

A-1012

INSTYTUT TURYSTYKI
WARSZAWA

INSTYTUT EKONOMIKI OBROTU TOWAROWEGO
AKADEMII EKONOMICZNEJ W KRAKOWIE

**PRZESTRZENNE MODELE ANALIZY ROZWOJU
TURYSTYKI KRAJOWEJ**

(etap II)

temat III. 4.6. realizowane w ramach CPBP 08.06

**„TURYSTYKA JAKO CZYNNIK ROZWOJU
SPOŁECZNO-GOSPODARCZEGO”**

Autorzy:

Doc. dr hab. Tadeusz Grabiński
Dr Anna Nowakowska

Kraków — Warszawa 1987

S p i s t r e ś c i

Wstęp.	3
Część I. Metody analizy stosowane w badaniach.	8
1.1 Analiza korelacji.	8
1.2 Analiza regresji.	12
1.3 Analiza czynnikowa.	20
1.4 Wielowymiarowa analiza porównawcza.	44
Część II Wyniki badań empirycznych.	58
2.1 Kwantyfikacja układów analizy.	58
2.2 Analiza czynnikowa determinant krajowego ruchu turystycznego.	60
2.3 Wyznaczenie syntetycznych mierników - determinant krajowego ruchu turystycznego.	68
2.4 Ekonometryczne model przyczynowo-opisowe ruchu turystycznego w Polsce	71
2.5 Regionalizacja Polski na podstawie zmiennych agregato- wych.	85

W S T, E P

W raporcie przedstawione są wyniki badań uzyskanych w trakcie realizacji II etapu tematu III-4-6 "Przestrzenne modele analizy rozwoju turystyki krajowej" wchodzącego w skład CFBP 08.06 "Turystyka jako czynnik rozwoju społeczno-gospodarczego". Zgodnie z założonym harmonogramem prac przebiegały one w dwóch kierunkach.

Pierwszy z nich polegał na utworzeniu komputerowej bazy danych, zawierającej informacje statystyczne za lata 1974-1985 w przekroju województw ujęte w dwóch głównych działach: GOSPODARKA oraz TURYSTYKA. W ramach każdego działu wyróżniono następujące grupy tematyczne"

I. GOSPODARKA

1. Demografia
2. Zatrudnienie
3. Handel
4. Gastronomia
5. Komunikacja
6. Kultura
7. Zasobność
8. Ochrona środowiska
9. Osadnictwo

II. TURYSTYKA

1. Obiekty turystyczne
2. Miejsca noclegowe
3. Korzystający z noclegów
4. Udzielone noclegi
5. Walory turystyczne
6. Urządzenia turystyczne
7. Biura obsługi
8. Inne

Lista zmiennych znajdujących się w bazie zawiera blisko 200 pozycji, wliczając w to tylko zmienne o charakterze pierwotnym, dla których zbierano informacje ze źródeł statystycznych (Roczniki Statystyczne Województw, GUS). Do systemu administru-

jącego bazą dołączono pakiet programów pozwalających generować różnie zdefiniowane zmienne wynikowe (udziały procentowe województw, indeksy dynamiki o podstawie stałej i ruchomej, współczynniki lokalizacji według wybranych zmiennych np. powierzchni, liczby ludności itd.). W rezultacie istnieje możliwość ponad 10-krotnego zwiększenia ilości zmiennych dostępnych do bezpośredniego wykorzystania w trakcie obliczeń i analiz na podstawie założonej bazy danych. Przyjmując, że pojedyncza zmienna obejmuje 500 obserwacji (10 lat x 49 województw + Polska ogółem) daje to zbiór zawierający blisko 1 mln informacji statystycznych.

W ramach omawianego kierunku badań podjęto również prace (wykraczające poza zakres ujęty harmonogramem) polegające na adaptacji zbiorów taśmowych znajdujących się w dyspozycji GUS i zawierających informacje źródłowe w zakresie obejmującym dane znajdujące się w sprawozdaniu Kt-1. Przetransponowano na EMC CYBER-72, który to system jest wykorzystywany do obliczeń w temacie, pełne zbiory danych za lata 1983-1986 (informacje za lata poprzednie uległy likwidacji).

W efekcie uzyskano bezpośredni dostęp do danych źródłowych o bazie i ruchu turystycznym w przekroju województw, gmin i miejscowości oraz możliwość dowolnej ich agregacji według rodzajów obiektów, ich kategorii, gestorów, zakresu świadczonych usług oraz czasokresu użytkowania obiektów (według poszczególnych miesięcy).

Dane ze sprawozdań Kt-1 tylko częściowo będą wykorzystywane do celów podjętego tematu, tym niemniej mogą być przetwarzane dla potrzeb wykonawców innych tematów.

W następnym etapie prac w roku 1988 przewiduje się aktualizację bazy o dane dotyczące lat 1986 i 1987 oraz ewentualnie dalsze

rozszerzanie zbioru o nowe zmienne pierwotne. W tym też etapie nastąpi ostateczna redakcja bazy w postaci wydruku komputerowego zawierającego dane liczbowe wraz z elementami ich wizualizacji (wykresy tendencji rozwojowej, kartogramy przestrzennego rozmieszczenia zmiennych, histogramy, krzywe Lorenza itd.). Należy zauważyć, że utworzona baza pozwoli ^{na} tylko na realizację podjętego tematu, ale także może stanowić podstawę podejmowania innych badań ukierunkowanych na przestrzenne aspekty turystyki krajowej.

Uwzględniając duży nakład środków przeznaczonych na utworzenie komputerowej bazy danych GCSTUR wydaje się celowym rozważenie możliwości wyodrębnienia jej w postaci odrębnego podzadania badawczego. Produktem końcowym byłby tu system oprogramowania użytkowego, zbiory danych liczbowych oraz instrukcje ich wykorzystywania, ujęte w postaci maszynowych nośników informacji (dyskietki lub taśmy magnetyczne) i przystosowane do eksploatacji na ogólnie dostępnym sprzęcie komputerowym, czy nawet mikrokomputerowym (IBM PC XT/AT). System taki mógłby być przekazywany innym użytkownikom, dzięki czemu badania merytoryczne można byłoby rozpoczynać bez potrzeby żmudnych prac polegających na zbieraniu i weryfikacji danych statystycznych, a sprowadzających się tylko do uzupełnienia istniejącej bazy o nowe dane.

Drugi kierunek prac realizowanych w ramach tematu miał na celu przeprowadzenie analizy strukturalnych prawidłowości krajowego ruchu turystycznego oraz jego determinant w układzie województw. Korzystając z zebranego materiału statystycznego wyróżniono cztery układy zmiennych charakteryzujących:

- A) poziom infrastruktury turystyki krajowej,
- B) wielkość zanieczyszczenia środowiska naturalnego,
- C) walory wypoczynkowe i turystyczne,
- D) bazę krajowego ruchu turystycznego,

które traktowano jako czynniki kształtujące wielkość krajowego

ruchu turystycznego.

Przy pomocy dwóch alternatywnych podejść: analizy czynnikowej oraz wielowymiarowej analizy porównawczej skwantyfikowano realizacje mierników syntetycznych oraz głównych składowych, określających determinanty ruchu turystycznego w przekroju województw dla roku 1983.

Informacje te wykorzystano do:

- 1) rejonizacji Polski w jednowymiarowych przestrzeniach mierników syntetycznych oraz czynników głównych,
- 2) konstrukcji ekonometrycznych modeli przyczynowo-opisowych charakteryzujących związki między wielkością krajowego ruchu turystycznego a wyróżnionymi jego determinantami.

Przeprowadzona analiza poza określeniem rejonów turystycznych w Polsce oraz oszacowaniem modeli ekonometrycznych krajowego ruchu turystycznego pozwoliła także dokonać porównawczej oceny przydatności metod WAP i analizy czynnikowej w badaniach przestrzennych.

Raport składa się z dwóch części - metodologicznej i empirycznej. W pierwszej części przedstawiono założenia stosowanych metod, zwracając szczególną uwagę na charakterystykę analizy czynnikowej, jako że pozostałe procedury zostały szczegółowo omówione w innej pracy Autorów pt. "Ruch turystyczny i jego uwarunkowania w ujęciu przestrzennym", AE Kraków, Instytut Turystyki, Warszawa 1985. Dlatego też w przypadku tych metod ograniczono się do przytoczenia podstawowych etapów postępowania.

Również i w części empirycznej wykorzystano niektóre elementy i wyniki otrzymane w w/w pracy, jakkolwiek stanowi ona odrębną całość oraz charakteryzuje się odmiennym podejściem do problematyki badań.

Uzyskane wyniki mają charakter wstępny i odnoszą się tylko do jednego roku (1983). W następnym etapie badań podobne analizy zostaną przeprowadzone dla innych punktów czasowych, co będzie stanowiło podstawę oceny zmian zachodzących w kształtowaniu się struktury przestrzennej ruchu turystycznego w Polsce i jego powiązań z wyróżnionymi determinantami.

Część I

METODY ANALIZY STOSOWANE W BADANIACH

1.1. Analiza korelacji

Analiza korelacji stanowi zespół metod i mierników służących do badania zjawiska współzależności między cechami statystycznymi. Chodzi tu o ustalenie, czy między rozpatrywanymi cechami występują związki przyczynowo-skutkowe lub przynajmniej symptomatyczne, a jeżeli tak to jaki jest stopień ich nasilenia (duży, średni, niewielki), charakter (prostoliniowy, krzywoliniowy) oraz kierunek (dodatni, czyli wprost proporcjonalny, czy też ujemny, czyli odwrotnie proporcjonalny).

Wyróżnia się następujące sposoby badania związków:

- porównanie szeregów statystycznych cech,
- budowa i analiza tablic korelacyjnych,
- sporządzanie i analiza wykresu (diagramu) korelacji,
- wyznaczenie i interpretacja mierników analitycznych.

Pierwsze trzy sposoby odgrywają rolę pomocniczą i nie są zbyt rozpowszechnione w praktyce, natomiast ostatni sposób stanowi podstawową metodę analizy korelacji. Spotykane w literaturze współczynniki korelacji można podzielić:

- ze względu na liczbę uwzględnionych jednocześnie cech na:
proste i wielorakie (częstkowe),

- z punktu widzenia rodzaju związku na:
prostoliniowe i krzywoliniowe,
- w zależności od charakteru cech na współczynniki służące do analizy współzależności zmiennych: ilościowych, rangowych i jakościowych.

Dalsze rozważania ograniczone zostaną do współczynników wyznaczonych na podstawie cech o charakterze ilościowym. Współczynniki korelacji dla cech rangowych lub jakościowych w badaniach strukturalnych mają mniejsze pole zastosowań i będą omówione w trakcie ich wykorzystania.

Jednym z najważniejszych mierników w analizie korelacji jest prosty współczynnik korelacji liniowej dla dwóch cech ilościowych X_i oraz X_j dany wzorem:

$$r_{ij} = \frac{\text{cov}(X_i, X_j)}{s_i s_j} \quad (1.1)$$

gdzie $\text{cov}(X_i, X_j)$ jest kowariancją zmiennych:

$$\text{cov}(X_i, X_j) = \frac{1}{n} \sum_{k=1}^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j) \quad (1.2)$$

przy czym $\bar{x}_i, \bar{x}_j$ to średnie arytmetyczne, s_i, s_j - odchylenia standardowe zmiennych X_i, X_j , natomiast n jest liczbą obserwacji.

Współczynnik (1.1) zawiera się w przedziale $\langle -1, 1 \rangle$. Jego znak wskazuje na kierunek korelacji (dodatni lub ujemny), natomiast wartość bezwzględna na stopień nasilenia związku - korelacja jest tym większa im moduł współczynnika korelacji bliższy jest 1, i odwrotnie.

W przypadku, gdy w analizie uwzględnionych jest wiele zmiennych ilościowych X_j ($j=1, \dots, m$) wyznaczane są współczynniki korelacji cząstkowej i wielorakiej, których podstawą jest macierz prostych współczynników korelacji liniowej:

$$R = [r_{ij}], \quad (i, j=1, \dots, m) \quad (1.3)$$

przy czym $r_{ii}=1$, natomiast m jest liczbą cech przyjętych do badań.

Współczynniki korelacji cząstkowej dane są wzorem:

$$r_{ij \cdot 1, 2, \dots, m} = - \frac{R_{ij}}{\sqrt{R_{ii} R_{jj}}}, \quad (i, j=1, \dots, m; i \neq j) \quad (1.4)$$

gdzie R_{ij} , R_{ii} , R_{jj} są dopełnieniami algebraicznymi elementów r_{ij} , r_{ii} , r_{ji} macierzy (1.3). Współczynnik (1.4) charakteryzuje stopień współzależności między zmiennymi X_i , X_j , przy wyeliminowaniu wpływu na te cechy wszystkich pozostałych zmiennych znajdujących się w zbiorze $\{X\}$. Zakres zmienności i sposób interpretacji współczynników cząstkowych jest identyczny jak współczynników korelacji prostej.

Współczynniki korelacji wielorakiej określane są wzorem:

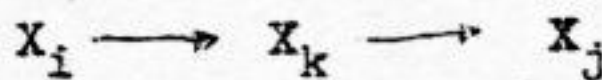
$$r_{i \cdot 1, 2, \dots, m} = \sqrt{1 - \frac{|R|}{R_{ii}}} \quad (i=1, \dots, n) \quad (1.5)$$

gdzie $|R|$ jest wyznacznikiem macierzy R , natomiast R_{ii} dopełnieniem algebraicznym elementu r_{ii} . Współczynnik (1.5) zawiera się w przedziale $\langle 0, 1 \rangle$ i przyjmuje tym większą wartość im większa jest współzależność między zmienną X_i a jednocześnie wszystkimi pozostałymi zmiennymi występującymi w zbiorze

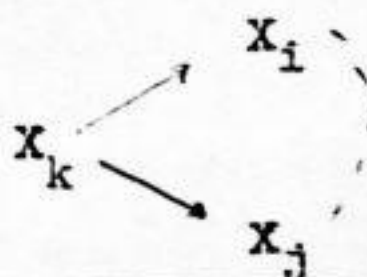
analizowanych cech $\{X\}$.

Omówione dotychczas współczynniki odnoszą się do przypadku analizy związków między cechami (dwoma lub wieloma) o charakterze ilościowym przy założeniu, że zależności mają postać prostoliniową. W wielu sytuacjach np. przy zależności między popytem i dochodami konsumentów, liczbą ludności kraju i czasem, założenie to nie jest prawdziwe. Wykorzystuje się wtedy współczynniki korelacji krzywoliniowej, które definiowane są w oparciu o pojęcia z zakresu analizy regresji i przedstawione zostaną w następnym punkcie opracowania.

Interpretując współczynniki korelacji należy zwrócić uwagę na zjawisko tzw. korelacji pozornej. Wyraża się ono w tym, że niekiedy obserwuje się wyraźne statystyczne skorelowanie cech, podczas gdy merytoryczna analiza zjawisk opisywanych przez te cechy wyklucza występowanie związków przyczynowych między nimi. Przykładami korelacji pozornej może być stwierdzona empirycznie ujemna zależność między stopą urodzeń a liczbą śmiertelnych wypadków drogowych w Europie, dodatnia między liczbą lekarzy a poziomem dochodu narodowego, ujemna między liczbą prawników a poziomem dochodu narodowego, dodatnia między liczbą bocianów a liczbą urodzeń w krajach skandynawskich itd. Przyczyną korelacji pozornej cech X_i , X_j , między którymi nie ma związku przyczynowego jest zazwyczaj trzecia cecha X_k (niekiedy jest nią zmienna czasowa), z którymi obie badane cechy są skorelowane. Występuje wtedy zjawisko przechodniości korelacji powodujące korelację pozorną pośrednią:



lub tzw. quasi-korelację



przy czym w powyższych schematach strzałki wskazują na związki przyczynowe, natomiast linie przerywane na pozorne skorelowanie cech.

1.2. Analiza regresji

Analiza regresji podobnie jak analiza korelacji jest zestawem statystycznych metod służących do badania związków między zjawiskami. Podstawowym narzędziem analizy jest tu funkcja regresji, którą w ogólnej postaci można zapisać następująco:

$$Y = f(X_1, \dots, X_k, \xi) \quad (1.6)$$

gdzie Y jest zmienną objaśnianą (zależną, endogeniczną, predyktatą, regresatą), $X_1, \dots, X_k$ - zbiorem zmiennych objaśniających (niezależnych, egzogenicznych, predyktorów, regresorów), natomiast ξ - składnikiem losowym.

W szczególnym przypadku zbiór zmiennych objaśniających może być jednoelementowy, co prowadzi do regresji dwóch zmiennych (Y, X) .

Zgodnie z równaniem (1.6) zmienna Y stanowi funkcję zmiennych $X_1, \dots, X_k$ oraz składnika losowego, który odzwierciedla wpływ na zmienną objaśnianą wszystkich innych zmiennych objaśniających nie uwzględnionych explicite w równaniu.

W zależności od postaci analitycznej funkcji (1.6) wyróżnia się regresję prostoliniową i krzywoliniową. W przypadku gdy przyjęte zostanie założenie o liniowym charakterze zależności między zmienną Y a zmiennymi zbioru $\{X\}$ to równanie regresji przyjmie następującą postać:

$$Y = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k + \xi \quad (1.7)$$

gdzie α_0 - wyraz wolny równania natomiast α_j ($j=1, \dots, k$) to współczynniki regresji.

Oceny współczynników regresji uzyskuje się najczęściej przy pomocy metody najmniejszych kwadratów zgodnie ze wzorem:

$$a = (X^T X)^{-1} X^T y \quad (1.8)$$

gdzie X macierz realizacji zmiennych objaśniających o wymiarach $(n \times k+1)$, natomiast y - wektor realizacji zmiennej objaśnianej, przy czym pierwsza kolumna macierzy X związana jest z wyrazem wolnym równania (1.7) i zawiera same jedynki.

Współczynnik regresji a_j wskazuje o ile zmieni się wartość zmiennej Y w przypadku jednostkowej zmiany zmiennej X_j , pod warunkiem, że pozostałe zmienne objaśniające będą na ustalonym poziomie.

Oszacowane równanie regresji (1.7) można wykorzystać do celów analitycznych dla wykazania i porównania siły oddziaływań poszczególnych zmiennych objaśniających (traktowanych jako determinanty) na zmienną objaśnianą (wynikową). Innym kierunkiem wykorzystania równania regresji jest prognozowanie poziomu zmiennej objaśnianej Y dla zadanych wartości zmiennych objaśniających:

$$x_0 = [1, x_1^0, \dots, x_k^0] \quad (1.9)$$

Prognozę zmiennej Y dla wektora x_0 uzyskuje się stosując wzór:

$$y_p = x_0 a \quad (1.10)$$

Poza oceną i interpretacją współczynników regresji w trakcie analizy regresji wyznacza się również miary dopasowania równania (1.7) do danych empirycznych. Określają one jaki stopień jest zgodności empirycznych realizacji zmiennej objaśnianej y_i z jej

realizacjami teoretycznymi, wynikającymi z oszacowanego równania regresji:

$$\hat{y} = X a \quad (1.11)$$

Wiarygodność wniosków analitycznych i precyzja prognoz jest tym większa im większa jest dokładność odwzorowania zmiennej Y przez równanie regresji. Stopień zgodności funkcji regresji z danymi empirycznymi zależy od:

- faktu istnienia związków przyczynowo-skutkowych między zmienną objaśnianą a zmiennymi objaśniającymi oraz siły ich natężenia (poziom diagnostyczności zmiennych X_j),
- właściwego doboru wektora zmiennych objaśniających (chodzi o wyeliminowanie zjawiska ich współliniowości, i uwzględnienie w równaniu zmiennych X_j wzajemnie nieskorelowanych, nie powtarzających tych samych informacji,
- umiejętnego ustalenia postaci analitycznej funkcji regresji (nie zawsze występujące zależności mają charakter prostoliniowy - wprost lub odwrotnie proporcjonalnych),
- błędów pomiarowych zmiennych występujących w równaniu regresji,
- rzetelności danych statystycznych i adekwatności zebranych realizacji zmiennych w stosunku do merytorycznego znaczenia zjawisk, które zmienne te powinny odzwierciedlać.

Do podstawowych mierników dopasowania funkcji regresji do danych empirycznych należy:

- średni błąd szacunku (odchylenie standardowe reszt):

$$S_e = \left[\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1} \right]^{1/2} \quad (1.12)$$

gdzie $y_i, \hat{y}_i$ - to zaobserwowane i teoretyczne realizacje

zmiennej Y , natomiast n oraz k - to liczba obserwacji i parametrów równania,

- współczynnik zmienności resztowej:

$$V_e = \frac{S_e}{\bar{y}} \cdot 100 \quad (2.13)$$

gdzie $\bar{y}$ to średnia arytmetyczna zmiennej objaśnianej,

- współczynnik determinacji:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.14)$$

Średni błąd szacunku S_e wskazuje o ile średnio biorąc różnią się między sobą teoretyczne i empiryczne wartości zmiennej Y . Współczynnik V_e wyraża w % wielkość średniego błędu szacunku w stosunku do średniej arytmetycznej Y i stanowi minimalne oszacowanie względnego błędu prognozy na podstawie danego równania regresji. Parametr ten umożliwia porównywanie stopnia dopasowania różnych funkcji regresji, na co nie pozwala miernik S_e , który jest mianowany i wyrażony w tych samych jednostkach pomiarowych co zmienna objaśniana Y .

Współczynnik determinacji R^2 zawiera się w przedziale $\langle 0, 1 \rangle$ i po przemnożeniu przez 100 mówi ile procent całkowitej zmienności zmiennej Y zostało objaśnione przez przyjęty do równania zestaw zmiennych $\{X\}$. Dopełnienie do 1 współczynnika R^2 wskazuje jaką część zmienności zmiennej Y jest efektem oddziaływania zmiennych nieuwzględnionych w równaniu i nosi nazwę współczynnika nieokreśloności ψ^2 (jest to względna ocena

wielkości składnika losowego ϵ_i).

Pierwiastek kwadratowy współczynnika determinacji jest równoważny współczynnikowi korelacji wielorakiej określonego wzorem (1.5), przy założeniu, że zmienna X_i odgrywa rolę zmiennej objaśnianej a pozostałe zmienne zbioru są zmiennymi objaśniającymi. Powiązanie analizy regresji z analizą korelacji ma miejsce także w przypadku współczynników korelacji cząstkowej (1.4) oraz współczynników regresji (1.8). Współczynniki te bowiem dają się wyrazić wzorem:

$$r_{ij \cdot 1, 2, \dots, k} = \sqrt{a_i a_j} \quad (i, j=1, \dots, k; i \neq j) \quad (1.15)$$

Tak więc w przypadku oszacowania krzywoliniowych funkcji regresji wzory (1.14) i (1.15) pozwalają określić odpowiednie krzywoliniowe współczynniki korelacji wielorakiej i cząstkowej.

W analizie poprawności funkcji regresji wykorzystuje się również test Studenta weryfikujący ^{użyty} hipotezę o nieistotnie różnych od zera wartościach ocen parametrów a_j .

W przypadku odrzucenia tej hipotezy można przyjąć, że zmienna X_j w istotny sposób oddziałuje na zmienną Y . W poprawnej funkcji regresji wszystkie lub prawie wszystkie parametry powinny być istotnie różne od zera. Sprawdzianem wysuniętej hipotezy jest statystyka Studenta:

$$t_j = \frac{a_j}{D_j} \quad (j=1, \dots, k) \quad (1.16)$$

gdzie D_j jest średnim błędem szacunku parametru a_j , będącym pierwiastkiem kwadratowym z elementów głównej przekątnej macierzy:

$$D = S_e^2 (X^T X)^{-1} \quad (1.17)$$

Jeżeli wartość statystyki (1.16) przekroczy co do modułu wartość krytyczną t_{α} odczytaną z tablic rozkładu Studenta dla założonego poziomu istotności α oraz $n-k-1$ stopni swobody to można przyjąć, że dany parametr funkcji regresji jest istotnie różny od zera.

W trakcie weryfikacji równania regresji sprawdza się również hipotezę o istotności jednocześnie wszystkich ocen parametrów. Wykorzystuje się w tym celu statystykę:

$$F = \frac{n - k}{k - 1} \frac{R^2}{1 - R^2} \quad (1.18)$$

dla której wartość krytyczną F_{α} odczytuje się z tablic rozkładu Fishera-Snedecora dla poziomu istotności α oraz $k-1$ i $n-k$ stopni swobody. Jeżeli $F > F_{\alpha}$ to można przyjąć, że wszystkie oceny parametrów funkcji regresji są istotnie różne od zera. Warto nadmienić, że w przypadku funkcji regresji dwóch zmiennych $Y = f(X, \xi)$ statystyka F może być także wykorzystana do sprawdzenia czy współczynnik korelacji prostej r_{YX} jest istotnie różny od zera.

Istotnym elementem poprawności równania regresji jest też własność koincydencji, polegająca na zgodności znaków współczynników korelacji prostej r_{YX_j} ze znakiem współczynników regresji cząstkowej a_j . Jeżeli koincydencja znaków tych parametrów nie zachodzi to trudno jest wyciągnąć na ich podstawie jakiekolwiek wnioski merytoryczne. Warunki koincydencji odnoszą się tylko do istotnych ocen parametrów oraz funkcji z więcej niż jedną zmienną objaśniającą.

W przypadku regresji dwóch zmiennych sposób estymacji parametrów równania regresji, miar jego dopasowania do danych empirycznych oraz ich interpretacja nie ulega zmianie, jak

również pozostają w mocy powiązania między współczynnikami korelacji prostej i wielorakiej a odpowiednimi parametrami funkcji regresji.

Jeżeli założenia o liniowym charakterze współzależności nie można przyjąć to wykorzystuje się krzywoliniowe funkcje regresji, z których w praktyce spotyka się funkcje sprowadzalne do postaci liniowej drogą odpowiednich transformacji zmiennych. Należą do nich m.in. funkcje wielomianowe, logarytmiczne, wykładnicze, hiperboliczne, przy czym do najczęściej stosowanych należy funkcja potęgowa:

$$Y = \alpha_0 X_1^{\alpha_1} \dots X_k^{\alpha_k} e^{\varepsilon} \quad (1.19)$$

Po linearyzacji funkcja ta przyjmuje postać:

$$\ln Y = \ln \alpha_0 + \alpha_1 \ln X_1 + \dots + \alpha_k \ln X_k + \varepsilon \quad (1.20)$$

co daje funkcję liniową względem logarytmów zmiennych i pozwala stosować metodę najmniejszych kwadratów dla oszacowania jej parametrów. Algorytm wyznaczania parametrów i miar dopasowania nie ulega tu zmianie, inaczej natomiast interpretuje się oceny współczynników regresji. Są to mianowicie elastyczności zmian zmiennej Y względem poszczególnych zmiennych objaśniających. Parametr α_j wskazuje o ile procent zmieni się Y (wzrośnie dla $\alpha_j > 1$ lub spadnie dla $\alpha_j < 1$) w przypadku 1-procentowego wzrostu poziomu zmiennej X_j , pod warunkiem, że pozostałe zmienne objaśniające nie ulegną zmianie.

Funkcje regresji stanowią zewnętrzną formę modelu ekonometrycznego, którego celem jest opis związków ilościowych między analizowanymi zmiennymi. Z punktu widzenia walorów poznawczych wyróżnia się m.in.:

- modele przyczynowo-skutkowe, gdzie poszczególne równania odzwierciedlają obiektywnie istniejące związki przyczynowo-skutkowe między zmiennymi (zmienna Y wyraża skutek natomiast zmienne X_j traktowane są jako przyczyny),
- modele symptomatyczne, w których między zmiennymi objaśniającymi a zmienną objaśnianą nie ma związku przyczynowego a ma miejsce jedynie silny związek korelacyjny (modele takie wykorzystywane są głównie do celów prognostycznych),
- modele tendencji rozwojowej, gdzie jedyną zmienną objaśniającą jest zmienna czasowa lub jej funkcja (modele te służą do opisu zmian w czasie danej zmiennej wyodrębniając trend, wahania prognostyczne i wahania przypadkowe, prowadząc do prognoz zmiennej Y).

Według sposobu ujęcia czynnika czasu modele ekonometryczne dzieli się na:

- statyczne, oparte na szeregach przekrojowych i uwzględniające zmienne odnoszące się do tego samego okresu czasu,
- dynamiczne, oparte na szeregach czasowych lub przekrojowo-czasowych i uwzględniające w zbiorze zmiennych objaśniających zmienną czasową lub zmienne ^{endogeniczne} ~~ekonomiczne~~ z różnymi opóźnieniami czasowymi.

Ze względu na złożoność powiązań między zmiennymi endogenicznymi rozróżnia się:

- modele proste, w których nie zakłada się związków między różnymi zmiennymi endogenicznymi odnoszącymi się do tego samego okresu (przykładem modelu prostego jest model złożony tylko z jednego równania),
- modele rekurencyjne, gdzie zmienne endogeniczne można ułożyć w łańcuch, którego dany element zależy od zmiennych egzogenicznych, zmiennych opóźnionych w czasie, zmiennych endogenicznych poprzedzających go w łańcuchu i jednocześnie nie zależy

- od zmiennych endogenicznych następujących po sobie,
- modele o równaniach współzależnych, w których przyjmuje się istnienie sprzężeń zwrotnych między nieopóźnionymi zmiennymi endogenicznymi.

1.3. Analiza czynnikowa

Założenia i cele analizy czynnikowej

Analiza czynnikowa stanowi zespół metod i procedur statystycznych pozwalających na sprowadzenie dużej liczby badanych zmiennych do znacznie mniejszej liczby wzajemnie niezależnych czynników lub składowych głównych. Zachowują one stosunkowo dużą część informacji tkwiących w zmiennych pierwotnych, a jednocześnie każdy z nich jest nośnikiem innych treści merytorycznych.

Punktem wyjściowym analizy czynnikowej jest macierz współczynników korelacji między cechami pierwotnymi. W przypadku, gdy w zbiorze tych cech istnieją pewne odrębne grupy zmiennych wzajemnie wysoko skorelowanych to oznacza, że powtarzają one te same informacje. Każda taka grupa może być zastąpiona pojedynczym czynnikiem, będącym syntetycznym nośnikiem informacji zawartych w zmiennych tworzących daną grupę (jest to zmienna wyższego rzędu - metacecha). Analiza czynnikowa dostarcza sposobów i procedur umożliwiających wykrywanie, wyznaczanie i interpretację czynników głównych, odzwierciedlających strukturę powiązań korelacyjnych między rozpatrywanymi cechami, a zarazem leżących u podłoża zaobserwowanych korelacji.

Z formalnego punktu widzenia analiza czynnikowa polega na transformacji wielowymiarowej przestrzeni cech pierwotnych

w nową przestrzeń o mniejszej liczbie wymiarów-czynników poprzez ortogonalizację i rotację układu współrzędnych.

Analiza czynnikowa poza funkcją redundacyjną (zmniejszenie liczby zmiennych charakteryzujących badane zjawiska bez istotnej straty wnoszonych przez nie informacji) może spełniać także funkcje poznawcze, wyrażające się w możliwości wysuwania hipotez dotyczących istnienia i natury ogólnych prawidłowości kształtujących zachodzące związki między zjawiskami, jak również dostarcza narzędzi weryfikacji prawdziwości tych hipotez.

Wyodrębnione czynniki główne mogą być wykorzystane w badaniach m.in. do:

- opisu zjawisk z punktu widzenia nowych kategorii społeczno-ekonomicznych zdefiniowanych przez czynniki główne,
- porządkowania i klasyfikacji obiektów w ortogonalnych przestrzeniach czynnikowych,
- konstruowania ekonometrycznych modeli przyczynowo-opisowych odzwierciedlających związki między zmiennymi endogenicznymi a zbiorem wzajemnie niezależnych czynników w roli determinant,
- ustalenia wag określających znaczenie, jakie należy przypisać poszczególnym zmiennym i czynnikom w trakcie analiz.

Algorytm analizy czynnikowej

Problem analizy czynnikowej można zapisać w postaci następującego układu równań liniowych:

$$\begin{cases} Z_1 = \sum_{l=1}^k w_{1l} F_l + w_1 U_1 \\ \vdots \\ Z_m = \sum_{l=1}^k w_{ml} F_l + w_m U_m \end{cases} \quad (1.21)$$

lub krócej w zapisie macierzowym:

$$Z = WF + U \quad (1.22)$$

gdzie: $Z = [Z_j]^T$, $F = [F_l]^T$; $W = [w_{jl}]$;

$$U = [w_j U_j]^T; \quad (j=1, \dots, m; l=1, \dots, k; k \leq m)$$

przy czym:

- Z_j - j-ta standaryzowana zmienna pierwotna,
- F_l - l-ty czynnik wspólny (główny),
- U_j - j-ty czynnik swoisty (specyficzny),
- w_{jl} - ładunek l-tego czynnika głównego w j-tej zmiennej,
- w_j - ładunek j-tego czynnika specyficznego,
- m, k - liczba zmiennych pierwotnych i czynników głównych,

pod warunkiem, że czynniki główne i swoiste są zestandaryzowane i wzajemnie ortogonalne.

Celem analizy czynnikowej jest wyrażenie realizacji standaryzowanych zmiennych pierwotnych:

$$z_{pj} = \frac{x_{pj} - \bar{x}_j}{s_j} \quad (p=1, \dots, n; j=1, \dots, m) \quad (1.23)$$

gdzie x_{pj} - to wartość p-tej obserwacji j-tej zmiennej pierwotnej, $\bar{x}_j$, s_j - to średnia arytmetyczna i odchylenie standardowe j-tej zmiennej, n - liczba obserwacji; w formie liniowej kombinacji k czynników wspólnych oraz m czynników swoistych ze współczynnikami w_{j1} oraz w_j . Rozwiązanie analizy czynnikowej polega na wyznaczeniu ładunków czynnikowych w_{j1} oraz w_j na podstawie wartości zmiennych pierwotnych x_{pj} .

Punktem wyjścia obliczeń jest macierz korelacji:

$$R = [r_{ij}] \quad (i, j=1, \dots, m) \quad (1.24)$$

gdzie:

$$r_{ij} = \frac{1}{n} Z'Z = \frac{1}{n} \sum_{p=1}^n z_{pi} z_{pj} \quad (1.25)$$

Elementy głównej przekątnej macierzy korelacji interpretuje się jako współczynniki korelacji danej zmiennej z nią samą lub jako całkowitą wariancję tej zmiennej (zmienne po standaryzacji mają wariancję równą jedności).

Wariancję całkowitą j-tej zmiennej można podzielić na dwie składowe, z których jedna wyraża tzw. zasób zmienności wspólnej h_j^2 , będący tą częścią wariancji, która jest objaśniana przez czynniki główne:

$$h_j^2 = \sum_{l=1}^k w_{jl}^2 \quad (1.26)$$

natomiast druga składowa w_j^2 nazywana jest "swoistością" i stanowi pozostałość po odjęciu zasobu zmienności wspólnej od wariancji całkowitej

$$w_j^2 = 1 - h_j^2 \quad (1.27)$$

W analizie czynnikowej chodzi o takie ustalenie parametrów w_{j1} oraz w_j , aby maksymalnie wyeliminować wpływ czynników swoistych (w_j) na rzecz czynników głównych (w_{j1}). Uzyskuje się to poprzez zastąpienie w macierzy R elementów głównej przekątnej (całkowitych wariancji zmiennych) zasobami zmienności wspólnej, co prowadzi do tzw. zredukowanej macierzy korelacji:

$$\tilde{R} = \begin{cases} \tilde{r}_{ij} = r_{ij} & \text{dla } i \neq j \\ \tilde{r}_{ij} = h_j^2 & \text{dla } i=j \end{cases} \quad (1.28)$$

Parametry h_j^2 szacuje się przy pomocy jednej z następujących metod:

- jako maksymalny współczynnik korelacji danej zmiennej z pozostałymi:

$$h_j^2 = \max_i \{ r_{ij} \} \quad (i \neq j) \quad (1.29)$$

- jako średnią arytmetyczną współczynników korelacji danej zmiennej z pozostałymi:

$$h_j^2 = \frac{1}{m-1} \sum_{\substack{i=1 \\ i \neq j}}^m r_{ij} \quad (1.30)$$

- według wzoru:

$$h_j^2 = \frac{r_{ij} r_{sj}}{r_{is}} \quad (1.31)$$

gdzie r_{ij} , r_{sj} to pierwsze dwa największe co do wielkości współczynniki korelacji j -tej zmiennej z pozostałymi.

Algorytm analizy czynnikowej można ująć w trzy etapy (por. rys.1):

- określenie ładunków czynnikowych w_{jl} ($j=1, \dots, m$; $l=1, \dots, k$),
- rotacja czynników i ich interpretacja,
- obliczanie realizacji czynników głównych F_l ($l=1, \dots, k$).

Procedura wyznaczania ładunków czynnikowych składa się z powtarzających się kolejno kroków, przy czym w każdym z nich otrzymuje się ładunki jednego czynnika głównego.

W pierwszym kroku, wychodząc ze zredukowanej macierzy korelacji $\tilde{R}$, ustala się ładunki pierwszego czynnika F_1 w drodze maksymalizacji udziału tego czynnika w wariancji całego zbioru cech, tzn. maksymalizując funkcję:

$$V_1 = \sum_{j=1}^k w_{j1} \quad (1.32)$$

przy pomocy metody mnożników Lagrange'a, przy założeniu, że

$$\tilde{R} = W W^T \quad (1.33)$$

Wyznaczone wartości ładunków pierwszego czynnika:

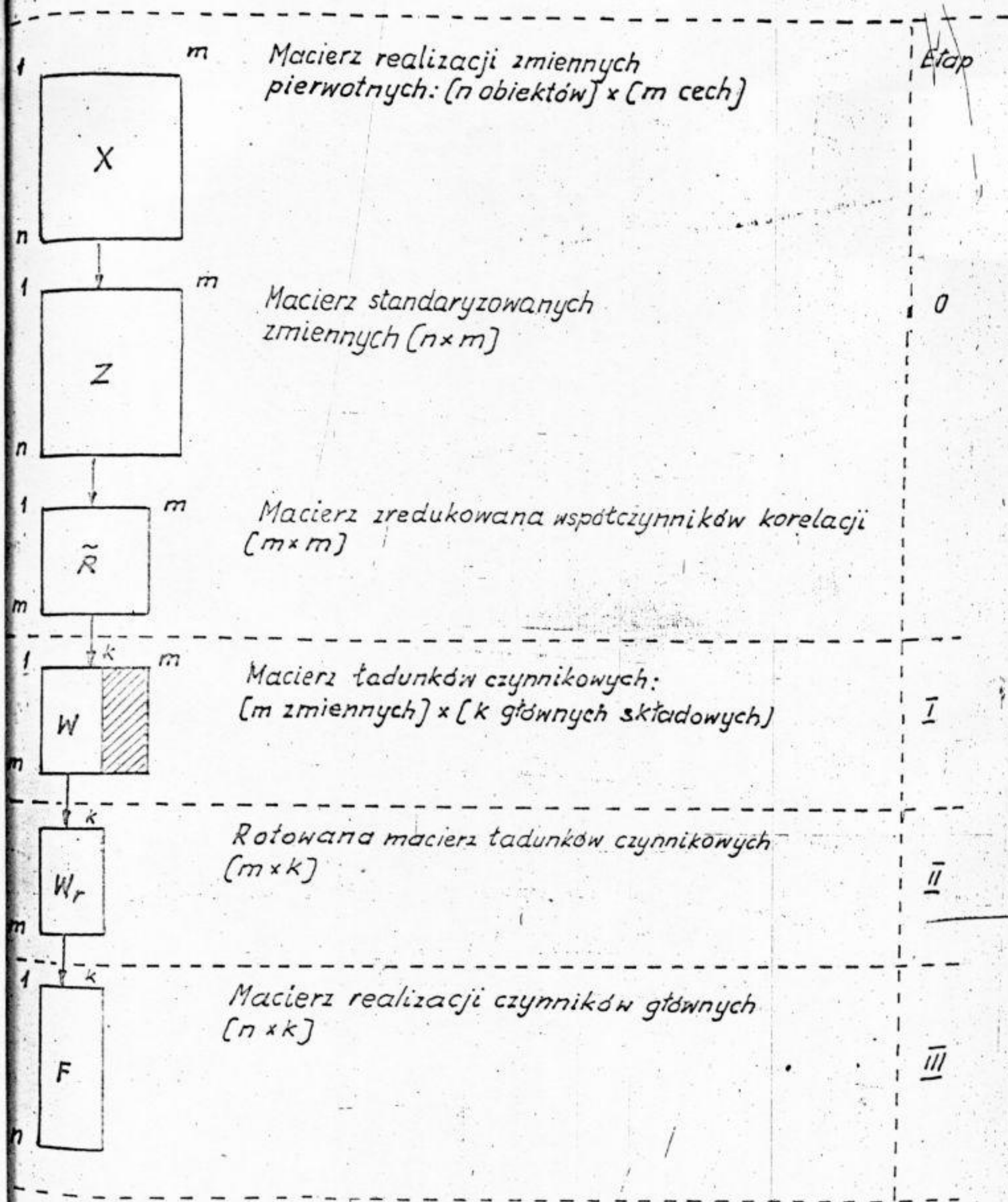
$$w_1 = [w_{j1}] \quad (1.34)$$

która po podstawieniu za $\tilde{R}$ stanowi podstawę do ustalenia w analogiczny sposób ładunków drugiego czynnika głównego F_2 .

Operacje te powtarza się dla dalszych czynników głównych, aż do osiągnięcia żadanego, łącznego stopnia wyjaśniania przez nie wariancji zmiennych (np. 90%), względnie do momentu, gdy udział danego czynnika nie spadnie poniżej ustalonego a priori

służą do określenia macierzy pozostałości korelacyjnej

$$\tilde{R}_1 = \tilde{R} - w_1 w_1^T \quad (1.35)$$



1. Etapy analizy czynnikowej

poziomu (np. 5%)¹.

Poszukiwanie warunkowego ekstremum funkcji typu (1.32) przy pomocy metody mnożników Lagrange'a sprowadza się do wyznaczenia wszystkich niezerowych i uporządkowanych malejąco wartości własnych macierzy $\tilde{R}$

$$\lambda = [\lambda_1] \quad (l=1, \dots, k) \quad (1.36)$$

oraz odpowiadających im wektorów własnych:

$$A_1 = [a_{j1}] \quad (j=1, \dots, m) \quad (1.37)$$

Ładunki czynnikowe l-tego czynnika głównego w j-tej zmiennej oblicza się ze wzoru:

$$w_{j1} = \sqrt{\lambda_1} \frac{a_{j1}}{\left[\sum_{j=1}^m a_{j1}^2 \right]^{1/2}} \quad (1.38)$$

i interpretuje jako współczynniki korelacji między zmienną Z_j oraz czynnikiem F_1 .

Wartości własne λ_1 stanowią podstawę oceny zasobu zmienności całkowitej wyczerpywanej przez kolejne czynniki główne. Zasoby te w wyrażeniu względnym dane są wzorem:

$$h_j = \frac{\lambda_j}{\sum_{j=1}^m \lambda_j} 100 = \frac{\lambda_j}{m} 100 \quad (1.39)$$

¹Zgodnie z algorytmem postępowania udziały kolejnych czynników głównych w wyjaśnieniu wariancji zmiennych są coraz to mniejsze.

Suma k kolejnych zasobów:

$$H_k = \sum_{j=1}^k \hat{h}_j \quad (1.40)$$

określa (w %) udział zmienności całkowitej wyczerpany przez k pierwszych czynników głównych.

Macierz ładunków czynnikowych pozwala na merytoryczną identyfikację wyodrębnionych czynników głównych. Czynnik F_1 określony jest przez zestaw tych zmiennych pierwotnych, dla których obserwuje się szczególnie wysokie, co do modułu wartości ładunków w_{j1} ($j=1, \dots, m$). Nie zawsze jednak ~~uzyteczna~~ w wyniku przytoczonej poprzednio procedury udaje się wyodrębnić rozłączne grupy zmiennych "sprzężonych" z poszczególnymi czynnikami głównymi. Często zdarza się bowiem, że te same zmienne są wysoko skorelowane z różnymi czynnikami, co utrudnia ich jednoznaczną identyfikację. W tym przypadku zachodzi konieczność tzw. rotacji czynników, stanowiącej drugi etap analizy czynnikowej.

Podstawą rotacji jest fakt, że warunek (1.33) spełnia nie tylko dana macierz ładunków W , ale także inna macierz W_1 określona wzorem:

$$W_1^T = B W^T \quad (1.41)$$

gdzie B jest dowolną macierzą ortogonalną². Okoliczność ta

²Elementy tej macierzy określają wielkości kątów, o jakie należy obrócić układ ortogonalnych osi przestrzeni czynników głównych z położenia zadanego macierzą W do położenia W_1 .

stanowi o niejednoznaczności analizy czynnikowej wyrażającej się w tym, że tę samą macierz korelacji można wygenerować przy pomocy nieskończenie wielu różnych macierzy ładunków. Inaczej mówiąc, analiza czynnikowa pozwala jedynie na określenie przestrzeni czynników, w której można wybierać różne zbiory ortogonalnych osi.

Celem rotacji jest wybór takiego położenia osi przestrzeni czynników głównych, aby odpowiadająca mu macierz ładunków zezwalała na stosunkowo łatwą interpretację czynników. Macierz ta powinna spełniać kryteria tzw. prostej struktury sformułowane przez L.L.Thurstone'a, do których m.in. zalicza się:

- co najmniej jeden zerowy ładunek³ w każdym wierszu macierzy W,
- co najmniej k zerowych ładunków w każdej kolumnie macierzy W, to znaczy k^2 w całej macierzy lub $(k/m)100\%$ ogólnej liczby możliwych ładunków,
- obecność takich grup cech, aby dla dowolnej pary czynników zero⁴ ładunki jednego czynnika odpowiadały niezerowym ładunkom drugiego.

Powyższe kryteria sprowadzają się do tego, aby w zbiorze cech można było wyróżnić rozłączne grupy, z których każda zawierałaby cechy o wysokich ładunkach dla jednego czynnika, średnich - dla kilku innych czynników oraz zerowych - dla pozostałych czynników. W praktyce rzadko można doprowadzić do struktury czynnikowej spełniającej kryteria struktury prostej. Doprowadziło to do koncepcji rotacji nieortogonalnych, w których osie czynni-

³Ładunki czynnikowe jako oceny współczynników korelacji mogą być testowane z punktu widzenia istotności ich różnic względem wartości 0. Za ładunek zerowy uważa się taki, którego wartość nie przekracza co do modułu wartości krytycznej r_{α} odczytywanej z tablic rozkładu Studenta dla zadanego poziomu istotności α oraz liczby stopni swobody $n-2$.

kowe po rotacji mogą tworzyć układy ukośnokatne. Rozwiązanie takie jest jednak krytykowane z uwagi na jego subiektywność, jeszcze bardziej kłopotliwą interpretację czynników oraz złożoną formalizację matematyczną. Dlatego też analiza czynnikowa rozwinięła się w kierunku poszukiwań algorytmów prowadzących do ortogonalnych struktur czynnikowych najbardziej zbliżonych do struktury prostej.

Do najczęściej stosowanych procedur w tym zakresie⁴ należy algorytm VARIMX opracowany przez H.F.Kaysera w roku 1958. W metodzie tej szuka się takiej macierzy transformacji B, aby zmaksymalizować liczbę zerowych ładunków w każdej kolumnie macierzy czynników oraz jednocześnie zmaksymalizować korelacje między minimalną liczbą zmiennych a każdym z wyodrębnionych czynników. W metodzie tej (jak zresztą i w każdej innej metodzie rotacji osi) zasób zmienności całkowitej wyczerpywany przez wszystkie czynniki H_k pozostaje bez zmiany, ale zasoby zmienności poszczególnych czynników $\hat{h}_1$ oraz związane z nimi ładunki różnią się w stosunku do struktury wyjściowej.

W trzecim etapie algorytmu analizy czynnikowej wyznacza się realizacje czynników głównych korzystając ze wzoru macierzowego:

$$F = W^T Z \quad (1.42)$$

wykorzystując warunek, jaki spełnia macierz ładunków czynnikowych W:

$$W^T W = I \quad (1.43)$$

gdzie I jest macierzą jednostkową. Macierz F o wymiarach $n \times k$ zawiera H_k procent informacji tkwiącej w wyjściowej macierzy

⁴Inne, rzadziej wykorzystywane metody tzw. Quartimax'owe podali: Ferguson, Neuhaus, Wrigley oraz Saunders.

danych i składa się ze wzajemnie ortogonalnych czynników głównych. Stanowi ona zwykle punkt wyjścia dalszej analizy statystycznej, (opis, ekonometryczne modele przyczynowo-skutkowe, grupowanie i porządkowanie metodami taksonomicznymi itd.), w której poszczególne obiekty scharakteryzowane są przez relatywnie niewielką liczbę czynników, przy czym każdy z nich wnosi do rozwiązań różny i nie powtarzający się zakres informacji. Równocześnie parametry $\hat{h}_l$ ($l=1, \dots, k$) wyznaczają wagi określające znaczenie, jakie należy przykładać w toku analizy poszczególnym czynnikom głównym.

Metoda głównych składowych

Analiza głównych składowych opracowana przez H. Hotellinga w 1933 roku jest metodą transformacji zmiennych pierwotnych we względnie ortogonalne, nowe zmienne tzw. główne składowe. Cele tej metody są więc zbieżne z analizą czynnikową jakkolwiek między nimi istnieje szereg różnic metodologicznych⁵.

U podstaw analizy czynnikowej tkwi założenie o podziale wariancji całkowitej na dwie części: wariancję wspólną i specyficzną. Natomiast analiza głównych składowych nie uwzględnia struktury wariancji. Czynniki główne w analizie czynnikowej generują w zasadzie kowariancję między zmiennymi pierwotnymi (są to bowiem po ich standaryzacji współczynniki korelacji),

⁵Niekiedy termin "analiza czynnikowa" używany jest w ogólniejszym znaczeniu, obejmującym zarówno metodę głównych składowych, omówioną poprzednio analizę czynnikową właściwą oraz inne procedury, pominięte w opracowaniu z uwagi na ich malejący zakres zastosowań np.: metodę dwuczynnikową Holzingera, czynników dwubiegunowych Burta, centroidalną Thurstone'a. W pracy pod pojęciem analizy czynnikowej rozumie się tylko analizę czynnikową w węższym znaczeniu, zgodną z algorytmem opisanym w poprzednim podrozdziale.

podczas, gdy czynniki swoiste wyjaśniają wariancję właściwą dla danej zmiennej. Analiza czynnikowa jest więc analizą struktury macierzy kowariancji z zasobami wariancji wspólnej na głównej przekątnej. W przypadku analizy głównych składowych punktem wyjścia nie jest zredukowana macierz korelacji, lecz pierwotna macierz korelacji z jedynkami na głównej przekątnej.

Analiza głównych składowych jest przede wszystkim procedurą ortogonalizacji zmiennych ułatwiającą budowę teoretycznego modelu opisującego strukturę i związki między badanymi zjawiskami. Analiza czynnikowa stanowi sama w sobie taki model posługując się jedynie ortogonalizacją zmiennych pierwotnych. Widoczne to jest zwłaszcza przy porównaniu modelu regresji wielorakiej z analizą czynnikową, gdzie ma miejsce szereg podobieństw. Czynniki główne odpowiadają zmiennym niezależnym (objaśniającym), ładunki czynnikowe - współczynnikom regresji, czynniki swoiste - resztom regresji, zasób zmienności wspólnej - współczynniki determinacji wielorakiej itd.

Podstawowe równanie analizy głównych składowych ma następującą postać macierzową:

$$Z = WF \quad (1.44)$$

i różni się od równania analizy czynnikowej (Y) ^{1.22} tylko brakiem czynników swoistych U . W konsekwencji liczba głównych składowych równa jest liczbie zmiennych pierwotnych m , podczas gdy na ogół liczba czynników głównych $k < m$ ⁶. W praktyce nie ma to jednak

⁶Wielkości k oraz m powinny ponadto być większe od liczby obserwacji (obiektów) n . W przeciwnym przypadku zarówno k jak i m równe są $n-1$.

dużego znaczenia - gdyż zarówno liczbę czynników jak i składowych głównych ogranicza się w rozważaniach do kilku najważniejszych i jednocześnie nie sprawiających trudności w interpretacji.

Porównanie charakterystycznych cech analizy czynnikowej i metody głównych składowych zebrano w tabl.1.1. Obok dotychczas omówionych, podstawowych różnic między rozważanymi procedurami w tablicy tej przytoczono również inne, o mniejszym znaczeniu poznawczym i praktycznym.

Algorytm określania głównych składowych jest analogiczny jak w analizie czynnikowej, z tą różnicą że macierzą wyjściową jest pierwotna macierz korelacji. Nie ulega zmianie także sposób wyznaczania i interpretacji macierzy ładunków oraz zasobów informacji zawartych w poszczególnych składowych. W metodzie głównych składowych zazwyczaj pomijany jest problem rotacji, zwłaszcza w przypadku gdy wykorzystuje się tę metodę wyłącznie do celów ortogonalizacji i redukcji, zbiorów zmiennych wyjściowych.

Interpretacja geometryczna analizy czynnikowej

W interpretacji geometrycznej zagadnienia analizy czynnikowej poszczególne zmienne i czynniki traktuje się jako wektory. Macierz współczynników korelacji R wyznacza tzw. "konfigurację" wektorów - zmiennych. Długości tych wektorów równe są jedności (zmienne po standaryzacji mają wariancję równą 1) i są reprezentowane przez elementy głównej przekątnej macierzy $R(r_{jj}=1)$. Poszczególne współczynniki korelacji równe są natomiast cosinusowi kąta zawartego między daną parą wektorów - zmiennych:

Tablica 1.1

Porównanie analizy czynnikowej z metodą głównych składowych

Analiza czynnikowa	Metoda głównych składowych
- Ch.Spearman (1904), L.L.Thurstone (1913)	- H.Hotteliny(1933)
- orientacja kowariancyjna: punktem wyjścia jest zredukowana macierz korelacji	- orientacja wariancyjna: punktem wyjścia jest zwykła macierz korelacji
- <u>model otwarty</u> : obok wariancji cech uwzględniona jest także wariancja nieobjaśniania, wynikająca z cech pominiętych w badaniu oraz tej części zmienności, która jest wywołana losowością obserwacji	- <u>model zamknięty</u> : uwzględniana jest wyłącznie wariancja badanych cech
- zmienne pierwotne są liniową funkcją nieobserwowalnych czynników głównych oraz pojedynczej, utajonej zmiennej specyficznej	- zmienne pierwotne są liniową funkcją nieobserwowalnych składowych głównych
- buduje się model teoretyczny zjawiska i sprawdza czy jest on zgodny z danymi empirycznymi	- wychodzi się od obserwacji empirycznych, a następnie buduje teoretyczny model zjawiska
- zmienne pierwotne muszą być losowe	- zmienne pierwotne nie muszą być losowe
- zmienne powinny mieć rozkład normalny lub być doprowadzone do takiej postaci drogą odpowiednich transformacji	- normalność zmiennych nie jest konieczna
- równanie podstawowe: $Z = WF + U$	- równanie podstawowe: $Z = WF$

$$r_{ij} = \cos \alpha \quad (i, j=1, \dots, m; i \neq j) \quad (1.45)$$

Czynniki główne są interpretowane również jako wektory jednostkowe (czynniki są standaryzowane z wariancją równą 1), lecz w odróżnieniu od poprzedniej konfiguracji są one wzajemnie ortogonalne (przecinają się pod kątem prostym 90°). Konfiguracja czynników głównych tworzy więc osie prostokątnego układu współrzędnych.

Ładunki czynnikowe można zinterpretować jako rzuty wektorów-zmiennych na osie-czynniki. Ładunek czynnikowy w_{j1} jest bowiem równy współczynnikowi korelacji między zmienną X_j oraz czynnikiem F_1 , tzn. równy jest cosinusowi kąta zawartego między j -tym wektorem-zmienną oraz 1-tą osią-czynnikiem.

Istota analizy czynnikowej wyraża się więc we wzajemnym usytuowaniu dwóch konfiguracji wektorów: nieortogonalnego-zmiennych i ortogonalnego-czynników oraz w wyznaczaniu rzutów zmiennych na osie prostokątnego układu współrzędnych. Ponieważ wiele zmiennych jest wzajemnie silnie skorelowanych, odpowiadające im wektory położone są blisko siebie, dzięki czemu można je zastąpić jednym reprezentującym je czynnikiem głównym. W rezultacie liczba tych czynników jest znacznie mniejsza niż zmiennych pierwotnych.

Główną wadą analizy czynnikowej jest brak jej jednoznaczności, wynikający z możliwości dowolnego usytuowania w przestrzeni wielowymiarowej układu odniesienia, jakim jest układ osi-czynników. Przy danym położeniu konfiguracji wektorów-zmiennych wartości ładunków czynnikowych (rzutów tych wektorów na osie) zależą do umiejscowienia układu odniesienia. Sytuację taką ilustruje rys. 2, gdzie łatwo można zauważyć zmiany w wielkości

rzutów wektora OX_1 na poszczególne osie F_1, F_2 w zależności od ich położenia.

Problem rotacji sprowadza się do obrotu układu osi-czynników dokoła początku tego układu, w takim kierunku i stopniu, aby uzyskać maksymalnie prostą strukturę macierzy ładunków (tzn. taką, gdy zmienne pierwotne położone są możliwie najbliżej poszczególnych osi). Rotacja nieortogonalna polega na wyborze i umiejscowieniu w przestrzeni niekoniecznie prostopadłego układu osi. Przyjęcie ukośnokątnych osi czynnikowych jakkolwiek może poprawić strukturę ładunków, to jednak obarczone jest subiektywizmem i nie jest zalecane do stosowania.

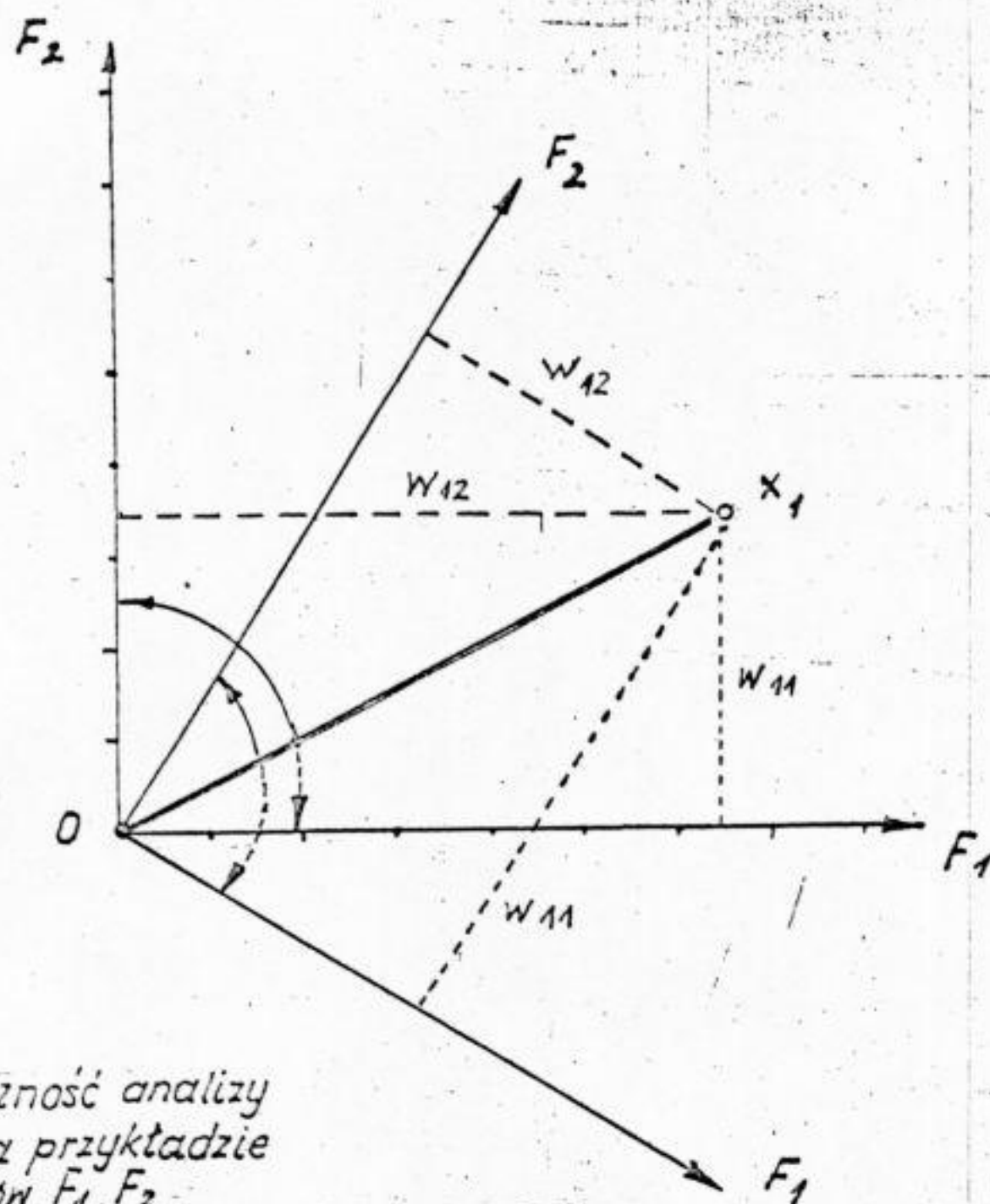
Powyższe rozważania zebrano w tablicy 1.2 zawierającej wybrane elementy interpretacji geometrycznej analizy czynnikowej.

Techniki analizy czynnikowej

W dotychczasowych rozważaniach przyjmowano założenie, że punktem wyjścia analizy czynnikowej jest macierz danych $X(n,m)$ typu obiekty x cechy. Rozpatrywane obiekty traktowane tu są jako zbiór punktów w przestrzeni wielowymiarowej wyznaczonej przez uwzględnione w badaniach cechy. Jest to rzeczywiście przypadek najczęściej spotykany w praktycznych zastosowaniach. Tym niemniej istota punktów oraz współrzędnych przestrzeni może być różna, co prowadzi do innych technik analizy czynnikowej. Związane to jest głównie z wprowadzeniem do badań ujęcia dynamicznego. Ogólnie biorąc podstawą analizy czynnikowej może być trójwymiarowa kostka danych:

$$X = [x_{ijt}] \quad (i=1, \dots, n; \quad j=1, \dots, m; \quad t=1, \dots, k) \quad (1.46)$$

gdzie n, m, k to odpowiednia liczba obiektów, cech oraz okresów czasu.



Rys. 2.

Niejednoznaczność analizy
czynnikowej na przykładzie
dwóch czynników F_1, F_2
oraz zmiennej X_1

Tablica 1.2

Elementy interpretacji geometrycznej analizy czynnikowej

Wyszczególnienie	Interpretacja
Macierz współczynników korelacji R	
- elementy diagonalne $r_{jj}=1$	długości wektorów-zmiennych pierwotnych (wariancje standaryzowanych zmiennych=1)
- elementy pozadiagonalne $r_{ij} = \cos \alpha$	cosinusy kątów zawartych między poszczególnymi wektorami-zmiennymi
Zredukowana macierz współczynników korelacji R	
- elementy diagonalne $\sqrt{h_j^2} \leq 1$	część długości wektorów-zmiennych pierwotnych (zasoby zmienności wspólnej)
- elementy pozadiagonalne $r_{ij} = \cos \alpha$	jak dla macierzy R
Czynniki główne F_1	- wektory o jednostkowej długości wzajemnie ortogonalne - osie prostokątnego układu współrzędnych (czynniki są standaryzowane z wartością = 1)
Ładunki czynnikowe w_{j1} $w_{j1} = r_{X_j F_1} = \cos \beta$	- rzuty wektorów-zmiennych na osie-czynniki - cosinusy kątów zawartych między zmienną X_j oraz czynnikiem F_1
Rotacja	obróć układu osi-czynników dokoła jego początku
Rotacja nieortogonalna	wybór i umiejscowienie ukośnokątnego układu osi-czynników

W zależności od sposobu podejścia do analizy czynnikowej w oparciu o macierze typu (1.46) lub jej przekroje wyróżnić można 5 rodzajów technik:

- cząstkowe,
- łączone,
- warstwowe,
- złożone,
- kompleksowe.

Do najprostszych, a zarazem najczęściej stosowanych należą techniki cząstkowe, których przegląd podany jest w tablicy 1.3. Jak można zauważyć punktem wyjścia w każdej z wymienionych technik jest inna warstwa kostki danych (1.46). Poszczególne przekroje tej kostki: obiekty, cechy i okresy mogą zamiennie wyznaczać ¹⁰²⁴ ~~obiekty~~ punktów, obrazy wymiarów przestrzeni lub też stanowić element stały w trakcie analizy czynnikowej. W każdym przypadku, inna jest istota i interpretacja wyodrębnionych czynników. Dla technik R oraz P są one metacechami, dla technik Q oraz S - grupami podobnych obiektów, natomiast dla technik Q oraz T - grupami podobnych okresów. W ramach poszczególnych par technik uzyskuje się typologię (cech, obiektów czy okresów) o różnej interpretacji merytorycznej w zależności od tego, jaki przekrój kostki danych określa obrazy punktów.

Warto zauważyć, że w technikach Q oraz O nie wykorzystuje się macierzy współczynników korelacji (których nie można wyznaczyć z uwagi na brak sumowalności szeregów składających się z realizacji różnoimiennych cech), lecz macierze mierników podobieństwa lub odległości - np. Euklidesa, miejskiej, Canberra itd. - między obiektami względnie okresami. Powoduje to konieczność pewnych modyfikacji w algorytmie numerycznym analizy czynnikowej nie zmieniając jednak go w sposób zasadniczy, jako że współ-

Charakterystyka cząstkowych technik analizy czynnikowej

Struktura danych	Element stały	Obrazy		Technika a.cz.	Podstawa obliczeń	Cel analizy czynnikowej
		punktów	wymiarów			
obiekty x cechy $X(n,m)$	okres	obiekty	cechy	R	macierz korelacji cech na podstawie szeregów przekrojowych $R(m,m)$	redukcja zbioru cech
cechy x obiekty $X(m,n)$	okres	cechy	obiekty	Q	macierz podobieństwa obiektów $P(n,n)$	typologia obiektów o podobnych relacjach wartości cech
okresy x cechy $X(k,m)$	obiekt	okresy	cechy	P	macierz korelacji cech na podstawie szeregów czasowych $R(m,m)$	redukcja zbioru cech
cechy x okresy $X(m,k)$	obiekt	cechy	okresy	O	macierz podobieństwa okresów $P(k,k)$	typologia okresów o podobnych relacjach wartości cech (periodyzacja)
obiekty x okresy $X(n,k)$	cecha	obiekty	okresy	T	macierz korelacji okresów na podstawie szeregów przekrojowych $R(k,k)$	ustalenie istotnych okresów czasu
okresy x obiekty $X(k,n)$	cecha	okresy	obiekty	S	macierz korelacji obiektów na podstawie szeregów czasowych $R(n,n)$	typologia obiektów o podobnych zmianach w czasie z p.w. danej cechy

czynniki korelacji stanowią przekształconą formę mierników podobieństwa.

Jak już wspomniano typową techniką analizy czynnikowej jest technika R. W literaturze spotkać można także szereg zastosowań techniki Q. Natomiast pozostałe techniki analizy czynnikowej wykorzystywane są w badaniach przestrzennych stosunkowo rzadko.

Techniki łączone nie stanowią w zasadzie odmiennego rodzaju analizy czynnikowej a sprowadzają się jedynie do krokowej realizacji par technik cząstkowych typu: $R \rightarrow Q$, $P \rightarrow O$, $T \rightarrow S$. W pierwszym kroku dokonuje się redukcji zbioru cech (R,P) lub okresów (T), a w drugim przeprowadza typologię obiektów (Q,S) lub okresów (O) na podstawie wyodrębnionych poprzednio czynników głównych.

W technikach warstwowych analizy czynnikowej (określanych też mianem technik M) rozpatruje się ciągi wyników analizy przy pomocy technik cząstkowych danego typu dla różnych elementów stałych - np. typologia obiektów z punktu widzenia tego samego zestawu cech techniką Q dla różnych okresów czasu. W drugim etapie badań dokonuje się analizy porównawczej uzyskanych wyników. Należy jednak zaznaczyć, że stosowanie technik warstwowych nie zawsze jest uzasadnione, zwłaszcza gdy natura czynników zmienia się w zależności od warstwy (np. w czasie), co prowadzi do braku porównywalności otrzymywanych serii rozwiązań czynnikowych. Tym niemniej technika M jest stosunkowo często wykorzystywana np. w geografii miast, czy w badaniach nad rozwojem ekonomicznym. Biorąc pod uwagę zgłaszane zastrzeżenia co do zasadności omawianej techniki analizy czynnikowej opracowana została specjalna metoda rotacji przydatna w studiach

porównawczych⁷.

Podstawą technik złożonych analizy czynnikowej są dwuwymiarowe macierze danych uzyskane z kostki trójwymiarowej $X(n,m,k)$ w drodze połączenia dwóch jej przekrojów w jeden. Wykaz możliwych technik złożonych podany jest w tablicy 1.4. Ich zaletą w porównaniu z technikami cząstkowymi i warstwowymi jest ujmowanie danego aspektu badanych zjawisk w powiązaniu z jednocześnie dwoma pozostałymi. Pozwala to na znaczne wzbogacenie analizy merytorycznej i umożliwia porównywanie otrzymywanych układów czynnikowych zarówno w czasie jak i przestrzeni. Wadę technik złożonych stanowią duże rozmiary macierzy korelacji lub podobieństwa w technikach RT, PS, QO. Należy zauważyć, że praktyczne wykorzystanie omawianych technik jest znikome.

Najbardziej ogólne podejście reprezentuje technika kompleksowa, w której analizie czynnikowej podlega cała trójwymiarowa kostka danych. Punktem wyjścia obliczeń jest tu łączna macierz kowariancji dla zbiorów zmiennych w przestrzeni i w czasie. W rezultacie uzyskuje się trójkierunkową macierz ładunków czynnikowych $W(n',m',k')$ określających istotne cechy, obiekty i okresy, przy czym elementy rdzeniowe tej macierzy wyznaczają siłę zależności między wyodrębnionymi czynnikami a poszczególnymi składowymi każdego z trzech przekrojów wyjściowej kostki danych $X(n,m,k)$. Technika kompleksowa analizy czynnikowej stanowi podstawę budowy modelu czasoprzestrzennego, którego zarówno założenia

⁷Por. Schilderink J.H.P., Factor Analysis to Developed and Developing Countries, Studies on Economics 1, Tilburg Institute of Economics, Groningen 1970.

Charakterystyka złożonych technik analizy czynnikowej

Struktura danych	Obrazy		Podstawa obliczeń	Cel analizy czynnikowej	Technika
	punktów	wymiarów			
obiekty x cechookresy $X(n, mxk)$	obiekty	cechy w różnych okresach	mierniki korelacji cech i okresów na podstawie szeregów przekrojowych $R(mxk, mxk)$	jednoczesna redukcja cech i okresów	RT
cechookresy x obiekty $X(mxk, n)$	cechy w różnych okresach	obiekty	macierz podobieństwa obiektów $P(n, n)$	typologia obiektów z p.w. podobieństwa cech w różnych obrazach	QS
okresy x cechoobiekty $X(k, nxm)$	okresy	cechy dla różnych obiektów	macierz korelacji cech i obiektów na podstawie szeregów czasowych $R(nxm; nxm)$	redukcja cech wraz z typologią obiektów	PS
cechoobiekty x okresy $X(nxm, k)$	cechy dla różnych obiektów	okresy	macierz podobieństwa okresów $P(k, k)$	typologia okresów o podobnych relacjach wartości cech w różnych okresach	OT
obiektookresy x cechy $X(nxk, m)$	obiekty w różnych okresach	cechy	macierz korelacji cech na podstawie szeregów przekrojowo-czasowych $R(m, n)$	redukcja cech	RP
cechy x obiektookresy $X(m, nxk)$	cechy	obiekty w różnych okresach	macierz podobieństwa obiektów i okresów $P(nxk; nxk)$	typologia obiektów i okresów (periodyzacja łączna)	QO

teoretyczne jak i praktyczne próby konstrukcji można spotkać w literaturze przedmiotu⁸.

1.4. Wielowymiarowa analiza porównawcza

Statystyczna analiza porównawcza (SAP) zwana też wielowymiarową analizą porównawczą (WAP) jest stosunkowo nowym działem statystyki, znajdującym zastosowanie w szeregu dyscyplinach nauki - ekonomii, socjologii, psychologii, demografii, biologii, nauk rolniczych itd., przy czym w zakresie dyscyplin ekonomicznych WAP wykorzystywana jest m.in. do takich zagadnień, jak: teoria rozwoju, planowanie i gospodarka regionalna, analiza jakości, sterowanie procesami gospodarczymi itp.

Ogólna definicja WAP stwierdza, że jest to zestaw metod statystycznych służących do celowego wyboru informacji o elementach danej zbiorowości i wykrywaniu prawidłowości we wzajemnych relacjach między tymi elementami. Definicja bardziej operacyjna mówi, że WAP jest zestawem metod i technik porównywania obiektów wielocechowych.

W ramach WAP rozpatrywane są m.in. takie zagadnienia, jak:

- hierarchizacja obiektów i ich zbiorów ujmowanych w wielowymiarowych przestrzeniach cech z punktu widzenia pewnej charakterystyki, której nie można zmierzyć w sposób bezpośredni (poziom rozwoju społeczno-gospodarczego, standard życia,

⁸Por. Bivaud R., Modelowanie geograficznych układów czaso-przestrzennych, PAN, Oddział w Poznaniu, seria: Geografia t.VII, PWN, 1981; Tucker L.R., Implications of Factor Analysis of Three-way Matrices for Measurement of Change, w pr. Problems in Measuring Change, ed. C.W.Harris, Madison 1963.

- stopień jakości wyrobów, efektywność działalności przedsiębiorstwa, przydatność różnych wariantów lokalizacyjnych, itp.),
- wyodrębnienie jakościowo jednorodnych podzbiorów obiektów, zawierających elementy podobne do siebie z punktu widzenia wyróżnionej charakterystyki agregatowej a jednocześnie niepodobne do elementów należących do innych podzbiorów,
 - wybór cech najbardziej diagnostycznych, szacowanie współczynników ich ważności oraz ocena kierunku i siły wpływu, jaki wywierają poszczególne zmienne na poziom rozwoju ekonomicznego, społecznego, technicznego, itd.,
 - metody normowania i agregacji różnoimiennych zmiennych,
 - analiza charakteru i siły współzależności pomiędzy zbiorami zmiennych,
 - przeprowadzenie szczegółowej analizy prawidłowości dynamiczno-strukturalnych (izokwenty rozwoju, optymalne, proporcjonalne, harmonijne i selektywne ścieżki oraz strategie rozwoju),
 - określenie wzorców rozwoju w procesie predykcji metodą analogii oraz ustalenie wielkości opóźnień pomiędzy poszczególnymi obiektami z punktu widzenia syntetycznej miary rozwoju,
 - automatyczna redukcja liczby zmiennych objaśniających w modelach regresji.

Podstawą wielowymiarowej analizy porównawczej jest macierz obserwacji:

$$X = [x_{ij}] \quad (i=1, \dots, n; \quad j=1, \dots, m) \quad (1.47)$$

gdzie n , m to liczba obiektów oraz liczba zmiennych, przy czym macierz (1.47) dotyczy określonego okresu czasu. Rozszerzeniem WAP jest dynamiczna wielowymiarowa analiza porównawcza (DWAP), która zajmuje się analizą statystyczną danych ujętych w postaci macierzy trójwymiarowej:

$$X = [x_{ijt}] \quad (i=1, \dots, n; j=1, \dots, m; t=1, \dots, k) \quad (1.48)$$

gdzie k jest liczbą uwzględnionych w badaniach okresów czasu.

W praktyce często zachodzi potrzeba prowadzenia wielowariantowych analiz porównawczych, w których poszczególne warianty różnią się uwzględnionymi zestawami cech diagnostycznych. Warianty te charakteryzują zazwyczaj różne układy (sektory) społeczno-ekonomiczne, np.: przemysł-rolnictwo-handel-transport, gospodarka-demografia, miasto-wieś, warunki produkcji-nakłady-efekty, ruch-baza-walory turystyczne, itp. Sektorową, wielowymiarową analizę porównawczą można realizować zarówno w ujęciu statycznym (SWAP), w którym wykorzystuje się trójwymiarową macierz danych:

$$X = [x_{ijl}] \quad (i=1, \dots, n; j=1, \dots, m; l=1, \dots, s) \quad (1.49)$$

gdzie s jest liczbą wyróżnionych sektorów, jak i w ujęciu dynamicznym (SDWAP), w którym podstawą rozważań jest macierz czterowymiarowa:

$$X = [x_{ijlt}] \quad (i=1, \dots, n; j=1, \dots, m; l=1, \dots, s; t=1, \dots, k) \quad (1.50)$$

Wzajemne powiązania między przedstawionymi wersjami statystycznej analizy porównawczej ująć można w postaci następującego schematu:



Łatwo zauważyć, że zakres rozszerzonych wersji WAP jest znacznie większy, ponieważ wszystkie problemy można tu analizować nie tylko w ujęciu przestrzennym, ale i dynamicznym.

W praktyce wersje te stosowane są stosunkowo rzadko, co wynika z dużej ich pracochłonności, jak i wymogu dysponowania obszernymi bazami danych, zawierającymi porównywalne, kompletne i wiarygodne informacje statystyczne.

Podstawowym narzędziem analitycznym WAP są syntetyczne mierniki wyznaczone według ogólnego schematu:

$$X \rightarrow \tilde{X} \rightarrow C \rightarrow W \rightarrow N \rightarrow A \rightarrow Q \rightarrow Q' \quad (1.51)$$

gdzie:

- X - macierz realizacji zmiennych wyjściowych (pierwotnych),
- $\tilde{X}$ - macierz zredukowana realizacji zmiennych diagnostycznych (finalnych),
- C - kierunek preferencji zmiennych diagnostycznych w stosunku do zmiennej syntetycznej,
- W - systemy wag dla zmiennych diagnostycznych,
- N - sposoby normalizacji zmiennych,
- A - formuły agregacji zmiennych znormalizowanych,
- Q - realizacje zmiennej syntetycznej (agregatowej),
- Q' - mierniki syntetyczne.

Zmienna syntetyczna Q pozwala uporządkować liniowo zbiór obiektów, od obiektu najlepszego do najgorszego (lub odwrotnie) z punktu widzenia charakterystyki agregatowej. Obecnie przedstawione zostaną podstawowe problemy związane z poszczególnymi etapami wyznaczania zmiennych syntetycznych.

Kryteria doboru zmiennych diagnostycznych

Problem doboru zmiennych diagnostycznych występuje nie tylko w statystycznej analizie porównawczej, ale także i taksonomii numerycznej, analizie dyskryminacyjnej, modelowaniu

ekonometrycznym, teorii rozpoznawania obiektów itd. Z uwagi na różną specyfikę tych dyscyplin kryteria i metody doboru zmiennych diagnostycznych mimo szeregu podobieństw, nie zawsze są jednakowe. W przypadku WAP najczęściej uważa się, że zmienne diagnostyczne powinny:

- ujmować najbardziej istotne własności analizowanych zjawisk i dokładnie je reprezentować,
- być prosto, jasno i ściśle zdefiniowane,
- być logicznie ze sobą powiązane,
- zawierać duży ładunek informacji,
- być bezpośrednio lub pośrednio mierzalne, co sprowadza się do istnienia wiarygodnych i stosunkowo łatwo osiągalnych danych statystycznych,
- być wyrażone w jednostkach naturalnych a nie wartościowych, raczej w postaci wskaźników natężenia niż w wielkościach absolutnych,
- charakteryzować się wysoką zmiennością przestrzenną (postulat dużego stopnia informacyjności),
- nie być wzajemnie wysoko skorelowane (postulat braku redundancji),
- być wysoko skorelowane ze zmiennymi niediagnostycznymi oraz zmienną syntetyczną,
- umożliwiać wzajemną kontrolę,
- nie opisywać zjawisk i procesów specyficznych (np. wynikających ze społecznego podziału pracy).

Problem doboru zmiennych diagnostycznych nie jest łatwy i wymaga obszernych analiz statystycznych. Część z podanych kryteriów powinna być uwzględniona przy konstruowaniu wstępnej, w miarę możliwości szerokiej listy zmiennych, zawierających kandydatki na zmienne diagnostyczne. Część kryteriów natomiast

może stanowić punkt wyjścia do opracowania algorytmów, służących do wyboru zmiennych diagnostycznych, które tworzą finalną listę cech.

Algorytmy takie są stosunkowo łatwe do sformalizowania i sprowadzają się do wyznaczenia odpowiednich parametrów statystycznych oraz ich analizy z punktu widzenia "właściwego" kształtowania się ich poziomu względnie zmian w czasie. Kryterium minimalizacji wzajemnego skorelowania zmiennych diagnostycznych i maksymalizacji ich skorelowania ze zmiennymi wyeliminowanymi z listy kandydatek, uwzględnia się poprzez przeprowadzenie analizy taksonomicznej cech, wykorzystując procedury ich grupowania.

Określenie charakteru zmiennych

U podstaw procedur porządkowania liniowego tkwi konieczność rozróżniania wśród wyjściowego zbioru zmiennych diagnostycznych $\{X\}$ trzech podzbiorów. Pierwszy $\{X_S\}$ zawiera stymulanty, drugi $\{X_D\}$ - destymulanty, natomiast trzeci $\{X_N\}$ - nominanty. Zakłada się, że zbiór zmiennych diagnostycznych został poprawnie wyspecyfikowany, to znaczy, że nie zawiera zmiennych neutralnych (obojętnych) nie mających związku merytorycznego ze zmienną wyjścia (syntetycznym miernikiem rozwoju).

Ogólnie biorąc przez stymulantę rozumie się taką zmienną, której wysokie wartości są pożądane z punktu widzenia charakterystyki agregatowej, natomiast niskie wartości są niepożądane (np. ze względu na jakość życia stymulantami są wskaźniki liczby lekarzy, samochodów, telewizorów, powierzchni mieszkalnej na osobę). Dla destymulant sytuacja jest odwrotna - ze względu

na standard życia, taki charakter mają np. wskaźniki inflacji, bezrobocia, śmiertelności niemowląt, itd. W przypadku nominant pożądaną są "normalne" poziomy zmiennych (ze względu na poziom życia nominantami mogą być np.: stopa oprocentowania kredytów, przyrost naturalny, udział inwestycji w dochodzie narodowym). Wszelkie odchylenia nominant od poziomu "normalnego" są zjawiskiem negatywnym z punktu widzenia charakterystyki agregatowej.

Określenie charakteru zmiennych powinno odbywać się w zasadzie na przesłankach pozastatystycznych (merytorycznych). W przypadku braku odpowiedniej teorii ekonomicznej wyjaśniającej charakter zmiennych dobrze jest posłużyć się metodą ocen ekspertów. Poprawność określenia charakteru zmiennych można zweryfikować ex post sprawdzając kierunek skorelowania zmiennych diagnostycznych ze zmienną syntetyczną. Skorelowanie to powinno być jednakowego znaku dla zmiennych danego typu (np. dodatnie dla stymulant, a ujemne dla destymulant, lub odwrotnie). Dla zmiennych typu nominant nie powinno się obserwować istotnych korelacji ze zmienną agregatową.

Innym formalnym kryterium oceny charakteru zmiennych może być założenie, że wszystkie stymulanty powinny być ze sobą dodatnio skorelowane, podobnie jak wszystkie destymulanty między sobą. Współczynniki korelacji między stymulantami i destymulantami powinny być ujemne. Natomiast między nominantami a tak stymulantami jak i destymulantami nie powinno się obserwować istotnych zależności korelacyjnych.

Systemy wag

Poglądy na zagadnienie estymacji współczynników wagowych dla zmiennych diagnostycznych nie są jednolite. Dotyczy to

zarówno tego, czy w ogóle należy stosować systemy wag, jak i samego sposobu określania wag. Warto zauważyć, że omówiony poprzednio problem doboru zmiennych diagnostycznych zawiera w sobie ważenie zmiennych, przy czym przyjmowane są tu wagi dwojakiego rodzaju - zerowe dla cech odrzuconych z listy wyjściowej oraz jedynkowe dla zmiennych zaliczanych do listy finalnej. W tym kontekście celowość ważenia cech nie jest na ogół kwestionowana.

Jeżeli chodzi o sposoby określania wag to można tu wyróżnić dwa podejścia. Jedno opiera się na informacjach pozastatystycznych i realizowane jest metodą ocen ekspertów (indywidualnych lub zespołowych), natomiast drugie podejście wykorzystuje wyłącznie informacje tkwiące w analizowanym materiale statystycznym.

W praktyce częściej stosuje się drugie podejście, przy czym zazwyczaj współczynniki wagowe proponuje się ustalać odwrotnie proporcjonalnie do błędu pomiaru zmiennych względnie wprost proporcjonalnie do ich wariancji. System wag oparty na wariancjach zmiennych dany jest wzorem:

$$\alpha_j = \frac{V_j}{\sum_{j=1}^m V_j} \quad (j=1, \dots, m) \quad (1.52)$$

gdzie V_j to współczynnik zmienności j -tej zmiennej.

Należy jednak zaznaczyć, że w badaniach empirycznych przeważają na ogół systemy wag stałych.

Metody normalizacji zmiennych

Warunkiem umożliwiającym wyznaczenie zmiennych agregatowych jest normalizacja zmiennych diagnostycznych mająca na celu:

- 1) doprowadzić różnoimienne cechy do wzajemnej porównywalności (postulat addytywności),
- 2) ujednolicić charakter cech, przekształcając destymulanty na stymulanty lub odwrotnie (postulat jednolitej preferencji),
- 3) wyeliminować z obliczeń wartości niedodatnie (postulat dodatniości),
- 4) zastąpić zróżnicowane zakresy zmienności poszczególnych cech zakresem stałym (postulat stałości rozstępu lub stałości wartości ekstremalnych).

Najważniejszym z wymienionych postulatów jest postulat addytywności, bez spełnienia którego nie jest możliwe wyznaczenie zmiennych syntetycznych. Spotykane w literaturze procedury normalizacji realizujące ten postulat można sprowadzić do następującego wzoru:

$$x_i' = \left(\frac{x_i - A}{B} \right)^p \quad (i=1, \dots, n) \quad (1.53)$$

gdzie x_i , x_i' to wyjściowe i znormalizowane wartości i -tej realizacji zmiennej, n - liczba obserwacji, natomiast A, B oraz p to parametry, które mogą przyjmować różne wartości, na przykład:

- 1) dla przekształceń ilorazowych: $A=0$, $p=1$, natomiast B może być równe maksymalnej ($x_{\max}$), minimalnej ($x_{\min}$), średniej ($\bar{x}$) lub sumie ($\sum x_i$) wartości zmiennej,
- 2) dla standaryzacji: $A=0$ lub $A=\bar{x}$, parametr B równy jest odchyleniu standardowemu zmiennej (s_x), natomiast $p > 0$,

3) dla unitaryzacji: $A=0$, $A=x_{\min}$ lub $A=\bar{x}$, parametr $p > 0$, natomiast $B = x_{\max} - x_{\min}$.

W praktyce najczęściej korzysta się ze standaryzacji na 0-1, w której $A=\bar{x}$, $B=s_x$, $p=1$. Wartości tak standaryzowanej zmiennej mają średnią równą zero oraz odchylenie standardowe równe 1.

Innym zalecanym sposobem normalizacji jest unitaryzacja ($A=0$, $B=x_{\max} - x_{\min}$, $p=1$) względnie przekształcenie ilorazowe ($A=0$, $B=\sum x_i$, $p=1$).

Ujednolicenia charakteru zmiennych można dokonać przy ich specyfikacji (najczęściej polega to na zastosowaniu przekształcenia odwrotnościowego, które zmienia charakter zmiennej na przeciwny), względnie opierając się na wartościach znormalizowanych. W tym ostatnim przypadku korzysta się z następującego przekształcenia:

$$x_i'' = \begin{cases} x_i' & \text{dla } X \in \{X_S\} \\ -x_i' & \text{dla } X \in \{X_D\} \end{cases} \quad (1.54)$$

gdzie $\{X_S\}$ oraz $\{X_D\}$ to zbiór stymulant i destymulant.

Wyeliminowanie ujemnych wartości zmiennej znormalizowanej uzyskuje się poprzez dodanie do wszystkich wartości x_i'' stałej δ :

$$\delta = - \min_{i,j} \{x_{ij}''\} \quad (1.55)$$

pod warunkiem, że $\min \{x_{ij}''\} \approx 0$.

Formuły agregacji zmiennych

Większość formuł agregacji zmiennych można podzielić na dwie grupy:

- bezwzorcowe,

- wzorcowe.

Metody bezwzorcowe sprowadzają się do operacji uśrednienia znormalizowanych wartości zmiennych x'_{ij} z uwzględnieniem przyjętych współczynników wagowych α_j . Formuły bezwzorcowe odpowiadają różnym postaciom wartości przeciętnej, na przykład:

- średniej arytmetycznej:

$$q_i = \sum_{j=1}^m \alpha_j x'_{ij} \quad (i=1, \dots, n) \quad (1.56)$$

- średniej geometrycznej:

$$q_i = \prod_{j=1}^m (x'_{ij})^{\alpha_j} \quad (i=1, \dots, n) \quad (1.57)$$

- średniej harmonicznej:

$$q_i = \left[\sum_{j=1}^m \frac{\alpha_j}{x'_{ij}} \right]^{-1} \quad (i=1, \dots, n) \quad (1.58)$$

gdzie q_i wartości zmiennej syntetycznej, n, m - liczba obiektów zmiennych diagnostycznych.

Istota metod wzorcowych polega na wyznaczaniu odległości poszczególnych obiektów od obiektu modelowego, przy czym w zależności od sposobu zdefiniowania współrzędnych tego obiektu wzorcowe metody dzielą się na dwie grupy. Pierwsza opiera się na współrzędnych równych górnemu biegunowi zbioru obiektów (wzorcu rozwoju), natomiast druga - na dolnym biegunie zbioru (antywzorcu rozwoju). W zakresie procedur wzorcowych można m.in. wykorzystać następujące formuły odległości:

- odległość Minkowskiego:

$$q_i = \left[\sum_{j=1}^m \alpha_j |x'_{ij} - x'_{oj}|^p \right]^{1/p} \quad (i=1, \dots, n) \quad (1.59)$$

która dla $p=1$ daje odległość miejską (Hamminga) a dla $p=2$ - odległość Euklidesa;

- odległość Jeffreysa-Matusita:

$$q_i = \sum_{j=1}^m \alpha_j (\sqrt{x'_{ij}} - \sqrt{x'_{oj}}) \quad (i=1, \dots, n) \quad (1.60)$$

- odległość Braya-Curtisa:

$$q_i = \frac{\sum_{j=1}^m \alpha_j |x'_{ij} - x'_{oj}|}{\sum_{j=1}^m \alpha_j |x'_{ij} + x'_{oj}|} \quad (i=1, \dots, n) \quad (1.61)$$

- odległość Canberra:

$$q_i = \sum_{j=1}^m \alpha_j \frac{|x'_{ij} - x'_{oj}|}{x'_{ij} + x'_{oj}} \quad (i=1, \dots, n) \quad (1.62)$$

- odległość Clarka:

$$q_i = \left[\sum_{j=1}^m \alpha_j \left(\frac{|x'_{ij} - x'_{oj}|}{x'_{ij} + x'_{oj}} \right)^2 \right]^{1/2} \quad (i=1, \dots, n) \quad (1.63)$$

przy czym x'_{oj} to znormalizowane współrzędne obiektu modelowego (wzorca lub antywzorca rozwoju). Współrzędne te można określić na podstawie:

- ocen ekspertów,
- planów perspektywicznych lub programów rozwoju
- danych empirycznych.

Najwygodniejsza w praktyce jest ostatnia metoda, w której przy założeniu, że wszystkie zmienne diagnostyczne sprowadzono do jednakowego charakteru stymulant, współrzędne obiektu modelowego wyznacza się ze wzoru:

$$x'_{oj} = \begin{cases} \max_i \{x'_{ij}\} & \text{dla wzorca rozwoju} \\ \min_i \{x'_{ij}\} & \text{dla antywzorca rozwoju} \end{cases} \quad (1.64)$$

Wyznaczenie syntetycznych mierników rozwoju

Wartości zmiennej syntetycznej q_i pozwalają uporządkować liniowe obiekty od najbardziej rozwiniętego (położonego najbliżej wzorca rozwoju lub najdalej od antywzorca rozwoju względnie o średnio "najlepszych" wartościach zmiennych diagnostycznych) do obiektu najmniej rozwiniętego. Syntetyczne mierniki rozwoju dla i -tego obiektu uzyskuje się przy pomocy wzoru:

$$q'_i = \frac{q_i}{\|Q\|} \quad (i=1, \dots, n) \quad (1.65)$$

gdzie $\|Q\|$ jest normą zmiennej syntetycznej, za którą można przyjąć jedną z następujących wielkości:

- maksymalną wartość zmiennej Q :

$$\|Q\| = \max_i \{q_i\} \quad (1.66)$$

- statystyczną wartość maksymalną:

$$\|Q\| = \bar{q} + 2s_q \quad (1.67)$$

gdzie $\bar{q}$, s_q to średnia arytmetyczna i odchylenie standardowe zmiennej syntetycznej,

- sumę wartości zmiennej Q:

$$\|Q\| = \sum_{i=1}^n q_i \quad (1.68)$$

- rozstęp:

$$\|Q\| = \max_i \{q_i\} - \min_i \{q_i\} \quad (1.69)$$

Mierniki rozwoju (1.65) w przypadku stosowania wzorcowych formuł agregacji mają różne kierunki preferencji w zależności od rodzaju przyjętego obiektu modelowego. W celu zachowania jednakowego kierunku preferencji można dla mierników uzyskanych przy pomocy formuł odległości opartych na wzorcu rozwoju zastosować następujące przekształcenie:

$$q_i'' = 1 - q_i' \quad (i=1, \dots, n) \quad (1.70)$$

W rezultacie niezależnie od rodzaju wykorzystywanych formuł agregacji (wzorcowych jak i bezwzorcowych) oraz sposobu zdefiniowania obiektu modelowego otrzymuje się mierniki, których duże wartości (bliskie 1) świadczą o wysokim poziomie rozwoju obiektu, natomiast małe wartości (bliskie 0) - o niskim stopniu rozwoju.

Część II

WYNIKI BADAN EMPIRYCZNYCH

2.1 K^{nt}asyfikacja układów analizy

Celem przeprowadzonych badań empirycznych było ustalenie strukturalnych powiązań krajowego ruchu turystycznego w układzie województw z jego podstawowymi determinantami, do których zaliczono:

- A) poziom infrastruktury turystyki krajowej,
- B) wielkość zanieczyszczenia środowiska naturalnego,
- C) ocenę walorów wypoczynkowych i turystycznych
- D) bazę krajowego ruchu turystycznego.

Krajowy ruch turystyczny scharakteryzowany jest przy pomocy następujących zmiennych:

- E1 - korzystający z noclegów w obiektach turystycznych,
- E2 - udzielone noclegi w obiektach turystycznych,
- E3 - korzystający z noclegów w obiektach czasowo-wypoczynkowych,
- E4 - udzielone noclegi w obiektach czasowo-wypoczynkowych,
- E5 - korzystający z noclegów ogółem ($E5 = E1 + E3$)
- E6 - udzielone noclegi ogółem ($E6 = E2 + E4$)

Determinanty ruchu A-D również opisane są przez wieloelementowy zbiór zmiennych, których wykaz zawiera tabl. 2.1. Obok nazw, symboli i mian poszczególnych zmiennych przytoczono tu także

wartości odpowiadających im współczynników zmienności V.

Wyjściowe zmienne mają postać wskaźników dwojakiego typu:¹

- udziały procentowe województw w zakresie danej cechy w stosunku do jej wielkości ogółem, dla całej Polski:

$$x'_i = \frac{x_i}{\sum_{i=1}^n x_i} 100 \quad (2.1)$$

- relacje powyższych udziałów x'_i w stosunku do udziałów województw w zakresie wybranych cech bazowych x_i^0 (powierzchni, ludności, powierzchni obszarów wypoczynkowych, liczby miejsc noclegowych)

$$x''_i = \frac{x'_i}{x_i^0} \quad (2.2)$$

przy czym $n = 49$ natomiast x_i to wartości danej cechy w wyrażeniu naturalnym. Zmienne opisujące ruch turystyczny E1 - E6 są wskaźnikami typu (2.1).

Należy zaznaczyć, że dobór poszczególnych zmiennych-determinant nie był przypadkowy i wynikał z przeprowadzonej uprzednio przy pomocy metod taksonomicznych analizy diagnostyczności szerokiego, ponad 100-elementowego zbioru zmiennych². Stąd też w wykazie uwzględnionych cech nie ma szeregu zmiennych, które

¹Pomimo wspólnej podstawy wyjściowej wskaźniki te różnią się w swojej interpretacji i charakteryzują odmienne aspekty struktury województw.

²Szczegółowe omówienie rezultatów delimitacji zmiennych oraz wykorzystanych w tym celu metod znaleźć można w pracy Autorów: "Ruch turystyczny i jego uwarunkowania w ujęciu przestrzennym", AE Kraków, Instytut Turystyki, Warszawa, 1985

z merytorycznego punktu widzenia wstępnie zakwalifikowano do rozważań, a z drugiej strony niektóre z nich przyjęte są w "dwojakiej" postaci, zarówno określonej wzorem (2.1) jak i (2.2).

Dane wyjściowe zaczerpnięto z Roczników Statystycznych Województw, GUS oraz w zakresie walorów turystycznych z opracowania: "Ocena przyrodniczych walorów wypoczynkowych", Instytut Turystyki, Oddział we Wrocławiu, Wrocław 1983 (maszynopis), temat R.17.2.1. Wszystkie źródłowe informacje statystyczne wykorzystane w badaniach znajdują się w komputerowej bazie GOSTUR.

Prezentowane wyniki analiz odnoszą się tylko do roku 1983 i stąd też mają charakter wstępny. Szczególna uwaga zostanie zwrócona na wnioski o charakterze metodologicznym w celu ewentualnego uproszczenia procedury badań oraz wyeliminowania wariantów rozważań o niewielkiej wartości poznawczej. W rezultacie powinno to znacznie ułatwić przewidywaną w następnym etapie analizę dynamiczną, gdzie podobne badania będą prowadzone dla pozostałych lat okresu 1976-1987 w celu oceny zmian zachodzących w kształtowaniu się struktury przestrzennej krajowego ruchu turystycznego w powiązaniu z wyróżnionymi determinantami.

2.2 Analiza czynnikowa determinant krajowego ruchu turystycznego

W badaniach wykorzystano metodę głównych składowych według techniki R, co stanowi w praktyce podstawowy wariant analizy czynnikowej. Punktem wyjścia rozważań była macierz wskaźników X o wymiarach: 49 województw x 28 zmienne, wyrażających uwzględnione determinanty krajowego ruchu turystycznego (tabl.2.1).

Tablica 2.1

Wykaz cech charakteryzujących determinanty krajowego
ruchu turystycznego

Nr	Nr w grupie	Nazwa cechy	Miano	V%
		Udział województw w:		
1	A 1	- długości linii krajowych	%	61,3
2	A 2	- długości dróg utwardzonych	%	33,9
3	A 3	- liczbie zakładów gastronomicznych	%	91,0
4	A 4	- liczbie przystanków PKS /P	-	30,1
5	A 5	- powierzchni sklepów /P	-	108,6
6	A 6	- długości linii kolejowych /L	-	51,3
7	A 7	- długości dróg z siecią PKS /L	-	42,1
8	A 8	- liczbie miejsc konsumpcyjnych /L	-	41,2
9	B 1	- liczbie zakładów uciążliwych /P	-	141,8
10	B 2	- wielkości emisji pyłów /P	-	188,8
11	B 3	- wielkości emisji gazów /P	-	237,7
12	B 4	- wielkości odpadów uciążliwych /P	-	359,0
13	B 5	- wielkości emisji pyłów /L	-	128,9
14	B 6	- wielkości emisji gazów /L	-	143,6
15	C 1	- powierzchni obszarów wypoczynkowych I - III kat.	%	89,1
16	C 2	- powierzchni obszarów wypoczynkowych ogółem	%	58,6
17	C 3	- powierzchni lasów ochronnych /P	-	83,4
18	C 4	- liczbie rezerwatów /P	-	147,0
19	C 5	- liczbie pomników przyrody /P	-	94,8
20	C 6	- powierzchni obszarów wypoczynkowych IV kat./P	-	58,9
21	C 7	- powierzchni obszarów zagrożonych I - III kat./P	-	57,1
22	C 8	- powierzchni obszarów zagrożonych IV kat. /P	-	84,1
23	C 9	- powierzchni obszarów zagrożonych W	-	79,3
24	D 1	- liczbie miejsc noclegowych w obiektach turystycznych	%	85,4
25	D 2	- liczbie miejsc noclegowych w obiektach wczasowo-wypoczynkowych	%	81,2
26	D 3	- liczbie miejsc noclegowych w obiektach turystycznych /P	-	136,4
27	D 4	- liczbie miejsc noclegowych w obiektach turystycznych /T	-	52,6
28	D 5	- liczbie miejsc noclegowych w obiektach wczasowo-wypoczynkowych /T	-	84,5

Uwaga: P - udział województwa w powierzchni ogółem
 L - udział województwa w ogólnej liczbie ludności
 W - udział województwa w powierzchni obszarów wypoczynkowych ogółem
 T - udział województwa w liczbie miejsc noclegowych w obiektach turystycznych

Z uwagi na duże rozmiary nie przytoczono w pracy zarówno tej tablicy jak również macierzy danych standaryzowanych Z , macierzy współczynników korelacji R , oraz wektorów własnych A , które to elementy wyznaczały kolejne etapy obliczeń.

W tablicy 2.2 zamieszcza się jedynie uporządkowane malejąco wartości własne macierzy korelacji λ_j , oceny zasobu zmienności wyczerpywanej przez kolejne czynniki główne $\hat{h}_j$ (por. wzór 1.39) oraz skumulowane zasoby H_k (por. wzór 1.40) wskazujące ile procent zmienności całkowitej wyczerpuje kolejnych k czynników.

W tablicy 2.2 uwzględniono wstępnie 14 składowych głównych, wyczerpujących łącznie 95,9% zmienności całkowitej. Jak się jednak okazuje znaczenie wielu z nich jest nieduże (zasoby poniżej 5%) w związku z czym do dalszej analizy zdecydowano przyjąć tylko 5 pierwszych składowych, zawierających ponad 70% informacji zawartych w wyjściowych zmiennych, przy czym każdy z nich wyczerpuje co najmniej 5% zasobu zmienności całkowitej.

Tak duża liczba czynników niezbędnych do wyjaśnienia dostatecznie dużej części informacji wynika stąd, że wyjściowe cechy poddane były uprzednio zabiegowi redukcji metodami taksonomicznymi, w rezultacie czego pierwotny zbiór zawierał już cechy o możliwie małym skorelowaniu.

W tablicy 2.3 przytoczono macierz ładunków czynnikowych W o wymiarach 28×5 otrzymaną w drodze rotacji metodą VARIMAX. Ładunki te są współczynnikami korelacji między poszczególnymi cechami wyjściowymi a czynnikami głównymi i stanowią podstawę ich interpretacji merytorycznej.

Tablica 2.2

Wartości własne λ_j , zasoby $\hat{h}_j$ oraz skumulowane
 zasoby zmienności H_k w analizie czynnikowej
 determinant krajowego ruchu turystycznego

Nr	λ_j	$\hat{h}_j$	H_k
1	8,162	29,2	29,2
2	5,297	18,9	48,1
3	2,882	10,3	58,4
4	1,989	7,1	65,5
5	1,532	5,4	70,9
6	1,236	4,4	75,3
7	1,111	4,0	79,3
8	1,028	3,7	83,0
9	0,998	3,5	86,5
10	0,786	2,9	89,4
11	0,604	2,1	91,5
12	0,499	1,8	93,3
13	0,409	1,5	94,8
14	0,312	1,1	95,9

Tablica 2.3

Macierz ładunków czynnikowych W zawierająca współczynniki korelacji $r_{X_j F_l}$ ($j = 1, \dots, 28$; $l = 1, \dots, 5$) między wyjściowymi cechami X_j a czynnikami głównymi F_l

Nr k	Układ	F_1	F_2	F_3	F_4	F_5
1	A1	,40	-,77 *	,01	,04	-,01
2	A2	,06	-,76 *	-,18	-,06	,02
3	A3	,57	-,40	,57	,05	-,05
4	A4	,19	,12	,49	,17	,44
5	A5	,43	,04	,76 *	-,19	-,03
6	A6	-,12	-,54	-,44	,32	,19
7	A7	-,35	-,17	-,69 *	,25	,21
8	A8	,03	-,57	,33	,61	,10
9	B1	,73 *	,06	,48	-,21	-,08
10	B2	,80 *	-,01	,40	-,09	,15
11	B3	,89 *	-,03	,24	-,10	,17
12	B4	,88 *	-,12	,05	-,16	,12
13	B5	,40	,02	,08	,10	,53
14	B6	,66 *	,00	,02	,02	,51
15	C1	-,06	-,68 *	-,04	,42	-,40
16	C2	-,18	-,87 *	-,14	,11	-,16
17	C3	,44	-,42	,42	,12	-,00
18	C4	-,14	-,21	,01	,11	-,26
19	C5	-,06	,08	,80 *	,03	,12
20	C6	-,25	-,29	-,03	-,53	,56
21	C7	,71 *	-,12	,08	,41	-,03
22	C8	,13	,10	,05	-,23	,69 *
23	C9	,71 *	,26	,06	,18	,25
24	D1	,03	-,54	,47	,41	-,22
25	D2	,00	-,33	,03	,78 *	-,25
26	D3	,10	,01	,90 *	,19	,03
27	D4	,23	,06	,68 *	-,14	,19
28	D5	-,08	-,06	-,21	,78 *	-,03

W tym celu symbolem^{*} zaznaczono w tablicy 2.3 szczególnie wysokie co do modułu wartości ładunków, przyjmując jako wartość graniczną $|r_x| \approx 0.7$. Założenie wartości granicznej na niższym poziomie powoduje przeidentyfikowanie czynników poprzez fakt przyporządkowania każdemu z nich dużej liczby coraz to bardziej różnorodnych cech. Przyjęcie natomiast wyższej wartości granicznej prowadzi do skojarzenia czynników z małą liczbą cech, co również nie pozwala na właściwą ich interpretację.

Jak można zauważyć pierwszy, 29-procentowy czynnik główny F_1 można określić jako przemysłowe zanieczyszczenie środowiska. Wyznaczają go zmienne B_1, B_2, B_3, B_4, B_6 oraz C_7, C_9 , przy czym te ostatnie są ściśle związane z poziomem zanieczyszczenia.

Na drugi, 19-procentowy czynnik F_2 składają się po dwie cechy charakteryzujące wielkość infrastruktury turystyki A_1, A_2 oraz walorów wypoczynkowych C_1, C_2 . Czynnik ten opisuje więc dostępność i atrakcyjność turystyczna.

Trzeci 10-procentowy czynnik F_3 jest najbardziej niejednorodny, gdyż zawiera zarówno zmienne określające natężenie infrastruktury turystycznej A_5, A_7 , bazy turystycznej D_3, D_4 jak i liczbę pomników przyrody C_5 . Czynnik ten proponuje się nazwać bazą turystyczną, w odróżnieniu od bazy czasowo-wypoczynkowej, która stanowi treść czwartego 7-procentowego czynnika głównego F_4 , wyznaczonego przez zmienne D_2, D_5 .

Ostatni 5-procentowy czynnik F_5 wyznacza tylko zmienna C_8 , w związku z czym można go określić jako potencjalne zagrożenie środowiska.

Oceniając wyniki identyfikacji wyodrębnionych czynników głównych należy stwierdzić, że pokrywają się one w pewnym stopniu

ze znaczeniem merytorycznym poszczególnych grup zmiennych (A, B, C, D), jakkolwiek nie jest to zgodność pełna. Zmienne grupy C z definicji będące wypadkową walorów krajobrazowych, dostępności komunikacyjnej, czystości środowiska naturalnego itd., zostały rozdzielone między różne czynniki główne. Natomiast zmienne z grupy D stały się podstawą wyodrębnienia dwóch odrębnych czynników F_3 i F_4 . Oceny zasobów zmienności $\hat{h}_j$ wyznaczają wagi czynników i określają znaczenie jakie należy im nadawać w toku dalszych analiz. Warto tu odnotować fakt, że w przestrzennym rozmieszczeniu turystyki szczególnie dużą rolę odgrywa stopień zanieczyszczenia środowiska F_1 , oraz że dostępność i atrakcyjność turystyczna F_2 jest bardziej istotna niż szeroko rozumiana baza turystyki F_3 i F_4 .

Analizę czynnikową kończy wyznaczenie według wzoru (1.42) realizacji wyodrębnionych czynników. Wartości te zebrano w tabl. 2.4, a przy ich interpretacji warto zwrócić uwagę na fakt, że wysokim realizacjom czynnika F_2 odpowiada mała atrakcyjność turystyczna województwa i odwrotnie. Tak więc czynnik ten należałoby nazwać niedostępnością i nieatrakcyjnością turystyczną.

Ogólnie biorąc, analiza realizacji czynników głównych i porównanie ich z wyjściowymi wartościami zmiennych pozwala łatwiej zinterpretować istotę i znaczenie poszczególnych czynników. Dla przykładu widoczna tu jest różnica między czynnikami F_3 oraz F_4 . Nasycenie bazą turystyczną ma miejsce głównie w dużych aglomeracjach miejskich, podczas gdy baza wczasowo-wypoczynkowa zlokalizowana jest w województwach nadmorskich i województwach nowosądeckim.

Otrzymane wyniki niewątpliwie są zgodne z oczekiwaniami, pozwalając w sposób obiektywny wyrazić ilościowo złożone kategorie

Tablica 2.4

Realizacje wyodrębnionych czynników głównych $F_1 - F_5$
 traktowanych jako determinanty krajowego ruchu
 turystycznego

Nr	WOJEWÓDZTWO	F_1	F_2	F_3	F_4	F_5
1	St.warszawskie	5,66	2,06	17,83	-3,07	2,86
2	bialskopodlaskie	-4,92	4,42	-5,36	-2,32	2,37
3	białostockie	-4,17	,43	-3,51	-2,96	-1,06
4	bielskie	3,91	-, 54	9,07	3,87	-,08
5	bydgoskie	1,49	-7,46	-,64	3,11	-2,87
6	chełmskie	-2,23	5,60	-3,67	-1,28	2,59
7	ciechanowskie	-4,43	4,32	-3,58	-1,09	-,41
8	częstochowskie	-,44	-,62	-1,04	-1,67	,24
9	elbląskie	-2,77	-,58	-3,33	3,90	-1,50
10	gdańskie	1,32	-5,48	2,50	5,84	-4,66
11	gorzowskie	-4,39	-6,18	-5,08	,44	-,95
12	jeleniogórskie	4,57	-2,31	4,38	2,13	3,84
13	kaliskie	-3,69	,47	-1,23	-2,70	,63
14	katowickie	32,94	-5,06	17,49	-3,14	5,50
15	kieleckie	-,97	-3,80	-1,21	-1,09	-1,75
16	konińskie	2,19	4,18	-1,56	-,55	4,07
17	koszalińskie	-2,73	-7,82	-5,15	10,14	-4,25
18	m.krakowskie	16,02	3,52	13,54	,56	4,69
19	krośnieńskie	-3,28	-1,21	-,16	-,51	-2,74
20	legnickie	11,10	2,60	1,11	-1,31	6,60
21	leszczyńskie	-3,90	1,95	-3,66	,74	,16
22	lubelskie	-,76	3,08	-,25	-2,84	1,42
23	łomżyńskie	-6,08	3,37	-5,56	-2,08	-2,27
24	m.łódzkie	7,28	8,23	12,49	-5,26	-,46
25	nowosądeckie	-1,92	-4,63	6,14	6,81	-4,32
26	olsztyńskie	-4,00	-7,71	-4,35	3,44	-4,15
27	opolskie	,63	-4,28	-,32	-,79	3,22
28	ostrołęckie	-3,92	2,68	-3,75	-3,08	1,86
29	pilskie	-4,44	-3,35	-4,24	,41	-,99
30	piotrkowskie	-2,27	1,86	-2,58	-1,79	-,02
31	płockie	0,69	5,68	-1,67	-1,19	,78
32	poznańskie	1,16	-2,57	3,01	,43	-1,60
33	przemyskie	-4,29	3,19	-3,36	-2,38	-,93
34	radomskie	-,61	2,89	-2,28	-3,29	2,01
35	rzeszowskie	-2,77	4,04	,22	-3,18	-,36
36	siedleckie	-5,39	2,19	-3,95	-3,10	,50
37	sieradzkie	-2,08	4,54	-3,20	-1,70	1,35
38	skierniewickie	-1,75	5,90	-1,44	-3,30	1,11
39	ślupskie	-3,88	-3,43	-4,97	5,82	-3,16
40	suwalskie	-4,77	-7,39	-4,85	4,59	-4,46
41	szczecińskie	2,15	-7,42	-1,74	7,74	-2,47
42	tarnobrzeskie	1,07	,20	-1,78	-2,78	3,19
43	tarnowskie	-2,71	4,09	-,25	-3,45	,63
44	toruńskie	-,23	1,34	-1,13	,76	-1,13
45	wałbrzyskie	3,83	-,27	7,95	3,57	-,99
46	włocławskie	,38	4,43	-1,85	,53	-,51
47	wrocławskie	1,00	,03	3,65	-1,57	-,59
48	zamojskie	-4,77	1,86	-4,45	-1,67	-,92
49	zielonogórskie	-2,78	-7,07	-2,11	,37	,02

jakimi są determinanty krajowego ruchu turystycznego. Realizacja czynników głównych w dalszych badaniach stanowią podstawę regionalizacji turystyki krajowej w ujęciu województw, oraz są traktowane jako zmienne objaśniające w modelach ekonometrycznych opisujących kształtowanie się ruchu turystycznego w Polsce.

2.3 Wyznaczanie syntetycznych mierników - determinant krajowego ruchu turystycznego

Punktem wyjścia w wielowymiarowej analizie porównawczej jest określenie charakteru poszczególnych zmiennych opisujących województwa w ramach wyróżnionych układów badań. Posłużono się w tym celu analizą współczynników korelacji zamieszczonych w tabl.

2.5. Symbolem ^{*} oznaczono tu przypadki istotnych skorelowań ujemnych, gdzie wartości współczynników są mniejsze od wartości krytycznej, wynoszącej - 0,27 dla poziomu istotności $\alpha = 0,05$ oraz $n-1=48$ stopni swobody.

Jak łatwo zauważyć tylko w zakresie środowiska naturalnego (układ B) wszystkie zmienne mają wyraźnie jednoznaczny charakter. Zmienne te z uwagi na ich interpretację są destymulantami rozwoju turystyki. W pozostałych układach występują obok dodatnich także i ujemne skorelowania między zmiennymi - liczba istotnie ujemnych współczynników podana jest poniżej każdej z tablic. Świadczyć to może o niejednorodnym charakterze zmiennych należących do poszczególnych układów.

W przypadku infrastruktury turystyki (A) oraz bazy turystycznej (D) przyjęto kierując się interpretacją merytoryczną zmiennych, że są one stymulantami rozwoju turystyki. Założenie, że niektóre

ze zmiennych, np. A7 D5 mają charakter destymulant zmieniałoby wprowadzie kierunek ujemnych skorelowań na dodatnie, ale równocześnie spowodowałoby pojawienie się ujemnych współczynników korelacji w miejsce dotychczasowych - dodatnich.

Nieco inna sytuacja jest w przypadku walorów turystycznych (układ D). Jak można zauważyć zmiana charakteru zmiennych C7, C8, C9 na przeciwny w stosunku do pozostałych zmiennych C1-C6 spowoduje spadek ogólnej liczby ujemnych skorelowań z 14 do 2. Dlatego też przyjęto, że ostatnie trzy zmienne z tego układu są destymulantami rozwoju turystyki, co jest także zgodne z ich interpretacją, natomiast pozostałe zmienne traktowane są tu jako stymulanty.

W celu wyznaczenia mierników syntetycznych wykorzystano:

- system wag stałych, zgodnie z którym każda zmienna ma jednakowy wpływ na wielkość zmiennej syntetycznej,
- normalizację zmiennych wyjściowych na 0-1 według wzoru (1.53) przy założeniu, że $A = \bar{x}$, $B = s_x$ oraz $p=1$,
- wzorcową formułę agregacji zmiennych opartą na odległości województw od górnego bieguna zbioru (wzorca rozwoju), przy czym przyjęto tu formułę Euklidesa - wzór 1.59 dla $p=2$ - oraz empiryczny sposób określenia współrzędnych wzorca w/g wzoru 1.64,
- statystyczną wartość maksymalną zmiennej syntetycznej (1.67) jako normę oraz przekształcenie (1.70) prowadzące do ostatecznych realizacji mierników.

Przy wyborze powyższych rozwiązań kierowano się wynikami analizy ich efektywności przytoczonymi w pracy: T.Grabiński: "Wielowymiarowa analiza porównawcza w badaniach dynamiki zjawisk

Tablica 2.5

Współczynniki korelacji między analizowanymi zmiennymi w ramach
wyróżnionych układów badań A - D

A	A1	A2	A3	A4	A5	A6	A7	A8
A1	1	0,68	0,64	-0,07	0,17	0,50	-0,13	0,43
A2	0,68	1	0,21	0,12	-0,26	0,33	0,08	0,11
A3	0,64	0,21	1	0,32	0,70	-0,17	-0,62 *	0,41
A4	-0,07	-0,12	0,32	1	0,32	-0,42 *	-0,47 *	0,04
A5	0,17	-0,26	0,70	0,32	1	-0,42 *	-0,73 *	0,07
A6	0,50	0,33	-0,17	-0,42 *	-0,42 *	1	0,62	0,37
A7	-0,13	0,08	-0,62 *	-0,47 *	-0,73 *	0,62	1	-0,06
A8	0,43	0,11	0,41	0,04	0,07	0,37	-0,06	1
	0	0	1	2	2	2	3	0

B	B1	B2	B3	B4	B5	B6
B1	1	0,83	0,73	0,70	0,28	0,36
B2	0,83	1	0,90	0,76	0,61	0,57
B3	0,73	0,90	1	0,86	0,44	0,72
B4	0,70	0,76	0,86	1	0,31	0,62
B5	0,28	0,61	0,44	0,31	1	0,60
B6	0,36	0,57	0,72	0,62	0,60	1
	0	0	0	0	0	0

C	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	1	0,86	0,39	0,16	-0,20	-0,31 *	0,23	-0,46 *	-0,30 *
C2	0,86	1	0,29	0,12	-0,25	0,18	0,03	-0,30 *	-0,45 *
C3	0,39	0,29	1	-0,03	0,22	-0,12	0,37	-0,03	0,12
C4	0,16	0,12	-0,03	1	-0,09	-0,13	-0,04	-0,17	-0,12
C5	-0,20	-0,25	0,22	-0,09	1	0,01	0,13	0,11	0,16
C6	-0,31 *	0,18	-0,12	-0,13	0,01	1	-0,37 *	0,42	-0,27 *
C7	0,23	0,03	0,37	-0,04	0,13	-0,37 *	1	-0,07	0,74
C8	-0,46 *	-0,30 *	-0,03	-0,17	0,11	0,42	-0,07	1	0,41
C9	-0,30 *	-0,45 *	0,12	-0,12	0,16	-0,27 *	0,74	0,41	1
	3	2	0	0	0	3	1	2	3

D	D1	D2	D3	D4	D5
D1	1	0,57	0,58	-0,01	0,07
D2	0,57	1	0,13	-0,26	0,75
D3	0,58	0,13	1	0,63	-0,16
D4	-0,01	-0,26	0,63	1	-0,31 *
D5	0,07	0,75	-0,16	-0,31 *	1
	0	0	0	1	1

ekonomicznych", Zeszyty Naukowe AE Kraków, Seria specjalna: Monografie nr 61, Kraków 1984.

Otrzymane realizacje mierników syntetycznych w przekroju badanych układów A-D ^{zebrano} w tabl. 2.6.

Podobnie jak w przypadku czynników głównych znajdujących się w tabl. 2.4 tak i obecnie przytoczone realizacje mierników zostaną wykorzystane w dalszych rozważaniach jako zmienne objaśniające w ekonometrycznych modelach przyczynowo opisowych krajowego ruchu turystycznego oraz jako podstawa do regionalizacji Polski z punktu widzenia poszczególnych mierników odzwierciedlających w sposób syntetyczny determinanty ruchu turystycznego.

2.4 Ekonometryczne modele przyczynowo- opisowe ruchu turystycznego w Polsce

Uzyskane w poprzednich podrozdziałach realizacje czynników głównych F1-F5 oraz mierników syntetycznych $Q_A - Q_D$ wykorzystano w procesie modelowania ekonometrycznego krajowego ruchu turystycznego w ujęciu województw. Jako zmienne objaśniane przyjęto tu poszczególne wskaźniki natężenia ruchu turystycznego E1-E6, podczas gdy zmienne F1-F5 oraz $Q_A - Q_D$ spełniałyby rolę determinant ruchu. Prowadziło to do dwóch rodzajów ekonometrycznych modeli przyczynowo-opisowych:

- czynnikowych:

$$E_j = f(F_1, \dots, F_5, \xi_j) \quad (j=1, \dots, 6) \quad (2.4)$$

- syntetycznych

$$E_j = f(Q_A, Q_B, Q_C, Q_D, \xi_j) \quad (j=1, \dots, 6) \quad (2.5)$$

Nr	Województwo	Układ A	Układ B	Układ C	Układ D
1	Stoż.warszawskie	,853	,668	,844	,752
2	bialskopodlaskie	,068	,804	,260	,244
3	białostockie	,251	,799	,645	,137
4	bielskie	,732	,616	,810	1,142
5	bydgoskie	,834	,730	,888	,414
6	chełmskie	,024	,615	,359	,233
7	ciechanowskie	,234	,814	,304	,337
8	częstochowskie	,535	,721	,636	,202
9	elbląskie	,508	,782	,549	,667
10	gdańskie	,843	,769	,678	,975
11	gorzowskie	,409	,766	,960	,206
12	jeleniogórskie	,607	,520	,699	,884
13	kaliskie	,596	,795	,618	,116
14	katowickie	1,373	,352	,367	,425
15	kieleckie	,616	,754	,625	,341
16	konińskie	,338	,457	,304	,176
17	koszalińskie	,498	,804	,597	,911
18	m.krakowskie	,662	,407	,152	,648
19	krośnieńskie	,085	,776	,975	,296
20	legnickie	,446	,441	,007	,138
21	leszczyńskie	,450	,805	,344	,334
22	lubelskie	,393	,765	,368	,204
23	łomżyńskie	,006	,794	,594	,067
24	m.łódzkie	,220	,492	,511	,427
25	nowosądeckie	,473	,804	,764	1,466
26	olsztyńskie	,567	,812	,716	,406
27	opolskie	,955	,687	,502	,262
28	ostrołęckie	,103	,668	,612	,214
29	pilskie	,409	,812	,717	,234
30	piotrkowskie	,273	,729	,584	,396
31	płockie	,263	,687	,183	,297
32	poznańskie	,792	,779	,522	,527
33	przemyskie	,142	,810	,506	,006
34	radomskie	,316	,656	,481	,119
35	rzeszowskie	,397	,791	,539	,169
36	siedleckie	,312	,817	,583	,116
37	sieradzkie	,208	,764	,358	,127
38	skierniewickie	,203	,775	,308	,086
39	ślupskie	,309	,803	,640	,666
40	suwalskie	,305	,806	,949	,257
41	szczecińskie	,775	,675	,544	,831
42	tarnobrzeskie	,374	,565	,591	,145
43	tarnowskie	,299	,765	,503	,173
44	toruńskie	,554	,779	,371	,193
45	wałbrzyskie	,757	,704	,784	,754
46	włocławskie	,289	,734	,230	,284
47	wrocławskie	,775	,701	,724	,398
48	zamojskie	,298	,800	,426	,135
49	zielonogórskie	,616	,804	,974	,346

Realizacja syntetycznych mierników -determinat krajowego
ruchu turystycznego według układów A - D

Przed przystąpieniem do estymacji parametrów modeli (2.4) i (2.5) wyznaczono proste współczynniki korelacji liniowej między rozpatrywanymi zmiennymi. Współczynniki te zebrano w tabl. 2.7 przy czym w części a) tej tablicy znajdują się współczynniki między zmiennymi zbiorów E-F, w części b) - zbiorów E-Q natomiast w części c) - zbiorów F-Q. Symbolem * oznaczono tu maksymalne co do modułu wartości współczynników korelacji w danym wierszu tablic.

Jak można zauważyć zmienne E1 i E2 determinowane są głównie przez czynnik F3 określający bazę turystyczną natomiast zmienne E3 - E6 przez czynnik F4 opisujący bazę wczasowo-wypoczynkową. Zmienna syntetyczna Q_D stanowiąca wypadkową bazy turystycznej ogółem jest także najważniejszą determinantą krajowego ruchu turystycznego.

Warto odnotować fakt, że zmienna syntetyczna Q_B (zanieczyszczenie środowiska) nie jest istotnie skorelowana z ruchem turystycznym co może sugerować błędną jej konstrukcję. Pozostałe zmienne syntetyczne Q_A i Q_C zgodnie z oczekiwaniami wykazują istotną korelację dodatnią z wielkością krajowego ruchu turystycznego.

Kierunki i siła skorelowań czynników głównych z ruchem turystycznym również nie budzi zastrzeżeń merytorycznych. Czynniki F2 i F5 wykazują istotną ujemną korelację z wielkością ruchu natomiast pozostałe czynniki są na ogół skorelowane dodatnio ze zmiennymi E1-E6.

Zmienne "łączone" E5 i E6 charakteryzują się wyraźniejszymi związkami z wyróżnionymi determinantami ruchu aniżeli zmienne opisujące ruch w sposób bardziej szczegółowy F1-E4.

Tablica 2.7

Współczynniki korelacji liniowej między zmiennymi
ze zbiorów E - F, E - Q, F - Q

a)

Zmienna	F1	F2	F3	F4	F5
E1	0,40	-0,41	0,72 *	0,38	-0,16
E2	0,30	-0,52	0,55 *	0,55 *	-0,30
E3	0,02	-0,55	0,14	0,85 *	-0,47
E4	-0,00	-0,56	0,09	0,86 *	-0,48
E5	0,41	-0,60	0,80	0,89 *	-0,50
E6	0,32	-0,65	0,61	0,91 *	-0,55

b)

Zmienne	A	B	C	D
E1	0,63	-0,19	0,40	0,71 *
E2	0,58	-0,10	0,37	0,73 *
E3	0,32	0,08	0,33	0,86 *
E4	0,29	0,09	0,29	0,83 *
E5	0,70	0,03	0,48	0,90 *
E6	0,66	0,05	0,45	0,88 *

c)

Zmienne	A	B	C	D
F1	0,59	-0,91 *	-0,26	0,25
F2	-0,57	-0,11	-0,64 *	-0,44
F3	0,58	-0,65 *	0,03	0,48
F4	0,32	0,20	0,32	0,74 *
F5	0,06	-0,69 *	-0,54	-0,30

Część c) tabl.2.7 pozwala ustalić wzajemne powiązania między czynnikami głównymi a zmiennymi syntetycznymi. Łatwo można zauważyć, że zmienna Q_B nie jest jednorodna gdyż jest najsilniej skorelowana z trzema różnymi czynnikami głównymi F_1 , F_3 , F_5 i to z jednakowym, ujemnym kierunkiem, co w przypadku współczynnika $r_{Q_B F_3}$ trudno jest uzasadnić merytorycznie. Zmienna syntetyczna Q_C (walory turystyczne) jest najsilniej skorelowana z czynnikiem F_2 (atrakcyjność turystyczna) natomiast Q_D z czynnikiem F_4 (obydwa te agregaty opisują bazę turystyczną). Zmienna syntetyczna Q_A (infrastruktura turystyki) wykazuje istotne skorelowanie z większością czynników głównych przy czym kierunek tego skorelowania jest zgodny z oczekiwaniami.

Z kolei przystąpiono do budowy ekonometrycznych modeli przyczynowo-opisowych (2.4) i (2.5). Jako metodę pomocniczą wykorzystano tu metodę pojemności indywidualnych i integralnych Hellwiga w celu wyboru najbardziej optymalnego podzbioru determinant. W metodzie tej dla każdej możliwej kombinacji zmiennych objaśniających wyznacza się wskaźniki pojemności indywidualnej :

$$q_{1j} = \frac{r_j^2}{1 + \sum_{i \neq j} |r_{ij}|} \quad \begin{matrix} (i, j=1, \dots, k \\ l=1, \dots, 2^k - 1) \end{matrix} \quad (2.6)$$

gdzie r_j to współczynnik korelacji j-tej zmiennej objaśniającej ze zmienną objaśnianą, r_{ij} - to współczynnik korelacji między i-tą oraz j-tą zmienną objaśniającą, natomiast, k - liczba zmiennych objaśniających. Następnie dla każdej spośród $2^k - 1$ kombinacji zmiennych objaśniających ustala się wskaźnik pojemności integralnej:

$$G_1 = \sum_1 g_{1j} \quad (2.7)$$

który stanowi kryterium wyboru optymalnego podzbioru zmiennych. Wybiera się tu taką kombinację zmiennych objaśniających dla której G_1 jest największy, przy czym wybór ten dokonywany jest odrębnie w klasie kombinacji 1, 2, 3 i 4-elementowych, a ponadto uwzględnione są po 3 pierwsze "najlepsze" kombinacje.

Do analizy przyjęto liniowe funkcje regresji typu (1.7), których parametry szacowano metodą najmniejszych kwadratów według wzoru (1.8). Poza parametrami wyznaczono także odpowiadające im statystyki Studenta (1.16), statystykę F Fishera-Snedecora (1.18) oraz dwa podstawowe miary dopasowania funkcji do danych empirycznych: współczynnik zmienności resztowej V (wzór 1.13) oraz współczynnik korelacji wielorakiej R (wzór 1.14). Wartości powyższych parametrów zebrano w tabl. 2.8 - 2.12 zawierających odpowiednio informacje o:

- tabl. 2.8: modelach czynnikowych i syntetycznych z jedną zmienną objaśniającą,
- tabl. 2.9: metodach czynnikowych i syntetycznych z dwoma zmiennymi objaśniającymi,
- tabl. 2.10: modelach czynnikowych z trzema zmiennymi objaśniającymi,
- tabl. 2.11: modelach syntetycznych z trzema zmiennymi objaśniającymi,
- tabl. 2.12: modelach czynnikowych i syntetycznych z czterema zmiennymi objaśniającymi.

Tablica 2.8

Parametry modeli czynnikowych i syntetycznych ruchu
turystycznego z jedną zmienną objaśniającą

Ruch turyst.	Modele czynnikowe				Modele syntetyczne			
	parametr	V	R	F	parametr	V	R	F
E1	F3: 0,72 7,1	68,0	0,72	50,5	D: 0,71 6,9	68,9	0,71	47,8
	F2: -0,41 -3,1	89,3	0,41	9,6	A: 0,63 5,6	75,8	0,63	31,4
	F1: 0,41 3,1	89,5	0,41	9,4	C: 0,40 3,0	89,8	0,40	8,8
E2	F4: 0,55 4,5	77,6	0,55	20,4	D: 0,73 7,4	63,8	0,73	54,2
	F3: 0,54 4,3	78,4	0,54	18,9	A: 0,58 4,9	75,7	0,58	23,7
	F2: -0,53 -4,3	78,9	0,53	18,1	C: 0,37	86,4	0,37	7,4
E3	F4: 0,85 11,1	81,4	0,85	123,8	D: 0,86 11,7	78,4	0,86	137,1
	F2: -0,56 -4,6	128,8	0,56	21,2	C: 0,33 2,4	146,4	0,33	5,8
	F5: -0,46 -3,6	137,3	0,46	13,0	A: 0,32 2,3	146,9	0,32	5,4
E4	F4: 0,86 11,6	85,6	0,86	135,8	D: 0,83 10,3	93,6	0,83	105,9
	F2: -0,55 -4,5	141,0	0,55	20,3	C: 0,29 2,1	161,4	0,29	4,4
	F5: -0,48 -3,8	147,9	0,48	14,2	A: 0,29 2,1	161,6	0,29	4,3
E5	F4: 0,76 8,2	70,9	0,77	67,2	D: 0,93 16,4	42,5	0,92	269,9
	F2: -0,57 -4,8	90,4	0,57	23,1	A: 0,51 4,0	95,2	0,51	16,3
	F3: 0,42 3,1	100,4	0,42	9,9	C: 0,41 3,1	100,8	0,41	9,5
E6	F4: 0,84 10,7	62,7	0,84	115,7	D: 0,89 13,7	52,4	0,89	187,5
	F2: -0,61 -5,3	92,7	0,61	25,5	A: 0,44 3,3	104,9	0,44	11,2
	F5: 0,47 -3,6	103,2	0,47	13,1	C: 0,36 2,6	108,9	0,36	6,9

Tablica 2.9

Parametry modeli czynnikowych i syntetycznych ruchu
turystycznego z dwiema zmiennymi objaśniającymi

Ruch tur.	Modele czynnikowe					Modele syntetyczne				
	P a r a m e t r y		V	R	F	P a r a m e t r y		V	R	F
F1	F2: -0,45 -5,8	F3: 0,74 9,5	52,4	0,85	59,3	A: 0,37 3,5	D: 0,52 4,9	61,9	0,78	35,7
	F3: 0,76 9,6	F4: 0,44 5,6	52,9	0,84	57,6	C: 0,20 1,9	D: 0,65 6,2	67,2	0,73	26,9
	F3: 0,87 10,1	F5: -0,45 -5,3	54,4	0,84	53,3	B: -0,12 -1,1	D: 0,70 6,8	68,8	0,72	24,7
E2	F3: 0,59 6,7	F4: 0,60 6,8	55,8	0,80	42,0	A: 0,29 2,6	D: 0,59 5,5	59,6	0,77	34,0
	F2: -0,56 -5,9	F3: 0,56 6,0	59,7	0,77	33,9	C: 0,16 1,5	D: 0,68 6,6	62,4	0,75	28,9
	F1: 0,35 3,1	F4: 0,58 5,1	71,4	0,65	16,8	B: -0,03 -0,3	D: 0,73 7,2	63,9	0,73	26,6
E3	F3: 0,21 3,0	F4: 0,90 12,2	75,2	0,88	76,9	B: 0,17 2,4	D: 0,88 12,5	74,8	0,88	78,2
	F1: 0,09 1,1	F4: 0,73 11,2	81,1	0,86	62,9	C: 0,07 0,9	D: 0,84 10,8	78,5	0,87	68,7
	F2: 0,06 0,5	F4: 0,89 8,4	82,0	0,85	61,1	A: -0,14 -1,7	D: 0,93 11,2	76,8	0,87	72,9
E4	F3: 0,16 2,3	F4: 0,88 12,3	81,9	0,88	76,6	B: 0,19 2,4	D: 0,85 11,0	89,1	0,85	61,3
	F1: 0,06 0,8	F4: 0,87 11,6	85,9	0,86	67,8	C: 0,04 0,4	D: 0,82 9,6	94,4	0,83	52,1
	F2: 0,08 0,8	F4: 0,92 9,0	85,9	0,86	67,8	A: -0,17 -1,8	D: 0,91 10,1	91,3	0,84	57,4
E5	F3: 0,49 7,8	F4: 0,81 13,0	46,8	0,91	107,7	C: 0,14 2,4	D: 0,88 15,7	40,5	0,93	151,9
	F1: 0,26 2,9	F4: 0,79 9,0	65,7	0,81	43,4	B: 0,07 1,2	D: 0,93 16,5	42,4	0,93	136,8
	F2: -0,09 -0,7	F4: 0,71 5,4	71,3	0,77	33,4	A: 0,06 1,0	D: 0,89 13,8	42,6	0,92	135,4
E6	F3: 0,35 5,8	F4: 0,87 14,5	48,0	0,91	115,8	B: 0,12 2,0	D: 0,91 14,2	50,7	0,90	101,3
	F1: 0,18 2,4	F4: 0,86 11,5	59,6	0,86	67,0	C: 0,09 1,3	D: 0,87 12,7	51,8	0,90	96,1
	F2: -0,05 -0,4	F4: 0,81 7,4	63,3	0,84	57,0	A: -0,01 -0,1	D: 0,90 11,8	52,8	0,89	91,8

Tablica 2.10

Parametry modeli czynnikowych ruchu turystycznego
z trzema zmiennymi objaśniającymi

Ruch tur.	P a r a m e t r y m o d e l i			V	R	F
E1	F2: -0,28 -2,7	F3: 0,76 10,2	F4: 0,25 2,5	49,6	0,87	46,1
	F1: -0,56 -5,5	F2: -0,49 -8,0	F3: 1,19 11,7	40,9	0,91	74,6
	F2: -0,32 -3,9	F3: 0,83 10,9	F5: -0,29 -3,3	47,5	0,88	51,7
E2	F2: -0,27 -2,3	F3: 0,59 7,0	F4: 0,41 3,6	53,2	0,83	32,6
	F1: -0,33 -2,4	F3: 0,86 6,1	F5: 0,60 7,1	53,2	0,83	32,7
	F1: -0,47 -3,3	F2: -0,59 -6,9	F3: 0,93 6,6	54,2	0,82	31,0
E3	F1: -0,22 -2,0	F3: 0,39 3,4	F4: 0,87 12,6	73,0	0,89	55,8
	F2: 0,06 0,6	F3: 0,21 3,0	F4: 0,91 9,2	75,7	0,88	50,7
	F1: 0,09 1,2	F2: -0,07 0,7	F4: 0,91 8,5	81,6	0,86	41,6
E4	F1: -0,18 -1,6	F3: 0,31 2,7	F4: 0,87 12,5	80,7	0,88	53,6
	F2: 0,09 0,9	F3: 0,16 2,3	F4: 0,94 9,5	82,1	0,88	51,1
	F1: 0,07 0,9	F2: 0,09 0,9	F4: 0,93 9,0	85,9	0,87	45,4
E5	F2: -0,08 -0,9	F3: 0,49 7,8	F4: 0,75 8,8	46,9	0,91	71,8
	F1: -0,36 -4,0	F3: 0,77 8,7	F4: 0,81 14,9	40,7	0,93	100,5
	F1: 0,25 2,9	F2: -0,04 -0,4	F4: 0,76 6,2	66,3	0,81	28,5
E6	F2: -0,04 -0,5	F3: 0,35 5,8	F4: 0,84 10,1	48,4	0,91	76,1
	F1: -0,26 -2,9	F3: 0,56 6,1	F4: 0,87 15,5	44,7	0,93	91,9
	F1: 0,18 2,4	F2: -0,02 -0,2	F4: 0,84 8,0	60,3	0,86	43,7

Tablica 2.11

Parametry modeli syntetycznych rachunku turystycznego
z trzema zmiennymi objaśniającymi

Ruch tur.	Parametry modeli			V	R	F
E1	A: 0,36 3,4	C: 0,17 1,8	D: 0,48 4,5	60,6	0,79	25,9
	A: 0,37 3,2	B: -0,01 -0,1	D: 0,53 4,9	62,6	0,78	23,3
	B: -0,22 -2,1	C: 0,28 2,6	D: 0,60 5,8	64,9	0,76	20,7
E2	A: 0,27 2,5	C: 0,13 1,4	D: 0,56 5,1	59,1	0,78	23,7
	A: 0,31 2,7	B: 0,07 0,7	D: 0,58 5,4	60,0	0,77	22,6
	B: -0,10 -0,9	C: 0,19 1,7	D: 0,66 6,2	62,5	0,75	19,5
E3	B: 0,17 2,2	C: 0,01 0,1	D: 0,88 11,4	75,4	0,88	51,0
	A: -0,15 -1,8	C: 0,08 1,0	D: 0,91 10,7	76,6	0,88	49,2
	A: -0,09 -1,0	B: 0,14 1,9	D: 0,92 11,3	74,8	0,88	52,5
E4	B: 0,20 2,4	C: -0,04 -0,5	D: 0,87 10,3	89,9	0,85	40,2
	A: -0,17 -1,9	C: 0,05 0,62	D: 0,90 9,6	91,9	0,85	37,8
	A: -0,11 -1,1	B: 0,15 1,9	D: 0,90 10,1	88,9	0,86	41,5
E5	A: 0,05 0,8	C: 0,13 2,3	D: 0,86 13,5	40,6	0,93	100,8
	B: 0,03 0,4	C: 0,13 2,1	D: 0,89 15,2	40,9	0,93	99,4
	A: 0,10 1,5	B: 0,10 1,7	D: 0,88 13,8	41,8	0,93	94,5
E6	B: 0,11 1,6	C: 0,05 0,7	D: 0,89 12,9	51,0	0,90	66,8
	A: -0,02 -0,2	C: 0,09 1,3	D: 0,87 11,3	52,4	0,90	62,8
	A: 0,05 0,6	B: 0,14 2,0	D: 0,88 12,0	51,1	0,90	66,7

Tablica 2.12

Parametry modeli czynnikowych i syntetycznych ruchu
turystycznego z czterema zmiennymi syntetycznymi

Ruch tur.	Parametry modelu				V	R	F				
E1	A:	0,31 2,7	B:	-0,09 -0,9	C:	0,21 2,0	D:	0,48 4,4	60,7	0,80	19,6
	F1:	-0,52 -5,4	F2:	-0,36 -4,4	F3:	1,17 12,1	F4:	0,19 2,4	38,9	0,92	63,1
	F1:	-0,51 -3,9	F2:	-0,46 -5,8	F3:	1,17 10,8	F5:	-0,05 -0,6	41,3	0,91	55,2
E2	A:	0,28 2,3	B:	0,01 0,1	C:	0,13 1,2	D:	0,56 5,0	59,7	0,78	17,4
	F1:	-0,41 -3,1	F2:	-0,34 -3,1	F3:	0,91 7,1	F4:	0,37 3,4	48,7	0,86	31,7
	F2:	-0,24 -2,1	F3:	0,65 7,5	F4:	0,32 2,6	F5:	-0,22 -2,0	51,5	0,84	27,2
E3	A:	-0,09 -1,0	B:	0,13 1,5	C:	0,03 0,3	D:	0,92 10,9	75,5	0,88	38,6
	F1:	-0,22 -1,9	F2:	0,03 0,3	F2:	0,39 3,3	F4:	0,89 9,1	73,7	0,89	41,0
	F2:	0,07 0,7	F3:	0,23 3,0	F4:	0,88 8,2	F5:	-0,06 -0,6	76,2	0,88	37,7
E4	A:	-0,10 -1,0	B:	0,16 1,8	C:	-0,02 -0,2	D:	0,90 9,8	89,8	0,86	30,5
	F1:	-0,17 -1,4	F2:	0,06 0,6	F3:	0,30 2,5	F4:	0,92 9,3	81,2	0,88	39,8
	F2:	0,10 1,0	F3:	0,18 2,3	F4:	0,91 8,4	F5:	-0,06 -0,6	82,7	0,88	37,9
E5	A:	0,07 1,0	B:	0,05 0,8	C:	0,11 1,8	D:	0,86 13,4	40,8	0,93	75,0
	F1:	-0,39 -4,4	F2:	-0,14 -1,9	F3:	0,79 9,1	F4:	0,71 9,7	39,5	0,94	80,7
	F1:	-0,36 -3,3	F3:	0,77 8,3	F4:	0,81 11,1	F5:	0,01 0,1	41,1	0,93	73,7
E6	A:	0,04 0,05	B:	0,12 1,6	C:	0,04 0,5	D:	0,08 11,5	51,5	0,90	49,3
	F1:	-0,28 -3,0	F2:	-0,09 -1,1	F3:	0,58 6,2	F4:	0,81 10,4	44,5	0,93	69,7
	F2:	-0,02 -0,3	F3:	0,39 6,1	F4:	0,78 8,7	F5:	-0,13 -1,6	47,5	0,92	59,9

W tablicach tych zamiast oryginalnych wartości ocen parametrów funkcji regresji α_j podano je w postaci przekształconej:

$$\beta_j = \alpha_j \frac{s_{x_j}}{s_Y} \quad (j=1, \dots, k) \quad (2.8)$$

gdzie s_{x_j} , s_Y to odchylenia standardowe j -tej zmiennej objaśniającej X_j oraz zmiennej objaśnianej Y . Parametry β_j są współczynnikami regresji standaryzowanej i z uwagi na znormalizowany zakres zmienności są bardziej przydatne do oceny porównawczej stopnia cząstkowego skorelowanie poszczególnych zmiennych objaśniających ze zmienną Y .

Wartości statystyk Studenta t_j podane są poniżej ocen parametrów β_j . Wartość krytyczna testu Studenta dla poziomu istotności $\alpha = 0,05$ oraz $n-k-1 \approx 44-47$ stopni swobody wynosi ok. 2,01. Wartość krytyczna testu Fishera-Snedecora F_α dla $\alpha = 0,05$ oraz stopni swobody $n-k \approx 45-47$ wynosi odpowiednio dla modeli:

- jednoczynnikowych : 19,5
- dwuczynnikowych: 8,6
- trzyczynnikowych: 5,7
- czteroczynnikowych: 4,5

Należy jednak zaznaczyć, że w praktyce aby można było przyjąć hipotezę o istotności równania regresji to empiryczna wartość testu Fishera-Snedecora powinna dwu- a nawet trzykrotnie przekraczać wartość krytyczną F_α .

Analizując otrzymane ekonometryczne modele krajowego ruchu turystycznego warto zwrócić uwagę na następujące wnioski:

- 1) Stopień dopasowania funkcji regresji do danych empirycznych jest bardzo słaby. Współczynniki zmienności resztowej V

przekraczają ogół 50%, a tylko w nielicznych przypadkach (zwłaszcza dla modeli dotyczących zmiennych "łącznych" E5 i E6) zawierają się one ~~nie~~ w granicach 40-50%.

2) Pomimo słabej zgodności modeli z zaobserwowanymi wielkościami krajowego ruchu turystycznego (co jak się wydaje można tłumaczyć dużym jego zróżnicowaniem przestrzennym) współczynniki korelacji wielorakiej R ^{krotność} ~~konkretnie~~ tują się na wysokim poziomie, często przekraczając wartość 0,9. Oznacza to, że przyjęte do modelu determinanty w zadawalający sposób określają kształtowane się ruchu turystycznego wyjaśniając ok. 80% całkowitej jego zmienności.

3) We wszystkich modelach cząsteczkowych tylko co najwyżej 3 współczynniki regresji są istotnie różne od zera. Modele te są więc mniej przydatnym narzędziem analizy ruchu turystycznego aniżeli funkcje regresji z mniejszą liczbą zmiennych objaśniających.

4) Również i ograniczenie rozważań do modeli jednoczynnikowych nie wydają się być najwłaściwszym rozważaniem z uwagi na tkwiące w nich zbyt duże uproszczenia. Prowadzi to do wniosku, że dla opisu kształtowania się krajowego ruchu turystycznego w ujęciu województw niezbędne są modele dwu- lub trzyczynnikowe.

5) Relatywnie najlepsze modele otrzymano dla zmiennych "łącznych" E5-E6, a w dalszej kolejności zmiennych charakteryzujących ruch w obiektach turystycznych E1-E2. Najgorsze modele odnoszą się do ruchu w obiektach czasowo-wypoczynkowych E3-E4.

6) Liczba korzystających z noclegów (zmienne E1, E3, E5) łatwiej poddaje się procesowi modelowania ekonometrycznego aniżeli liczba udzielonych noclegów (E2, E4, E6).

7) Najsilniej oddziałują na wielkość krajowego ruchu turystycznego zmienne określające bazę turystyczną (zmienna syntetyczna Q_D lub czynniki F_3, F_4). Jak można było przewidywać w modelach czynnikowych zmiennych E_1 i E_2 podstawową rolę odgrywa baza turystyczna F_3 , a w modelach zmiennych E_3 i E_4 - baza czasowo-wypoczynkowa F_4 .

8) Następną po bazie turystyki istotną determinantą ruchu turystycznego (E_1, E_2) w modelach syntetycznych jest infrastruktura turystyki (Q_A) a dla ruchu czasowo-wypoczynkowego (E_3, E_4) - zanieczyszczenie środowiska (Q_B). W przypadku modeli czynnikowych ruchu turystycznego (E_1, E_2) poza czynnikami F_3 i F_4 w równaniach uwzględnione są również dwa pierwsze czynniki F_1 i F_2 przy czym wpływ dostępności i atrakcyjności F_2 jest większy niż czynnika F_1 - zanieczyszczenie środowiska. W modelach czynnikowych ruchu czasowo-wypoczynkowego (E_3, E_4) nie obserwuje się istotnego oddziaływania czynnika F_2 natomiast czynnik F_1 ^{model} ~~model~~ zachowuje tu znaczenie jakkolwiek w mniejszym stopniu niż w przypadku ruchu turystycznego.

9) Analiza koincydencji znaków współczynników regresji cząstkowej ze współczynnikami korelacji prostej nie budzi zastrzeżeń, zwłaszcza gdy obydwa współczynniki są istotnie różne od zera. Niekiedy obserwuje się częściowy brak koincydencji, w sytuacji gdy jeden z tych współczynników jest statystycznie nieistotny a równocześnie drugi wskazuje na obecność zależności. Przypadki takie spotyka się jednak bardzo rzadko.

2.5 Regionalizacja Polski na podstawie zmiennych agregatowych

Obeenie przedstawia się wyniki wykorzystania otrzymanych poprzednio zmiennych agregatowych do regionalizacji Polski. Rozważania ograniczone zostaną do zmiennych syntetycznych jakkolwiek analogiczną analizę można byłoby przeprowadzić na podstawie realizacji czynników głównych.

Punktem wyjścia są tu uporządkowane malejąco wartości syntetycznych $Q_A - Q_D$ zebrane w tabl. 2.13. Obok mierników podano odpowiadające im symbole identyfikacji województw.

Uzyskane uporządkowania pozwalają na podział województw na grupy zawierające województwa o zbliżonym poziomie danego miernika syntetycznego. Algorytm grupowania jest w tym przypadku bardzo prosty, gdyż przebiega ono w przestrzeni jednowymiarowej. Podstawą podziału mogą być wartości następujących wskaźników:

$$p_i = \tilde{q}_{i+1} - \tilde{q}_i \quad (i=1, \dots, n-1) \quad (2.8)$$

$$p'_i = \frac{p_i}{\tilde{q}_i} \cdot 100 \quad (i=1, \dots, n-1) \quad (2.9)$$

gdzie $\tilde{q}_i$ to i -ta wartość uporządkowanego miernika syntetycznego.

Punkty w których obserwuje się szczególnie wysokie wartości powyższych mierników wyznaczają miejsca o małym zagęszczeniu województw. Miejsca te stanowią więc naturalne granice podziału zbioru województw, prowadząc do takich grup, że wchodzące w ich skład województwa mają zbliżone wartości mierników, a jednocześnie możliwie najbardziej różne względem województw z innych grup.

Tablica 2.13

Kolejność Województw według syntetycznych mierników -
determinant krajowego ruchu turystycznego (układy A - D)

NRK	KOD	UKŁAD A	KOD	UKŁAD B	KOD	UKŁAD C	KOD	UKŁAD D
1	KA	1,373	SE	,817	KS	,975	NS	1,466
2	OP	,955	CI	,814	ZG	,974	BB	1,142
3	WA	,853	PI	,812	GO	,960	GD	,975
4	GD	,843	OL	,812	SU	,949	KO	,911
5	BY	,834	PR	,810	BY	,888	JG	,884
6	PO	,792	SU	,806	WA	,844	SZ	,831
7	WR	,775	LE	,805	BB	,810	WB	,754
8	SZ	,775	ZG	,804	WB	,784	WA	,752
9	WB	,757	NS	,804	NS	,764	EL	,667
10	BB	,732	KO	,804	WR	,724	SL	,666
11	KR	,662	BP	,804	PI	,717	KR	,646
12	ZG	,616	SL	,803	OL	,716	PO	,527
13	KI	,616	ZA	,800	JG	,699	LD	,427
14	JG	,607	BK	,799	GD	,678	KA	,425
15	KL	,596	KL	,795	BK	,645	BY	,414
16	OL	,567	LO	,794	SL	,640	OL	,406
17	TO	,554	RZ	,791	CZ	,636	WR	,398
18	CZ	,535	EL	,782	KI	,625	PT	,396
19	EL	,508	TO	,779	KL	,618	ZG	,346
20	KO	,498	PO	,779	OS	,612	KI	,341
21	NS	,473	KS	,776	KO	,597	CI	,337
22	LE	,450	SK	,775	LO	,594	LE	,334
23	LG	,446	GD	,769	TG	,591	PL	,297
24	GO	,409	GO	,766	PT	,584	KS	,296
25	PI	,409	LU	,765	SE	,583	WL	,284
26	LU	,393	TA	,765	EL	,549	OP	,262
27	TG	,374	SI	,764	SZ	,544	SU	,257
28	KN	,338	KI	,754	RZ	,539	BH	,244
29	RA	,316	WL	,734	PO	,522	PI	,234
30	SE	,312	BY	,730	LD	,511	CH	,233
31	SL	,309	PT	,729	PR	,506	OS	,214
32	RZ	,307	CZ	,721	TA	,503	GO	,206
33	SU	,305	WB	,704	OP	,502	LU	,204
34	TA	,299	WR	,701	RA	,431	CZ	,202
35	ZA	,298	OP	,687	ZA	,426	TO	,193
36	WL	,289	PL	,687	TO	,371	KN	,176
37	PT	,273	SZ	,675	LU	,368	TA	,173
38	PL	,263	OS	,668	KA	,367	RZ	,169
39	BK	,251	WA	,668	CH	,359	TG	,145
40	CI	,234	RA	,656	SI	,358	LG	,138
41	LD	,220	BB	,616	LE	,344	BK	,137
42	SI	,208	CH	,615	SK	,308	ZA	,135
43	SK	,203	TG	,565	CI	,304	SI	,127
44	PR	,142	JG	,520	KN	,304	RA	,119
45	OS	,103	LD	,492	BP	,260	SE	,116
46	KS	,085	KN	,457	WL	,230	KL	,116
47	BP	,068	LG	,441	PL	,183	SK	,086
48	CH	,024	KR	,407	KR	,152	LO	,067
49	LO	,006	KA	,002	LG	,007	PR	,006

Podstawowym problemem jaki się tu pojawia jest kwestia ustalenia liczby grup na które ma być podzielony cały zbiór województw. Możliwe są tu różne rozwiązania formalne (np. przyjęcie wszystkich maksimów lokalnych wskaźników p_i lub p'_i , ustalenie progowej wartości tych wskaźników w oparciu o charakterystyki ich rozkładu itd). tym niemniej wydaje się, że do stosunkowo dobrych rezultatów prowadzą tu decyzje arbitralne wsparte oceną merytoryczną otrzymywanych grup.

W rozpatrywanym przypadku przyjęto, że liczba grup wynosi 5-6. Granica poszczególnych grup zaznaczono w tabl. 2.13 poziomymi kreskami. Przestrzenne umiejscowienie województw należących do wyróżnionych grup zilustrowane jest na komputerowych kartogramach zaznaczonych na rys. 3 i 4. Na rysunkach tych każda grupa województw oznaczana jest odrębnym symbolem graficznym.

Jak można zauważyć dla infrastruktury turystyki szczególnie wysokie wartości mierników syntetycznych obserwuje się w województwach: katowickim, opolskim, warszawskim, gdańskim i bydgoskim, natomiast szczególnie małe - w województwach: łomżyńskim, chełmskim, białkopodlaskim, krośnieńskim, ostrołęckim i przemyskim.

Najbardziej zagrożone środowisko naturalne mają województwa katowickie, krakowskie, legnickie, konińskie, łódzkie, tarnobrzskie oraz jeleniogórskie. Większość województw z tego punktu widzenia tworzy jednorodną grupę charakteryzującą się relatywnie niskimi wartościami mierników syntetycznych.

Grupę województw o największych walorach turystycznych i wczasowo-wypoczynkowych tworzą województwa: krośnieńskie,

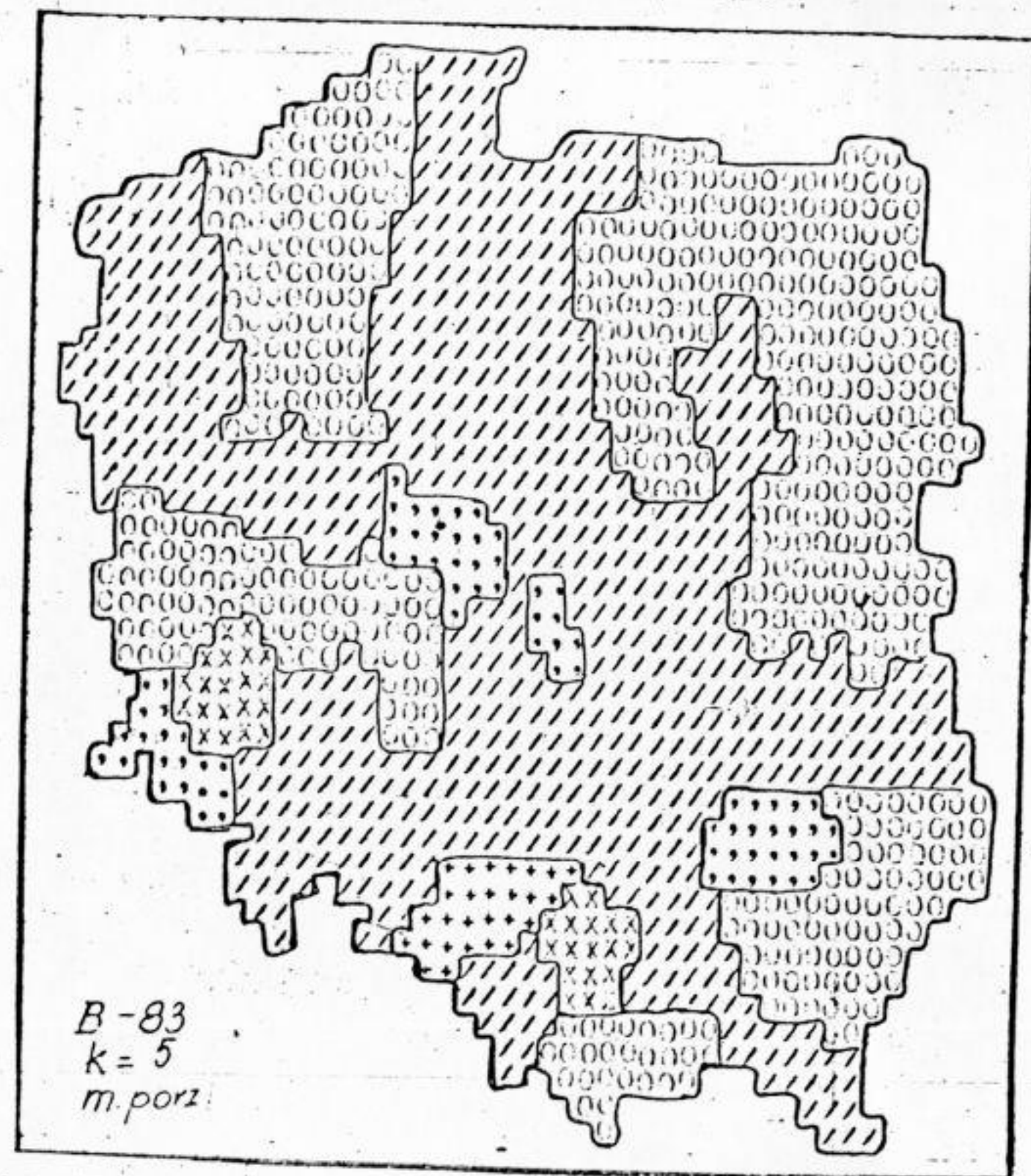
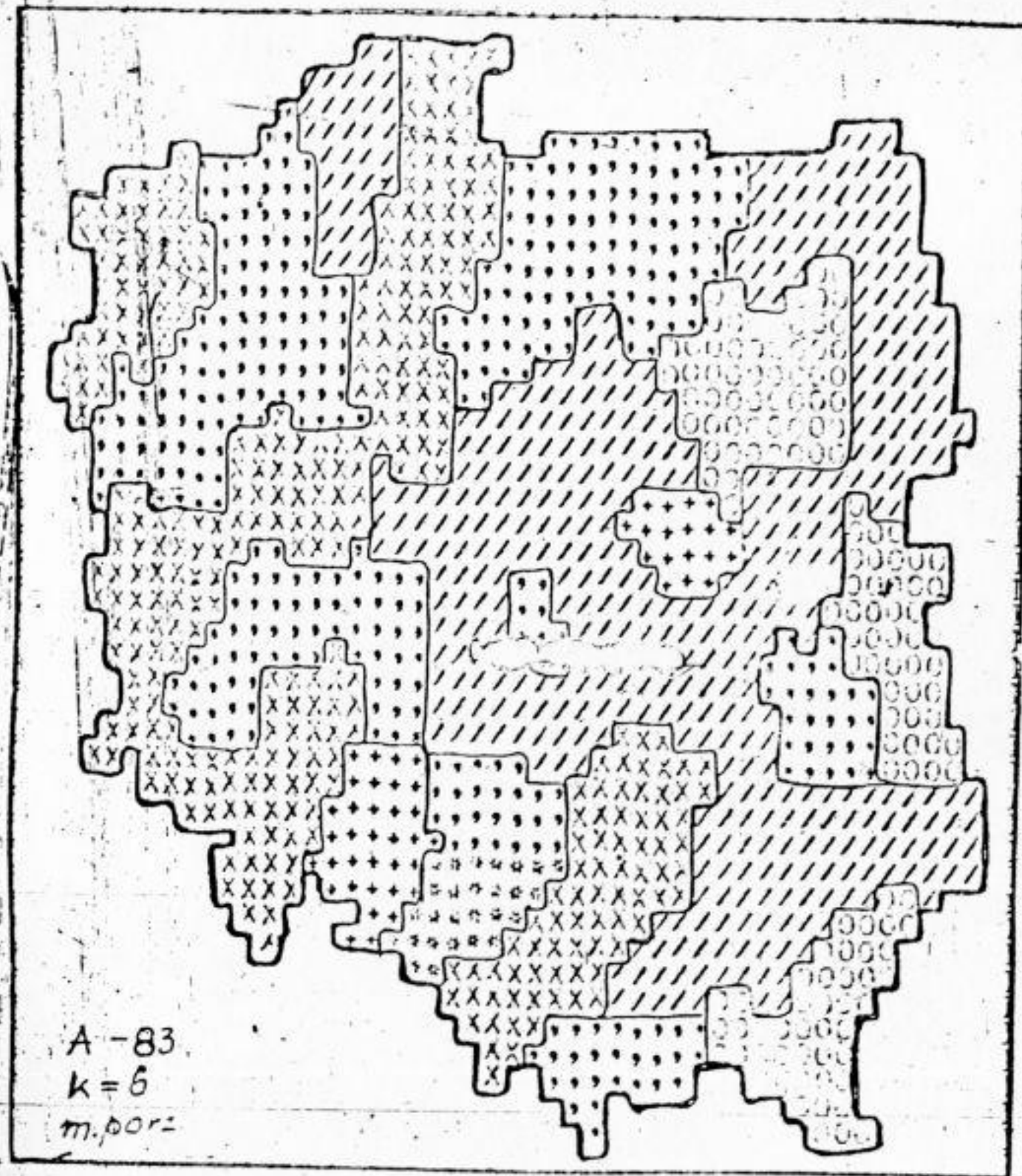
zielonogórskie, gorzowskie i suwalskie. Najmniej walorów występuje w województwach: legnickim, krakowskim, wrocławskim, białskopodlaskim, konińskim, ciechanowskim, skiernowickim.

Do województw o najbardziej rozwiniętej bazie turystycznej należą: nowosądeckie, bielsko-bialskie, gdańskie, koszalińskie, jeleniogórskie, szczecińskie. Natomiast województwa: przemyskie, łódzkie i skierniewickie mają najniższe syntetyczne mierniki bazy turystycznej.

Wymienione tu skrajne grupy województw w ramach każdego z badanych układów są zgodne z oczekiwaniami, co wskazuje na przydatność mierników syntetycznych do ustalenia rejonów turystyki krajowej. Do zalet tego podejścia należy niewątpliwie zaliczyć jego prostotę przy zachowaniu wielowymiarowości wyników grupowania.

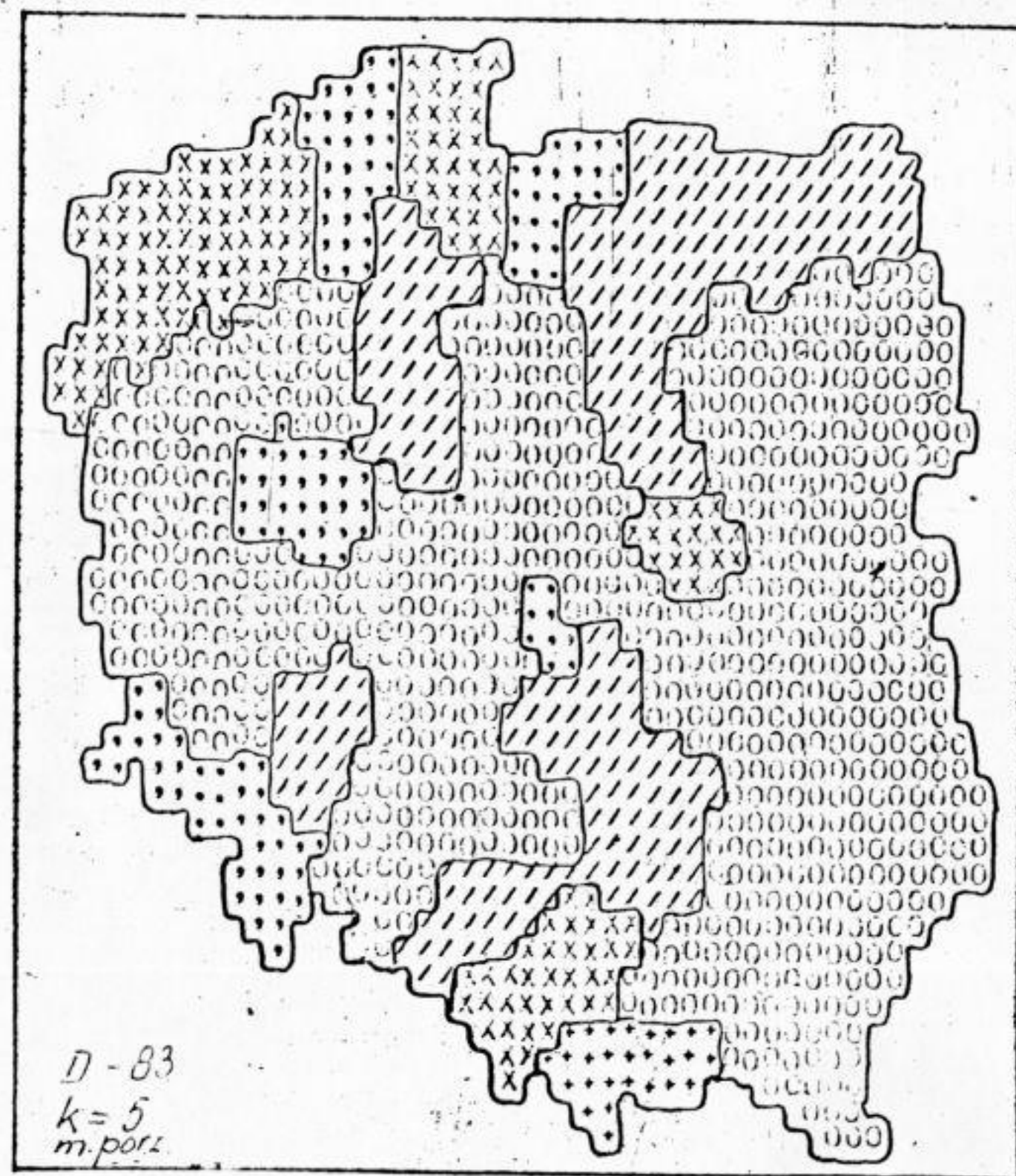
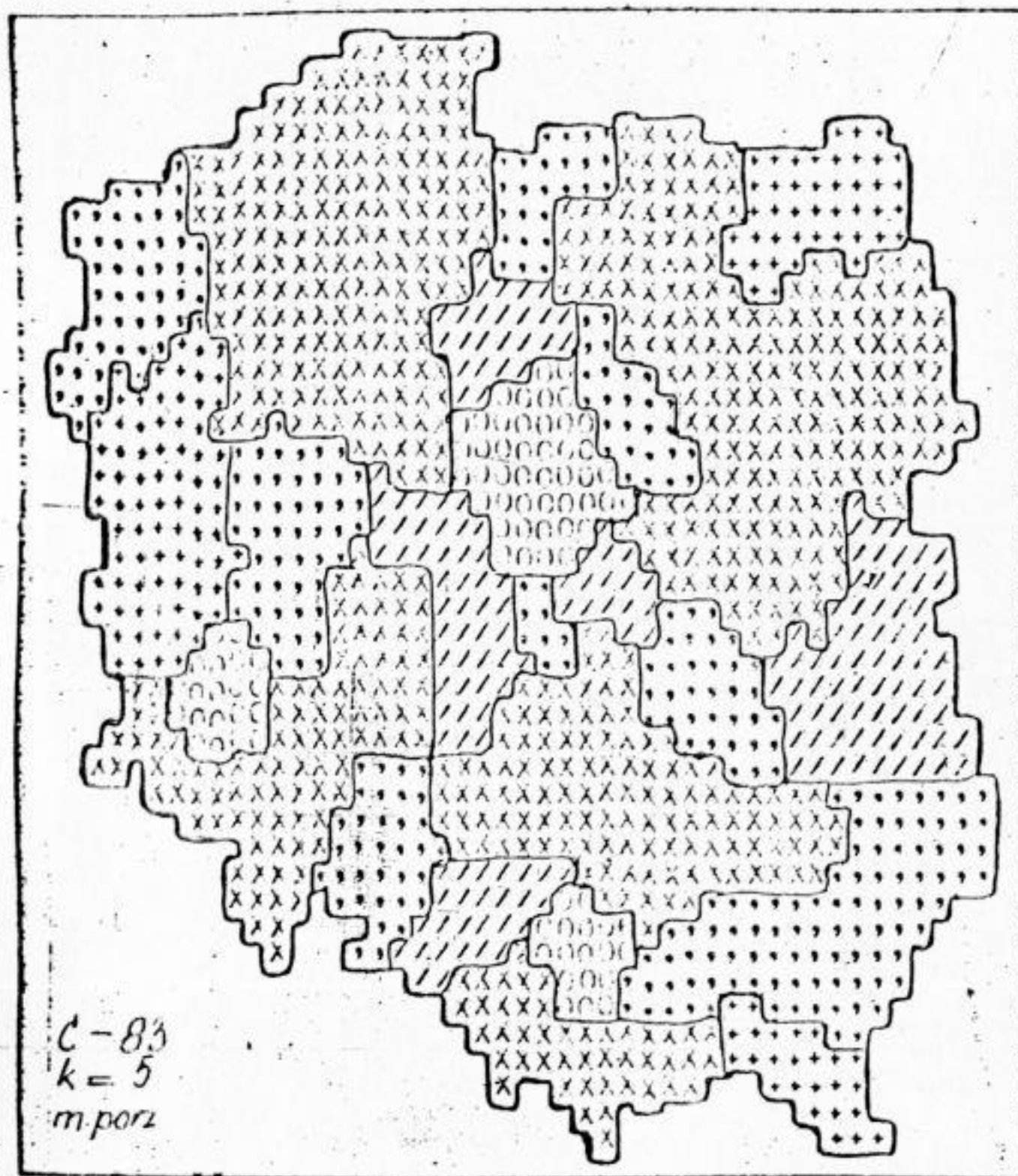
Dalsze problemy związane z regionalizacją Polski w zakresie krajowego ruchu turystycznego będą rozpatrywane w następnych etapach tematu. W szczególności podjęte zostaną kwestie:

- charakterystyki opisowej poszczególnych regionów,
- oceny stopnia jednorodności otrzymywanych grup województw w porównaniu z jednorodnością całego ich zbioru,
- analizy zgodności wyników rejonizacji w przekroju różnych układów,
- porównania rozbieżności w składach grup otrzymywanych na podstawie mierników syntetycznych, czynników głównych oraz ~~cech~~ cech pierwotnych.



Rys. 3. Grupy województw podobnych z punktu widzenia infrastruktury turystyki (A)

oraz zanieczyszczenie środowiska (B)



Rys. 4. Grupy województw podobnych z punktu widzenia walorów turystycznych (C) oraz bazy turystyki (D)