,4683	0,3752	0,1256	0,2188	0,2484
Poznań	0,3164	0,2842	0,4828	0,1499	0,2843	0,2232
Przemyśl	0,1581	0,4507	0,2477	0,0900	0,2034	0,1553
Radom	0,1571	0,4761	0,3190	0,1269	0,1481	0,1706
Rzeszów	0,2467	0,4495	0,3600	0,1001	0,2457	0,2608
Siedlce	0,1468	0,4767	0,2935	0,1323	0,1522	0,1622
Sieradz	0,1332	0,4774	0,2369	0,0880	0,1383	0,1383
Skierniewice	0,1675	0,4284	0,2941	0,0975	0,1578	0,1734
Słupsk	0,2001	0,3291	0,4650	0,1479	0,2522	0,1931
Suwałki	0,1543	0,5580	0,3253	0,1461	0,1686	0,1975
Szczecin	0,2611	0,3582	0,6562	0,1727	0,2322	0,2574
Tarnobrzeg	0,1669	0,6041	0,2922	0,1012	0,1841	0,2019
Tarnów	0,2346	0,5075	0,4313	0,1263	0,3213	0,2536
Toruń	0,1450	0,2031	0,2523	0,0880	0,2022	1,0000
Wałbrzych	0,1578	0,2504	0,3774	0,2512	0,1796	0,1732
Włocławek	0,1283	0,4450	0,2835	0,1083	0,1765	0,1683
Wrocław	0,4235	0,3197	0,5221	0,1437	0,2757	0,2090
Zamość	0,1200	0,4479	0,2269	0,0856	0,1322	0,1416
Zielona Góra	0,2844	0,2649	0,5409	0,1556	0,2534	0,2101

Źródło: obliczenia własne.

Tablica 5.36

Wartości funkcji przynależności do poszczególnych klas dla klasyfikacji rozmytej bezwzględnej - usługi bytowe

Nazwa województwa	Wartości funkcji przynależności do klasy					
	I	II	III	IV	V	VI
Warszawa	1,0000	0,0533	0,0625	0,2378	0,1167	0,0235
Biała Podlaska	0,0327	0,5657	0,2377	0,1725	0,3310	0,3621
Białystok	0,0792	0,6212	0,4188	0,3021	0,4508	0,2850
Bielsko-Biała	0,0711	0,3121	0,6274	0,2515	0,2925	0,2093
Bydgoszcz	0,1457	0,3481	0,3107	0,5051	0,4396	0,2315
Chełm	0,0291	0,4020	0,1863	0,1534	0,2367	0,2002
Ciechanów	0,0555	0,3685	0,2983	0,2105	0,4361	0,4136
Częstochowa	0,0554	0,2877	0,6628	0,2178	0,2445	0,2199
Elbląg	0,0684	0,6602	0,3720	0,2665	0,5017	0,3454
Gdańsk	0,2016	0,2652	0,2802	0,7374	0,4105	0,1780
Gorzów Wlkp.	0,0935	0,4094	0,2370	0,2832	0,5357	0,2352
Jelenia Góra	0,1324	0,3532	0,2613	0,3850	0,7538	0,2331
Kalisz	0,0557	0,3792	0,7483	0,2330	0,2993	0,2826
Katowice	0,1127	0,3946	0,2881	0,3520	0,4321	0,2038
Kielce	0,0668	0,5590	0,2954	0,2497	0,5495	0,3839
Konin	0,0264	0,5568	0,2832	0,1669	0,2591	0,2988
Koszalin	0,1254	0,3320	0,2322	0,3618	0,5188	0,2701
Kraków	0,2457	0,1871	0,2020	0,5366	0,3263	0,1530
Krosno	0,0186	0,4757	0,2686	0,1501	0,2266	0,2940
Legnica	0,1271	0,3604	0,2163	0,3575	0,5237	0,2111
Leszno	0,0623	0,3831	0,6956	0,2472	0,2846	0,2209
Lublin	0,1408	0,2654	0,2003	0,3484	0,4175	0,1950
Łomża	0,0278	0,6364	0,2835	0,1692	0,3083	0,3758
Łódź	0,3774	0,1241	0,1286	0,3977	0,2214	0,0750
Nowy Sącz	0,0315	0,3219	0,2992	0,1660	0,2849	0,5733
Olsztyn	0,0715	0,5215	0,2844	0,2569	0,5624	0,3622
Opole	0,1163	0,3657	0,4160	0,3792	0,4818	0,2225
Ostrołęka	0,0302	0,4756	0,2454	0,1652	0,2512	0,2184
Piła	0,0416	0,6625	0,3444	0,2049	0,3198	0,3449
Piotrków Tryb.	0,0492	0,5242	0,3639	0,2230	0,2940	0,3025
Płock	0,0098	0,2622	0,1570	0,1131	0,2155	0,5594
Poznań	0,1754	0,2199	0,2928	0,4621	0,2570	0,1236
Przemyśl	0,0375	0,6775	0,3243	0,1932	0,3551	0,4320
Radom	0,0626	0,6705	0,4412	0,2571	0,4447	0,3441
Rzeszów	0,0503	0,5322	0,4957	0,2288	0,3162	0,3072
Siedlce	0,0250	0,4148	0,2596	0,1569	0,3055	0,6727
Sieradz	0,0661	0,5313	0,3877	0,2642	0,4004	0,3661
Skierniewice	0,0677	0,3309	0,3618	0,2376	0,4128	0,3567
Słupsk	0,1095	0,3001	0,3214	0,3165	0,5160	0,2560
Suwałki	0,0530	0,5588	0,2582	0,2145	0,3361	0,2449
Szczecin	0,1409	0,2536	0,2115	0,3402	0,5205	0,2319
Tarnobrzeg	0,0334	0,6032	0,2619	0,1764	0,3410	0,3647
Tarnów	0,0271	0,5707	0,3681	0,1734	0,2832	0,3820
Toruń	0,1149	0,4610	0,3948	0,4107	0,5931	0,2785
Wałbrzych	0,1161	0,3834	0,2346	0,3444	0,5496	0,2377
Włocławek	0,0767	0,5354	0,4014	0,2819	0,5152	0,2913
Wrocław	0,2431	0,2153	0,2020	0,5714	0,3409	0,1243
Zamość	0,0318	0,5381	0,2239	0,1684	0,3196	0,3392
Zielona Góra	0,1070	0,3405	0,2779	0,3121	0,6702	0,2660

Źródło: obliczenia własne.

W tablicach tych wyróżnione są, dla większej przejrzystości, stopnie przynależności /podobieństwa/ przekraczające 0,3. Analiza tych zestawień pozwala wyróżnić województwa, których stopień podobieństwa do kilku klas przekracza przyjętą wartość krytyczną. Są to województwa, które, mówiąc obrazowo, leżą w miejscu "rozmycia" się poszczególnych klas. Oprócz tego analiza stopni przynależności dla poszczególnych klas pozwala w pewnym sensie zidentyfikować te klasy.

Obecnie przedstawimy krótką, syntetyczną charakterystykę otrzymanych wyników.

Zwrócić należy uwagę, iż w przypadku każdej grupy usług wyodrębniły się klasy, dla których stopnie przynależności wielu województw przekraczają wielkość krytyczną 0,3. Poniżej klasy te zostaną przedstawione, przy czym wymienione zostaną jedynie województwa, których stopień przynależności do danej klasy jest największy lub jeden z największych

- dla usług kulturalno-rozrywkowych:
  klasa II - woj. bialskopodlaskie /0,6981/, chełmskie /0,6663/, konińskie /0,6535/, radomskie /0,8567/, sieradzkie /0,6817/, zamojskie /0,6373/, klasa województw środkowych i wschodnich;
  klasa III - woj. bydgoskie /05050/, gdańskie /0,6773/, katowickie /0,5150/, poznańskie /0,4958/, wrocławskie /0,6368/ - klasa województw, zawierających wielkie aglomeracje oprócz Warszawy i Łodzi;
  klasa IV - woj. elbląskie /0,5620/, gorzowskie /0,6714/, olsztyńskie /0,6295/, opolskie /0,6224/, pilskie /0,5583/, toruńskie /0,4903/, zielonogórskie /0,5786/ - klasa województw Ziem Zachodnich i Północnych;
- dla usług w zakresie ochrony zdrowia i opieki społecznej:

klasa II: woj. bydgoskie /0,6246/, kieleckie /0,6011/, krośnieńskie /0,6155/, pilskie /0,6363/, piotrkowskie /0,6773/, płockie /0,6302/, tarnowskie /0,6554/, toruńskie /0,7271/;
klasa III - woj. białostockie /0,5236/, gorzowskie /0,5941/, słupskie /0,6432/, suwalskie /0,5518/;
klasa IV: woj. gdańskie /0,5741/, lubelskie /0,4459/, poznańskie /0,5149/, szczecińskie /0,4159/, wrocławskie /0,5662/,
klasa V: woj. kaliskie /0,4111/, konińskie /1,0000/, skierniewickie /0,4020/;

- dla usług komunalno-mieszkaniowych:

klasa II: woj. chełmskie /0,6093/, ciechanowskie /0,6399/, kaliskie /0,7696/, płockie /0,6709/, skierniewickie /0,7105/;
klasa III: woj. gdańskie /0,6245/, krakowskie /0,7080/, szczecińskie /0,6727/,
klasa IV: woj. elbląskie /0,6221/, koszalińskie /0,5029/, legnickie /0,5840/, olsztyńskie /0,0236/, opolskie /0,5818/, słupskie /0,5098/, wałbrzyskie /0,7474/, zielonogórskie /0,5163/;
klasa V: woj. kieleckie /0,7997/, nowosądeckie /0,6006/, rzeszowskie /0,5625/, tarnowskie /0,5992/;

- dla usług w zakresie oświaty i wychowania:

klasa II: woj. chełmskie /0,5510/, ciechanowskie /0,5685/, krośnieńskie /0,5430/, łomżyńskie /0,5237/, nowosądeckie /0,6301/, suwalskie /0,5580/, tarnobrzeskie /0,6041/, tarnowskie /0,5075/;
klasa III: woj. elbląskie /0,6203/, gorzowskie /0,6238/, kaliskie /0,5079/, koszalińskie /0,6629/, pilskie /0,5283/, szczecińskie /0,6562/, wrocławskie /0,5221/, zielonogórskie /0,5409/;

- dla usług bytowych:

  klasa II: woj. białostockie /0,6212/, elbląskie /0,6602/, łomżyńskie /0,6364/, pilskie /0,6625/, przemyskie /0,6775/, radomskie /0,6705/, tarnobrzeskie /0,6032/;

  klasa III: woj. bielskobialskie /0,6274/, częstochowskie /0,0628 kaliskie /0,7463/, leszczyńskie /0,6956/;

  klasa IV: woj. bydgoskie /0,5051/, gdańskie /0,7374/, krakowskie /0,5366/, wrocławskie /0,5714/;

  klasa V: woj. jeleniogórskie /0,7538/, olsztyńskie /0,5624/, toruńskie /0,5931/, zielonogórskie /0,6702/;

  klasa VI: woj. nowosądeckie /0,5733/, płockie /0,5594/, siedleckie /0,6727/.

Pozostałe podzbiory rozmyte oparte są w zasadzie o klasy jednoelementowe.

Należy zwrócić uwagę, że poszczególne województwa mają największy stopień przynależności do tej klasy rozmytej, która powstała w oparciu o klasę w zwykłym sensie, do której należy dany obiekt. Widać to po prześledzeniu powyższego zestawienia i porównaniu go z klasyfikacjami otrzymanymi w 5.3. Wyjątkami w tym względzie są w grupie usług komunalno-mieszkaniowych województwa olsztyńskie i zielonogórskie oraz w grupie usług bytowych województwa olsztyńskie i toruńskie. Dla województw tych stopień przynależności do klasy rozmytej powstałej w oparciu o klasę /w zwykłym sensie/, do której należały te województwa, jest mniejszy niż stopień przynależności do pewnej innej klasy.

Powyżej zasygnalizowana zbieżność wynika z tego, iż klasyfikacja rozmyta tworzona opisanym w tym paragrafie sposobem jest w dużym stopniu zdeterminowana skonstruowaną uprzednio klasyfikacją w zwykłym sensie.

Województwa, dla których stopień przynależności jest największy, mogą być interpretowane jako województwa - reprezentanci klas. Zwrócić należy uwagę również, iż dla usług w zakresie oświaty i wychowania różnice między wartościami funkcji przynależności w poszczególnych obiektach są stosunkowo niewielkie. Potwierdza to jeszcze raz, iż z pewną rezerwą traktować należy rezultaty klasyfikacji dla tej grupy usług.

Innym ważnym elementem analizy klasyfikacji rozmytej jest identyfikacja województw "granicznych", tzn. województw, dla których stopnie przynależności do różnych klas są zbliżone, przy czym chodzi oczywiście o duże wartości funkcji przynależności.

Do takich województw w naszych badaniach należą:

- dla usług kulturalno-rozrywkowych: woj. białostockie, bielsko-bialskie, bydgoskie, częstochowskie, elbląskie, kieleckie, koszalińskie, legnickie, leszczyńskie, lubelskie, słupskie, toruńskie, wałbrzyskie, włocławskie;
- dla usług w zakresie ochrony zdrowia i opieki społecznej: woj. elbląskie, jeleniogórskie, kaliskie, katowickie, koszalińskie, łomżyńskie, nowosądeckie, olsztyńskie, opolskie, szczecińskie, wałbrzyskie, zielonogórskie;
- dla usług komunalno-mieszkaniowych: woj. białostockie, bielsko-bialskie, bydgoskie, gorzowskie, jeleniogórskie, katowickie, lubelskie, olsztyńskie, ostrołęckie, pilskie, przemyskie, radomskie, toruńskie, włocławskie, wrocławskie, zielonogórskie;
- dla usług w zakresie oświaty i wychowania: woj. bielsko-bialskie, bydgoskie, gdańskie, konińskie, krakowskie, lubelskie, płockie, rzeszowskie, tarnowskie;
- dla usług bytowych: woj. bydgoskie, ciechanowskie, katowickie, kieleckie, lubelskie, łódzkie, olsztyńskie, opolskie, sieradzkie, skierniewickie, słupskie, włocławskie.

Informacje powyższe, wynikające z klasyfikacji rozmytej są bogatsze od informacji wynikających z klasyfikacji w zwykłym sensie. Mogą mieć one znaczenie dla celów poznawczych przy podejmowaniu decyzji w sferze polityki społeczno-ekonomicznej kraju. Pozwalają one bowiem w większym stopniu identyfikować dysproporcje między województwami, co pozwoli na większą precyzję prowadzonej polityki.

5.6. Poziom usług a poziom urbanizacji i inwestycji.

Ostatnim etapem przestrzennych badań usług konsumpcyjnych dla ludności jest zastosowanie zaproponowanej w 4.3 procedury budowy modeli regresji przy zastosowaniu skonstruowanej klasyfikacji rozmytej, przy czym wykorzystana będzie rozmyta klasyfikacja względna, powstała z rozmytej klasyfikacji bezwzględnej przez wyznaczenie udziałów podobieństwa /wzór 4.3/.

Problem, który zostanie zbadany, dotyczy zależności między poziomem danej grupy usług a poziomem urbanizacji, a w dalszej kolejności poziomem urbanizacji i wielkością inwestycji w dziedzinie odpowiadającej danej grupie usług.

Na wstępie postawiona zostanie hipoteza, iż poziom urbanizacji w dużym stopniu determinuje poziom usług. Oczywiście nie jest to zależność przyczynowo-skutkowa ale w poziomie urbanizacji zawiera się również działanie wielu czynników kształtujących poziom usług. Można o tym przekonać się śledząc zestaw zmiennych przedstawiony w 5.2. Zależność jest zatem raczej symptomatyczna.

W badaniach naszych poziom urbanizacji mierzony jest zmienną Y - procent ludności miejskiej.

Poziom danej grupy usług może być natomiast mierzony za pomocą miary rozwoju tej grupy usług /wartości tej miary rozwoju przedstawione są w tablicach 5.5, 5.9, 5.13, 5.17 i 5.21/ Miary rozwoju oznaczone są: Z_1, Z_2, Z_3, Z_4, Z_5, gdzie indeks oznacza numer grupy usług.

Na wstępie policzono współczynniki korelacji między Y a Z_1, Z_2, Z_3, Z_4, Z_5 wg tradycyjnej formuły Pearsona dla całego zbioru obiektów oraz rozmyte współczynniki korelacji osobno dla każdej klasy rozmytej wg wzoru /4.6/. Rezultaty przedstawione są w tablicy 5.37.

Tablica 5.37

Wartości współczynnika korelacji i rozmytych współczynników korelacji między zmiennymi charakteryzującymi poziom usług i poziom urbanizacji

Zmienne	Wartość współczynnika korelacji	Wartości rozmytych współczynników korelacji dla klasy					
		I	II	III	IV	V	VI
Y, Z_1	0,896	0,908	0,880	0,873	0,878	0,911	0,879
Y, Z_2	0,804	0,866	0,747	0,737	0,780	0,759	0,865
Y, Z_3	0,857	0,917	0,848	0,867	0,860	0,810	0,771
Y, Z_4	-0,006	0,088	-0,030	-0,020	0,215	-0,114	-0,130
Y, Z_5	0,796	0,871	0,753	0,737	0,809	0,755	0,729

Źródło: obliczenia własne.

Wartości współczynników korelacji i rozmytych współczynników korelacji wskazują, iż zależności miar rozwoju czterech grup usług i poziomu urbanizacji są silne. Brak jest praktycznie zależności w przypadku miary rozwoju usług w zakresie oświaty i wychowania /wynika to z małej zmienności zmiennych tej grupy/. Ta grupa usług zostaje w dalszych badaniach pominięta.

Dla pozostałych grup usług można wyciągnąć podobne wnioski.

Siła zależności w poszczególnych klasach /mierzona rozmytym współczynnikiem korelacji/ różni się, ale nieznacznie. Maksymalna różnica sięga 0,15.
W przypadku usług kulturalno-rozrywkowych zależność jest najsilniejsza w klasach I i V /zawierających silnie zurbanizowane województwa warszawskie i łódzkie/, natomiast nieco słabsza i prawie równa w pozostałych klasach.
Podobnie jest dla usług w zakresie ochrony zdrowia i opieki społecznej.
Dla usług komunalno-mieszkaniowych zależność jest najsilniejsza dla klasy, zawierającej woj. warszawskie, a w dalszej kolejności dla klas III i IV, zawierających województwa Ziem Zachodnich i Północnych oraz województwa skoncentrowane wokół wielkich aglomeracji, natomiast najsłabsze dla woj. krośnieńskiego.
Dla usług bytowych zależność jest najsilniejsza również dla woj. warszawskiego, a w dalszej kolejności dla klasy województw wielkich aglomeracji /klasa IV/.
Widać zatem, iż wraz ze wzrostem poziomu urbanizacji wzrasta również zależność poziomu usług od poziomu urbanizacji, a poza tym zależność ta jest nieco silniejsza w klasach zawierających województwa zachodnie, niż tych, które zawierają województwa środkowe i wschodnie.
Korzystając z tego, iż zależność między poziomem usług a ziomem urbanizacji jest dla czterech grup usług silna, zbudowano modele regresji rozmytej. Dla każdego modelu wyznaczone zostały również wartości rozmytego współczynnika zbieżności i rozmytego współczynnika zmienności losowej wg wzorów /4.7/ i /4.8/.
Wyniki są następujące:

- dla usług kulturalno-rozrywkowych:

w klasie I: $\tilde{Z}_1 = 0,3499 \cdot Y + 22,6155$

$\varphi^2 = 0{,}1974 \qquad W = 0{,}0665$

w klasie II: $\hat{Z}_1 = 0{,}2738 \cdot Y + 25{,}6028$

$\varphi^2 = 0{,}2254 \qquad W = 0{,}0559$

w klasie III: $\hat{Z}_1 = 0{,}2327 \cdot Y + 27{,}9363$

$\varphi^2 = 0{,}2440 \qquad W = 0{,}0617$

w klasie IV: $\hat{Z}_1 = 0{,}2812 \cdot Y + 26{,}1913$

$\varphi^2 = 0{,}2287 \qquad W = 0{,}0583$

w klasie V: $\hat{Z}_1 = 0{,}2429 \cdot Y + 28{,}0139$

$\varphi^2 = 0{,}1745 \qquad W = 0{,}0572$

w klasie VI: $\hat{Z}_1 = 0{,}3366 \cdot Y = 24{,}3405$

$\varphi^2 = 0{,}2409 \qquad W = 0{,}0666$

- dla usług w zakresie ochrony zdrowia i opieki społecznej:

w klasie I: $\hat{Z}_2 = 0{,}4054 \cdot Y = 25{,}1795$

$\varphi^2 = 0{,}2573 \qquad W = 0{,}0902$

w klasie II: $\hat{Z}_2 = 0{,}2285 \cdot Y + 31{,}9316$

$\varphi^2 = 0{,}4512 \qquad W = 0{,}0825$

w klasie III: $\hat{Z}_2 = 0{,}2845 \cdot Y + 31{,}7903$

$\varphi^2 = 0{,}4570 \qquad W = 0{,}0898$

w klasie IV: $\hat{Z}_2 = 0{,}2757 \cdot Y + 30{,}9850$

$\varphi^2 = 0{,}3949 \qquad W = 0{,}0906$

w klasie V: $\hat{Z}_2 = 0{,}2733 \cdot Y + 28{,}2019$

$\varphi^2 = 0{,}4237 \qquad W = 0{,}0867$

w klasie VI: $\hat{Z}_2 = 0{,}4885 \cdot Y + 21{,}6054$

$\varphi^2 = 0{,}2834 \qquad W = 0{,}1048$

- dla usług komunalno-mieszkaniowych:

w klasie I: $\hat{Z}_3 = 0{,}2585 \cdot Y + 12{,}4176$

$\varphi^2 = 0{,}1608 \quad W = 0{,}0780$

w klasie II: $\hat{Z}_3 = 0{,}2288 \cdot Y + 13{,}0088$

$\varphi^2 = 0{,}2816 \quad W = 0{,}0911$

w klasie III: $\hat{Z}_3 = 0{,}2089 \cdot Y + 14{,}9109$

$\varphi^2 = 0{,}2486 \quad W = 0{,}0845$

w klasie IV: $\hat{Z}_3 = 0{,}2146 \quad Y + 14{,}6955$

$\varphi^2 = 0{,}2609 \quad W = 0{,}0833$

w klasie V: $\hat{Z}_3 = 0{,}1971 \quad Y + 15{,}4264$

$\varphi^2 = 0{,}3470 \quad W = 0{,}0988$

w klasie VI: $\hat{Z}_3 = 0{,}1484 \quad Y + 18{,}6087$

$\varphi^2 = 0{,}4448 \quad W = 0{,}1103$

- dla usług bytowych:

w klasie I: $\hat{Z}_5 = 0{,}3019 \cdot Y + 3{,}3074$

$\varphi^2 = 0{,}2761 \quad W = 0{,}1306$

w klasie II: $\hat{Z}_5 = 0{,}1601 \cdot Y + 8{,}8949$

$\varphi^2 = 0{,}4349 \quad W = 0{,}1393$

w klasie III: $\hat{Z}_5 = 0{,}1649 \cdot Y + 10{,}2638$

$\varphi^2 = 0{,}4584 \quad W = 0{,}1351$

w klasie IV: $\hat{Z}_5 = 0{,}1885 \cdot Y + 8{,}9859$

$\varphi^2 = 0{,}3464 \quad W = 0{,}1289$

w klasie V: $\hat{Z}_5 = 0{,}1504 \cdot Y + 10{,}3601$

$\varphi^2 = 0{,}4352 \quad W = 0{,}1340$

w klasie VI: $\hat{Z}_5 = 0,1496 \cdot Y + 9,6542$

$\varphi^2 = 0,4737 \quad W = 0,1432$

Analiza dwóch zaproponowanych mierników dobroci modelu wskazuje, iż najlepsze modele otrzymano dla usług kulturalno-rozrywkowych. Szczególnie dotyczy to współczynnika wyrazistości W. Jeśli chodzi o pozostałe grupy usług, to niezłe wyniki dały również modele dla usług komunalno-mieszkaniowych, ale dotyczy to w zasadzie tylko czterech pierwszych klas. Pozostałe modele, zwłaszcza dla usług bytowych, należy traktować z dużą rezerwą. Porównując te rezultaty z wartościami rozmytych współczynników korelacji, widać, iż dobroć modelu jest większa, im wyższy jest rozmyty współczynnik korelacji zmiennej Z_i z Y. Jest to prawidłowość zgodna z tą, która zachodzi przy stosowaniu zwykłych modeli regresji z jedną zmienną objaśniającą.

Interpretację tych modeli można przeprowadzić w odniesieniu do modeli dla usług kulturalno-rozrywkowych; w pozostałych przypadkach wartości wyznaczonych współczynników należy traktować z ostrożnością.

Wartość współczynnika przy zmiennej Y należy /podobnie, jak w przypadku zwykłych modeli regresji/ interpretować jako wpływ zmiennej Y na odpowiednią zmienną Z_i.

Tak więc, zauważmy, że w przypadku usług kulturalno-rozrywkowych wpływ ten jest największy w klasie I, wynosi 0,3499, oraz w klasie VI - 0,3366, w pozostałych wynosi on poniżej 0,3, a najmniej w klasie III - 0,2327. Zwraca uwagę relatywnie niski wpływ w klasach rozmytych, które były konstruowane w oparciu o jednoelementową klasę - woj. łódzkie, oraz klasę, zawierającą aglomeracje miejskie oprócz Warszawy, Łodzi i Szczecina.

W następnej kolejności postawiona została hipoteza o zależności poziomu usług i poziomu inwestycji. Poziom inwestycji mierzony był przez zmienne:

V_1 - przeciętne roczne nakłady inwestycyjne w dziedzinie kultury i sztuki na 1 mieszkańca w latach 1978-79 w tys.zł.

V_2 - przeciętne roczne nakłady inwestycyjne w dziedzinie ochrony zdrowia i opieki społecznej na 1 mieszkańca w latach 1978-79 w tys.zł.

V_3 - przeciętne roczne nakłady inwestycyjne w dziedzinie gospodarki komunalnej i mieszkaniowej na 1 mieszkańca w latach 1975-79 w tys.zł.

V_4 - przeciętne roczne nakłady inwestycyjne w dziedzinie oświaty i wychowania na 1 mieszkańca w latach 1978-79 w tys.zł.

Przyjęcie w trzech przypadkach dwuletnich okresów mierzenia nakładów inwestycyjnych, a także pominięcie nakładów w sferze odpowiadającej usługom bytowym wyniknęło z braku danych.

Wartości współczynników korelacji i rozmytych współczynników korelacji między odpowiednimi miarami rozwoju a wielkościami nakładów inwestycyjnych przedstawia tablica 5.38.

Tablica 5.38

Wartości współczynnika korelacji i rozmytych współczynników korelacji między zmiennymi charakteryzującymi poziom usług i poziom inwestycji

Zmienne	Wartość współczynnika korelacji	Wartość rozmytych współczynników korelacji dla klasy					
		I	II	III	IV	V	VI
V_1, Z_1	0,539	0,796	0,391	0,514	0,407	0,466	0,586
V_2, Z_2	0,370	0,621	0,261	0,231	0,325	0,370	0,476
V_3, Z_3	0,578	0,798	0,500	0,572	0,486	0,528	0,503
V_4, Z_4	0,157	0,213	0,193	0,172	0,068	0,133	0,131

Źródło: obliczenia własne.

Widać, że i tutaj zależność dla usług w zakresie oświaty i wychowania jest słaba /choć nieco silniejsza niż w poprzednim przypadku/. Oprócz tego pominięte zostaną nakłady w dziedzinie ochrony zdrowia i opieki społecznej, gdyż tutaj w zasadzie tylko w klasie I /gdzie rozmyty współczynnik korelacji wynosi 0,621/ zależność nie jest słaba.

W pozostałych dwóch grupach zależność jest silniejsza, w granicach: dla usług kulturalno-rozrywkowych 0,391 - 0,796, dla usług komunalno-mieszkaniowych 0,486 - 0,798. Widać zatem, że zależność między poziomem usług a poziomem inwestycji w tej dziedzinie jest słabsza niż zależność między poziomem usług a poziomem urbanizacji, oraz posiada większe zróżnicowanie siły w poszczególnych klasach rozmytych.

Skonstruowane modele regresji poziomu usług względem poziomu urbanizacji i poziomu inwestycji są następujące:

- dla usług kulturalno-rozrywkowych :

w klasie I: $\hat{Z}_1 = 0,1952 \cdot Y + 0,0270 \cdot V_1 + 29,8118$

$\varphi^2 = 0,1236 \qquad W = 0,0526$

w klasie II: $\hat{Z}_1 = 0,2728 \cdot Y + 0,0029 \cdot V_1 + 25,5844$

$\varphi^2 = 0,2245 \qquad W = 0,0588$

w klasie III: $\hat{Z}_1 = 0,2148 \cdot Y + 0,0219 \cdot V_1 + 28,2211$

$\varphi^2 = 0,2292 \qquad W = 0,0598$

w klasie IV: $\hat{Z}_1 = 0,2781 \cdot Y + 0,0077 \cdot V_1 + 26,1693$

$\varphi^2 = 0,2252 \qquad W = 0,0579$

w klasie V: $\hat{Z}_1 = 0,2380 \cdot Y + 0,0190 \cdot V_1 + 27,7364$

$\varphi^2 = 0,1523 \qquad W = 0,0534$

w klasie VI: $\hat{Z}_1 = 0{,}2629 \cdot Y + 0{,}0481 \cdot V_1 + 26{,}6039$

$\varphi^2 = 0{,}2186 \qquad W = 0{,}0634$

- dla usług komunalno-mieszkaniowych:

w klasie I: $\hat{Z}_3 = 0{,}1737 \cdot Y + 0{,}0015 \cdot V_3 + 11{,}7833$

$\varphi^2 = 0{,}1524 \qquad W = 0{,}0759$

w klasie II: $\hat{Z}_3 = 0{,}2316 \cdot Y - 0{,}0001 \cdot V_3 + 13{,}2835$

$\varphi^2 = 0{,}2823 \qquad W = 0{,}0912$

w klasie III: $\hat{Z}_3 = 0{,}2041 \cdot Y + 0{,}0002 \cdot V_3 + 14{,}6079$

$\varphi^2 = 0{,}2475 \qquad W = 0{,}0843$

w klasie IV: $\hat{Z}_3 = 0{,}2118 \cdot Y + 0{,}0001 \cdot V_3 + 14{,}4842$

$\varphi^2 = 0{,}2603 \qquad W = 0{,}0832$

w klasie V: $\hat{Z}_3 = 0{,}1795 \cdot Y + 0{,}0008 \cdot V_3 + 13{,}6992$

$\varphi^2 = 0{,}3444 \qquad W = 0{,}0985$

w klasie VI: $\hat{Z}_3 = 0{,}1319 \cdot Y + 0{,}0008 \cdot V_3 + 17{,}0274$

$\varphi^2 = 0{,}4405 \qquad W = 0{,}1098$

Dla wszystkich powyższych modeli uwagę zwracają wartości współczynników przy zmiennych obrazujących poziom inwestycji. Są one we wszystkich przypadkach bliskie zera. Oznacza to, że w porównaniu z wpływem poziomu urbanizacji, wpływ poziomu inwestycji na poziom usług kulturalno-rozrywkowych i komunalno-mieszkaniowych jest nieistotny.

Porównując wartości współczynników zbieżności i zmienności losowej powyższych modeli z modelami bez zmiennych obrazujących poziom inwestycji, skonstruowanymi uprzednio, trzeba stwierdzić znikomą poprawę ich wartości. I w tym przypadku do zaakceptowania są właściwie tylko modele dla usług kulturalno-rozrywkowych.

Dla modeli tych, w porównaniu z modelami poprzednio skonstruowanymi obserwuje się nieznaczne zmiany wartości współczynników przy zmiennej Y, interpretowanymi jako wpływ poziomu urbanizacji na poziom usług kulturalno-rozrywkowych. Wyjątkiem są tutaj modele dla klasy I i VI, dla których nastąpiło obniżenie wartości współczynników.

Konkludując, należy stwierdzić, iż dysproporcje w poziomie usług między miastem a wsią nie są likwidowane przez odpowiednie inwestowanie. Jest to niewątpliwie sygnał dla polityków i decydentów.

WNIOSKI I PROBLEMY OTWARTE.

Przedstawiona praca wykazała przydatność stosowanych metod analizy wielowymiarowej do przestrzennych badań usług konsumpcyjnych. Przeprowadzone badania empiryczne pozwalają na sformułowanie kilku ogólnych wniosków odnośnie przestrzennego zróżnicowania usług konsumpcyjnych dla ludności w Polsce:

1. Usługi kulturalno-rozrywkowe, usługi w zakresie ochrony zdrowia i opieki społecznej, usługi komunalno-mieszkaniowe i usługi bytowe wykazują duże zróżnicowanie przestrzenne w przekroju województw.
2. Usługi w zakresie oświaty i wychowania wykazują małe zróżnicowanie przestrzenne w przekroju województw; wynika to w dużej części z określonej polityki oświatowej prowadzonej po II wojnie światowej.
3. Województwa zawierające wielkie aglomeracje miejskie wykazują wyższy poziom usług niż pozostałe województwa.
4. Województwa zaliczane do tzw. Ziem Zachodnich i Północnych wykazują wyższy poziom usług niż województwa Polski środkowej i wschodniej.
5. Dla większości grup usług obszary typu regionalnego pokrywają się w dużym stopniu z regionami; wskazuje to na zależność poziomu usług od położenia województw.
6. Istnieje dość silna zależność między poziomem usług a poziomem urbanizacji danego województwa; wynika stąd wniosek o relatywnym niedorozwoju usług w województwach o dużym udziale ludności wiejskiej.

7. Na poziom usług nie ma praktycznie wpływu poziom inwestycji w ostatnich latach.

Na tle powyższych wniosków nasuwa się jeszcze jedna konkluzja: występujące dysproporcje w poziomie usług, będące efektem rozwoju społeczno-gospodarczego w przeszłości, są likwidowane w niedostatecznym stopniu.

Jak większość prac metodycznych, tak i ta pozostawią pewne otwarte problemy teoretyczne. Część z nich wiąże się z metodami klasyfikacji. Jest ich olbrzymia ilość, ale brak jest propozycji odnośnie zespołu kryteriów, które pozwoliłyby ocenić te metody i dokonać przynajmniej ich częściowego wartościowania. Spowodowałoby to skoncentrowanie się badaczy wokół pewnej grupy najlepszych metod, przydatnych dla badań różnych rodzajów. Zaznaczamy, że chodzi tu o kryteria formalne, zdając sobie sprawę, że sama natura badanego problemu narzuca pewne wskazówki odnośnie stosowania określonej metody.

Innym niebagatelnym problemem jest sposób określania najwłaściwszej ilości klas, na jakie powinien zostać podzielony zbiór obiektów, w przypadku gdy nie ma tej informacji danej z góry. Zagadnienie to dotyczy zarówno metod klasyfikacji, jak i regionalizacji oraz klasyfikacji rozmytej /w pracy zaproponowano jeden taki sposób/.

Często do opisu badanych zjawisk złożonych stosowane są zbiory zawierające zmienne różnego typu /prezentowane na skali nominalnej, porządkowej, ilorazowej, ciągłe, dyskretne/. Nie został do tej pory dostatecznie rozwiązany problem odpowiedniej dla tego typu zbioru zmiennych funkcji podobieństwa czy odległości. Szereg otwartych problemów wiąże się również z narzędziami badawczymi bazującymi na pojęciu rozmytości. W pracy zaproponowano pewne procedury przydatne do analizy i prognozowania zjawisk. Wiąże się z tym wiele zagadnień. Należy do nich np. dobór postaci modelu przy konstruowaniu modeli dla klas rozmytych, wybór właściwych mierników dobroci prognozy, prognozy przedziałowe, itp.

LITERATURA.

1. Ajwazjan S.A., Bieżajewa J., Starowierow O.W. - Kłassifikacija mnogomiernych nabljudjenij, Statistika, Moskwa 1974.

2. Ajzerman M.A., Brawerman E.M., Rozonoer L.J. - Rozpoznawanie obrazów. Metoda funkcji potencjalnych, WN-T, Warszawa 1976.

3. Ałampiew P.M. - Ob objektivnoj osnowie ekonomiczeskowo rajonirowanja w SSSR, w: Woprosy rozmieszczenija proizwodstwa i ekonomiczeskowo rajonirowanja, s.239-253, Gospłanizdat, Moskwa 1960.

4. Anderberg C.M. - Cluster analysis for applications, Academic Press, New York 1973.

5. Anderson T.W. - An introduction to multivariate statistical analysis, Wiley, New York 1958.

6. Batchelor B.G. - Practical approach to pattern classification, Plenum Press, London 1974.

7. Benzecri J.P. i in. - L'analyse des données, t.1 - La taxinomie, t.2 - L'analyse des correspondances, Dunod, Paris 1976.

8. Berezowski St. - Wstęp do regionalizacji gospodarczej, SGPiS, Warszawa 1968.

9. Berry B.J. - A method of deriving multifactor uniform regions, w: Przegląd Geograficzny 2/1961, s.263-282.

10. Berry B.J. - A note concerning methods of classification, w: Annals of the Association of American Geographers, 3/1958, s.300-304.

11. Bertier P., Bouroche J.M. - Analyse des données multidimensionnelles, Presses Universitaires de France, Paris 1975.

12. Bezdek J.C. - Numerical taxonomy with fuzzy sets, w: Journal of Mathematical Biology 1/1974, s.57-71.

13.Bezdek J.C., Harris J.D. - Fuzzy partitions and relations: an axiomatic basis for clustering, w: Fuzzy sets and systems, 1/1978, s.111-127.

14.Bielecki Cz. - Ekonomika i planowanie rozwoju regionów, PWE, Warszawa 1974.

15.Bobek H. - Some remarks on basis concepts in economic regionalization, w: Economic regionalization, Proceedings of the 4th General Meeting of the Commission on Methods of Economic Regionalization of the IGU, s.17-24, Academia, Praga 1967.

16.Bobiński J. - Regionalizacja oświaty i kultury, w: Wiadomości Statystyczne 3/1969, s.22-25.

17.Bock R.D. - Multivariate statistical methods in behavioral research, McGraw Hill, New York 1975.

18.Borys T. - Metody normowania cech w statystycznych badaniach porównawczych, w: Przegląd Statystyczny 2/1978, s.227-239.

19.Boyce A.J. - Mapping diversity: a comparative study of some numerical methods, w: Cole A.J. /ed./ - Numerical taxonomy, s.1-30, Academic Press, New York 1969.

20.Bukietyński W., Hellwig Z., Królik U., Smoluk A. - Uwagi o dyskryminacji zbiorów skończonych, w: Prace Naukowe WSE Wrocław 21, s.111-122, Wrocław 1969.

21.Byfuglien J., Norgard A. - Konstruowanie regionów-porównanie metod, w: Przegląd Zagranicznej Literatury Geograficznej, 3-4/1976, s.111-143.

22.Chang S.S.L. - Applications of fuzzy sets theory to economics, w: Kybernetes 6/1977, s.203-207.

23.Chang S.S.L., Zadeh L.A. - On fuzzy mapping and control, w: IEEE Transactions on Systems, Man and Cybernetics SMC-2, 1972, s.30-34.

24.Chojnicki Z. - Podstawy teoretyczne zastosowania metod matematycznych w badaniach struktury przestrzennej rolnictwa, w: Metody matematyczne i taksonomiczne w badaniach struktury przestrzennej rolnictwa, Biuletyn KPZK PAN 61, s.7-41, PWN, Warszawa 1970.

25.Chojnicki Z., Czyż T. - Analiza czynnikowa w geografii, w: Metody ilościowe i modele w geografii, s.77-93, PWN Warszawa 1977.

26.Chojnicki Z., Czyż T. - Metody taksonomii numerycznej w regionalizacji geograficznej, PWN, Warszawa 1973.

27.Chojnicki Z., Czyż T. - Zmiany struktury regionalnej Polski w świetle przepływów towarowych w latach 1958-1966, Studia KPZK PAN 50, PWN, Warszawa 1972.

28.Cieślak M. - Modele zapotrzebowania na kadry kwalifikowane, PWE, Warszawa 1976.

29.Cormack R.M. - A review of classification, w: Journal of the Royal Statistical Society A 134.3/1971, s.321-353.

30.Cox D.R. - Note on grouping, w: Journal of the American Statistical Association 52.280/1957, s.543-547.

31.Czyż T. - Wyznaczanie regionów jednolitych metodą analizy czynników wielokrotnych, w: Przeglad Geograficzny 1/1967, s.135--160.

32.Czyż T. - Zastosowanie metod analizy czynnikowej do badania ekonomicznej struktury regionalnej Polski, Prace Geograficzne IG PAN 92, Ossolineum, Wrocław 1971.

33.Czyż T. - Zastosowanie metod i modeli matematycznych w geografii polskiej, w: Przegląd Geograficzny 1/1973, s.29-49.

34.Czyż T. - Zastosowanie metody czynnikowej w badaniach przestrzenno-ekonomicznych, w: Przegląd Geograficzny 3/1970, s.467-486.

35. Daszkowska M. - Ekonomika i organizacja usług, UG, Gdańsk 1978.

36. Domański R. - Geografia ekonomiczna, PWN, Warszawa-Poznań 1977.

37. Dlin A.M. - Matiematiczeskaja statistika w tiechnikie, Sowietskaja nauka, Moskwa 1958.

38. Duda R.O., Hart P.E. - Pattern classification and scene analysis, Wiley, New York 1973.

39. Dunn J.C. - Indices of partition fuzziness and detection of clusters in large data sets, w: Fuzzy automata and decision processes, s.271-284, North Holland, New York 1977.

40. Duran B.S., Odell P.L. - Cluster analysis, a survey, Springer Verlag, Berlin 1974.

41. Działalność usługowa w 1979. Informacje. Materiały Statystyczne, GUS, Warszawa 1980.

42. Dziewoński K. - Concepts and terms in the field of economic regionalization, w: Economic regionalization, Proceedings of the 4th General Meeting of the Commission on Methods of Economic Regionalization of the IGU, s.25-30, Academia, Praga 1967.

43. Dziewoński K. - Economic regionalization. A report of progress, w: Economic regionalization and numerical methods, Geographia Polonica 15, s.9-24, PWN, Warszawa 1968.

44. Dziewoński K. - Elementy teorii regionu ekonomicznego, w: Przegląd Geograficzny 4/1961, s.593-613.

45. Dziewoński K., Leszczycki S., Otremba E., Wróbel A. - Review of concepts and theories of economic regionalization, w: Methods of economic regionalization, Geographia Polonica 4, s.11-23, PWN, Warszawa 1964.

46. Ekonomika i organizacja usług, PTE, Poznań 1976.

47. Everitt B. - Cluster analysis, Heinemann, London 1974.

48.Fajferek A. - Region ekonomiczny i metody analizy regionalnej, PWE, Warszawa 1966.

49.Fischer M.M. - Regional taxonomy: a comparison of some hierarchic and non-hierarchic strategies, Lagos 1978.

50.Florek K., Łukaszewicz J., Perkal J., Steinhaus H., Zubrzycki S. - Taksonomia wrocławska, w: Przegląd Antropologiczny, 18/1951, s.193-211.

51.Friedman H.P., Rubin J. - On some invariant criteria for grouping data, w: Journal of the American Statistical Association, 62.320/1967, s.1159-1178.

52.Fu K.S. - Syntactic methods in pattern recognition, Academic Press, New York 1974.

53.Goldberger A.S. - Teoria ekonometrii, PWE, Warszawa 1975.

54.Goodman L.A., Kruskal W.H. - Measures of association for cross classifications, w: Journal of the American Statistical Association, I-49/1954, s.723-764, II-54/1959, s.123-163.

55.Gorelik A.Ł., Skripkin W.A. - Metody raspoznawanja, Wysszaja szkoła, Moskwa 1977.

56.Gorzelak G. - Dobór zmiennych w statystycznej analizie porównawczej- zagadnienia teoretyczne, w: Wiadomości Statystyczne, 3/1979, s.17-21.

57.Gorzelak G. - Dobór zmiennych w statystycznej analizie porównawczej - metody, w: Wiadomości Statystyczne 4/1979, s.17-21.

58.Gorzelak G. - Zastosowanie statystycznej analizy porównawczej do badania przestrzennej struktury rolnictwa w Polsce, praca doktorska, Wrocław 1980.

59.Gower J.C. - A comparison of some methods of cluster analysis, w: Biometrics 23/1967, s.623-637.

60.Gower J.C. - A general coefficient of similarity and some of its properties, w: Biometrics 27/1971, s.857-874.

61. Gower J.C. - Some distance properties of latent root and vector methods used in multivariate analysis, w: Biometrika 53/1966, s.325-338.

62. Gupta M.M., Ragade R.K. - Fuzzy set theory: introduction, w: Fuzzy automata and decision processes, s.105-131, North Holland, New York 1977.

63. Gusiew Ł.A., Smirnowa J.M. - Razmytyje množestwa. Teorija i priłożenija /obzor/, w: Awtomatika i tielemechanika 5/1973, s.66-83.

64. Harman J. - Modern factor analysis, Chicago University Press, Chicago 1960.

65. Hellwig Z. - Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr, w: Przegląd Statystyczny 4/1968, s.307-327.

66. Hellwig Z., Kania-Gospodarowicz A. - Zastosowanie analizy porównawczej w badaniach międzynarodowych, Z prac ZBSE 83, Warszawa 1975.

67. Isard W. - Metody analizy regionalnej. Wprowadzenie do nauki o regionach, PWN, Warszawa 1966.

68. Iszkowski J., Kalestyńska C. - Klasyfikacja usług, w: Wiadomości Statystyczne 5/1970, s.1-4.

69. Jagielski Z. - Analiza czynnikowa w badaniach ekologiczno-osadniczych, w: Metody ilościowe i modele w geografii, s.128-142, PWN, Warszawa 1977.

70. Jardine N., Sibson R. - Mathematical taxonomy, Wiley, New York 1971.

71. Jelisiejewa J.J., Rukawisznikow W.O. - Gruppirowka, korelacija, raspoznawanje obrazow, Statistika, Moskwa 1977.

72. Jensen R.E. - A dynamic programming algorithm for cluster analysis, w: Operations Research 17.6/1969, s.1034-1057.

73. Johnson S.C. - Hierarchical clustering schemes, w: Psychometrika 32.3/1968, s.241-254.

74. Kaczmarek Z., Parysek J.J. - Zastosowanie analizy wielowymiarowej w badaniach geograficzno-ekonomicznych, w: Metody ilościowe i modele w geografii, s.94-127, PWN, Warszawa 1977.

75. Kaufmann A. - Introduction a la theorie des sous-ensembles flous a l'usage des ingenieurs, t.1 - elements theoriques de base, Masson, Paris 1977.

76. Kendall M.G. - Discrimination and classification, w: Krishnaiah P.R. /ed./ - Multivariate analysis, s.165-185, Academic Press, New York 1966.

77. Kendall M.G. - Multivariate analysis, Griffin, London 1975.

78. Kendall M.G., Stuart A. - Mnogomiernyj statisticzeskij analiz i wremiennyje rjady, Nauka, Moskwa 1976.

79. King B. - Step-wise clustering procedures, w: Journal of the American Statistical Association, 62.317/1967, s.86-101.

80. Klasyfikacja usług obowiązująca od 1976, Zeszyty Metodyczne GUS 16, GUS, Warszawa 1976.

81. Kolonko J. - Analiza dyskryminacyjna i jej zastosowania w ekonomii, PWE, Warszawa 1980.

82. Kolonko J., Stolarska E., Zadora K. - Prosta metoda dyskryminacji zbiorów skończonych, w: Przegląd Statystyczny 2/1970, s.173-186.

83. Kowalski A. - Charakterystyka niektórych metod taksonomicznych, PWN, Warszawa 1977.

84. Kucharczyk J. - Metody analizy skupień i ich realizacja w języku ALGOL 60, Instytut Informatyki UW, Wrocław 1978.

85.Lance G.N., Williams W.T. - A general theory of classificatory sorting strategies, 1.Hierarchical systems, w: The Computer Journal 9 1966/1967, s.373-380.

86.Lance G.N., Williams W.T. - A generalized sorting strategy for computer classifications, w: Nature 212/1966, s.218.

87.Lance G.N., Williams W.T. - Computer program for hierarchical polythetic classification - similarity analysis, w: The Computer Journal 9 1966/1967, s.60-64.

88.Lance G.N., Williams W.T. - Computer program for monothetic classification - association analysis, w: The Computer Journal 8 1965/1966, s.246-249.

89.Lange O. - Ekonomia polityczna t.1, PWN, Warszawa 1959.

90.Lawley D., Maxwell A. - Faktornyj analiz kak statisticzeskij metod, Mir, Moskwa 1967.

91.Leszczycki St. - Rozwój myśli geograficznej, w: Geografia powszechna t.1, s.20-56, PWN, Warszawa 1962.

92.Leszczycki St. - The tasks of economic regionalization, w: Methods of economic regionalization, Geographia Polonica 4, s.24-35, PWN, Warszawa 1964.

93.MacNaughton-Smith P., Williams W.T., Dale M.B., Mockett L.G. - Dissimilarity analysis: a new technique for hierarchical sub-division, w: Nature 202/1964, s.1034-1035.

94.Mała encyklopedia ekonomiczna, PWE, Warszawa 1976.

95.Markos G. - Teoretyczne podstawy geograficznego podziału pracy i regionalizacji gospodarki narodowej oraz ich znaczenie praktyczne, w: Przegląd Zagranicznej Literatury Geograficznej, 1/1956, s.40-71.

96.Marriott F.H.C. - Practical problems in a method of cluster analysis, w: Biometrics 27/1971, s.501-514.

97. McQueen J.B. - Some methods for classification and analysis of multivariate observations, Proooceedings Symp. Math. Statistics and Probability 5th Berkeley 1, s.281-297, University of California Press, Berkeley 1967.

98. Megee M. - Nowe dziedziny zastosowania analizy czynnikowej: sprawdzanie hipotez dotyczących rozwoju gospodarczego, w: Biuletyn KPZK PAN 34, s.187-209, PWN, Warszawa 1965.

99. Mnogomiernyj statisticzeskij analiz w socjalno-ekonomiczeskich issledovanjach, Nauka, Moskwa 1974.

100. Morrison D.F. - Multivariate statistical methods, McGraw Hill, New York 1976.

101. Mucznik J.B., Żukowskaja W.M. - Faktornyj analiz w socjalno-ekonomiczeskich issledovanjach, Statistika, Moskwa 1976.

102. Muszyński W. - Zastosowanie formalizmu rozmytości w teorii systemów zdarzeniowych, ICT PW, komunikat 595, Wrocław 1977.

103. Nagy G. - State of the art in pattern recognition, w: Proceedings of IEEE, 56/1968, s.836-862.

104. Negoita C.V., Ralescu D.A. - Applications of fuzzy sets to systems analysis, Birkhauser Verlag, Basel 1975.

105. Niewadzi Cz. - Rozwój sfery usług w Polsce i innych krajach socjalistycznych, CINTiE, Warszawa 1972.

106. Niewadzi Cz. - Usługi w gospodarce narodowej, PWE, Warszawa 1975.

107. Okóń J. - Analiza czynnikowa w psychologii, PWN, Warszawa 1964.

108. Parks J.M. - Classification of mixed mode data by R-mode factor analysis and Q-mode cluster analysis on distance function, w: Cole A.J. /ed./ - Numerical taxonomy, s.216-219, Academic Press, New York 1969.

109. Parysek J.J. - Zastosowanie taksonomicznej odległości Mahalanobisa w dynamicznych badaniach strukturalno-przestrzennych, w: Przegląd Geograficzny 2/1978, s.293-308.

110. Pluta W. - Grafowa metoda klasyfikacji cech I, w: Prace Naukowe WSE Wrocław 21, s.235-241, Wrocław 1969.

111. Pluta W. - Grafowa metoda klasyfikacji cech II, w: Prace Naukowe WSE Wrocław 35, s.195-207, Wrocław 1972.

112. Pluta W. - Metody wielowymiarowej analizy porównawczej w modelowaniu informacji ekonomicznej kombinatu przemysłowego, Prace Naukowe AE Wrocław 156, Wrocław 1979.

113. Pluta W. - Przyczynek do grafowej metody klasyfikacji cech I, w: Prace Naukowe WSE Wrocław 33, s.145-154, Wrocław 1972.

114. Pluta W. - Wielowymiarowa analiza porównawcza w badaniach ekonomicznych, PWE, Warszawa 1977.

115. Pociecha J., Woźniak M., Zając K. - Ilościowe metody badania usług, w: Kursokonferencja naukowa nt. Rynek usług i przedsiębiorstwo usługowe - organizacja, sterowanie, zarządzanie, zeszyt 12, Bydgoszcz 1978.

116. Podolec B., Zając K. - Ekonometryczne metody ustalania rejonów konsumpcji, PWE, Warszawa 1978.

117. Problems of economic region, Papers of the Conference on Economic Regionalization in Kazimierz, Geographical Studies 27, PWN, Warszawa 1961.

118. Racine J.B., Reymond H. - Analiza ilościowa w geografii, PWN, Warszawa 1977.

119. Rand W.M. - Objective criteria for the evaluation of clustering methods, w: Journal of the American Statistical Association, 66.336/1971, s.846-850.

120.Rao M.R. - Cluster analysis and mathematical programming, w: Journal of the American Statistical Association, 66.335/1971, s.622-627.

121.Raspoznawanje obrazow pri postrojenji ekonomiko-statisticzeskich modeliej, Nauka, Nowosybirsk 1975.

122.Rocznik statystyczny 1980, GUS, Warszawa 1980.

123.Rocznik statystyczny województw,1980, GUS, Warszawa 1980.

124.Roubens M. - Pattern classification and fuzzy sets, w: Fuzzy sets and systems 1/1978, s.239-253.

125.Rozin B.B. - Teoria rozpoznawania obrazów w badaniach ekonomicznych, PWE, Warszawa 1979.

126.Ruspini E.H. - A new approach to clustering, w: Information and Control 15/1969, s.22-32.

127.Ruspini E.H. - New experimental results in fuzzy clustering, w: Information Sciences 6/1973, s.273-284.

128.Ruspini E.H. - Numerical methods for fuzzy clustering, w: Information Sciences 2/1970, s.319-350.

129.Sebestyen G.S. - Decision making processes in pattern recognition, McMillan, New York 1962.

130.Siedlecka U. - Zastosowanie metody taksonomii stochastycznej do dyskryminacji zbiorów skończonych, w: Przegląd Statystyczny, 3/1976, s.275-288.

131.Skrócona klasyfikacja usług dla potrzeb planowania i sprawozdawczości w zakresie usług, GUS, Warszawa 1980.

132.Sneath P.H.A. - Evaluation of clustering methods, w: Cole A.J. /ed./ - Numerical taxonomy, s.257-267, Academic Press, New York 1969.

133.Sneath P.H.A., Sokal R.R. - Numerical taxonomy, Freeman and Co., San Francisco 1973.

134.Sneath P.H.A., Sokal R.R. - Numerical taxonomy, w: Nature 193/1962, s.855-860.

135.Strahl D. - Modelowanie zjawisk złożonych. Modele infrastruktury społecznej, Prace Naukowe AE Wrocław 158, Wrocław 1980.

136.Strohmeier A. - La classification automatique, les bases, Neuchatel 1977.

137.Styczeń M. - Niektóre właściwości podziału wrocławskiego, w: Metody matematyczne w socjologii, s.50-67, Ossolineum, Wrocław 1971.

138.Styczeń M. - Ogólna charakterystyka metod taksonomicznych, w: Wiadomości Statystyczne 8/1971, s.16-20.

139.Styś A. - Rynek usług w ujęciu przestrzennym, PWE, Warszawa 1977.

140.Styś A., Olearnik J., Kurczyna J. - Ekonomika i organizacja usług, AE, Wrocław 1978.

141.Tamura S., Higuchi S., Tanaka K. - Pattern classification based on fuzzy relations, w: IEEE Transactions on Systems, Man and Cybernetics, SMC-1, 1971, s.61-66.

142.Thomas M.D. - Kilka uwag w sprawie rozwoju i współczesnego zastosowania metody regionalnej w Stanach Zjednoczonych, w: Przegląd Geograficzny 2/1961, s.251-262.

143.Vinod H.D. - Integer programming and the theory of grouping, w: Journal of the American Statistical Association, 64.326/1969, s.507-519.

144.Ward J.H. - Hierarchical grouping to optimize an objective function, w: Journal of the American Statistical Association, 58.301/1963, s.236-244.

145.Whittlesey D. - Regionalnaja koncepcija i regionalnyj metod, w: Amierikanskaja geografija, s.39-80, Izdatielstwo innostrannoj litieratury, Moskwa 1957.

146. Winiarski B. - Polityka regionalna, PWE, Warszawa 1976.

147. Wishart D. - An algorithm for hierarchical classification, w: Biometrics 25/1969, s.165-170.

148. Wishart D. - Mode analysis: a generalization of nearest neighbour which reduces chaining effects, w: Cole A.J. /ed./ - - Numerical taxonomy, s.282-306, Academic Press, New York 1969.

149. Wishart D. - Numerical classification method for deriving natural classes, w: Nature 221/1969, s.97-98.

150. Wróbel A. - Pojęcie regionu ekonomicznego a teoria geografii, Prace Geograficzne IG PAN 48, PWN, Warszawa 1965.

151. Zadeh L.A. - Fuzzy sets, w: Information and Control, 8/1965, s.338-353.

152. Zadeh L.A. - Fuzzy sets and their application to pattern classification and cluster analysis, w: J. van Ryzin /ed./ - Classification and clustering, s.251-300, Academic Press, New York 1977.

153. Zadeh L.A. - Poniatje lingwisticzeskoj pieremiennoj i jewo primienienje k priniatju pribliżennych reszenij, Mir, Moskwa 1976.

154. Zając K. - Zarys metod statystycznych, PWE, Warszawa 1976.

155. Zasady metodyczne sprawozdawczości statystycznej z usług dla ludności, GUS, Warszawa 1975.

156. Żytkow G.N. - Niekatoryje metody awtomaticzeskoj kłassifikaciji, w: Strukturnyje metody opoznawanja i awtomaticzeskoje cztjenje, s.66-85, [redacted].