

KIWIEL



**POLSKA AKADEMIA NAUK**  
**Instytut Badań Systemowych**

# **WSPOMAGANIE DECYZJI**

# **SYSTEMY EKSPERCKIE**

pod redakcją

**Romana Kulikowskiego i Lucyny Bogdan**

Warszawa 1995

# METODY CZYNNIKOWE W BADANIACH REGIONALIZACYJNYCH W TRANSPORCIE\*

Barbara Baran-Jarosz, Hanna Bury  
Instytut Badań Systemowych PAN

W pracy rozpatrywany jest problem zastosowania wielozmiennej analizy statystycznej w badaniach regionalizacyjnych w dziedzinie transportu. Zjawiska występujące w zagadnieniach transportowych zaliczane są do zjawisk złożonych, opisywanych za pomocą wielu zmiennych. Zastosowano metody czynnikowe do analizy danych opisujących wybrane zjawiska społeczno-ekonomiczne w ujęciu regionalnym ze szczególnym uwzględnieniem elementów oddziałujących na system transportowy.

Przedstawiono przykład zastosowania metody składowych głównych do wielocechowej analizy i klasyfikacji typologicznej Polski w kategoriach zintegrowanych cech społeczno-ekonomicznych. Wynika to z założenia, że wielkość, strukturę i dynamikę generowanych w regionach potoków transportowych można rozważać jako funkcje czynników charakteryzujących stopień rozwoju społeczno-gospodarczego regionów.

Otrzymane wyniki wykorzystano do wyodrębnienia istotnych cech społeczno-gospodarczych warunkujących zapotrzebowanie na przewozy. Zaproponowano kilka zasad klasyfikacji typologicznej jednostek terytorialnych.

## 1. Metody czynnikowe i ich zastosowania

W metodach czynnikowych postuluje się zazwyczaj istnienie liniowych zależności między zmiennymi. Najwięcej informacji o liniowych zależnościach w zbiorze zmiennych zawiera macierz kowariancji zbioru informacji  $Y = [y_1, y_2, \dots, y_n]$ , gdzie  $y_j$  - wektor obserwacji j-tej cechy,  $y_{ij}$  - wartość j-tej cechy dla i-tej obserwacji,  $i=1, \dots, m$ ,  $j=1, \dots, n$ , n-liczba cech (zmiennych), m - liczba obserwacji.

---

\*Praca została zrealizowana w temacie PBZ-105-01 "Badanie potoków pasażerów i ładunków w komunikacji krajowej i międzynarodowej wraz z ich prognozami na lata 2005-2010, przy uwzględnieniu przejść granicznych".

Stosowane są dwa, często mylone, modele badawcze - model analizy czynnikowej oraz model składowych głównych.

### 1.1 Metoda analizy czynnikowej

W metodzie analizy czynnikowej zakłada się, że każdą zmienną obserwowalną można zapisać jako kombinację liniową podstawowych czynników wspólnych (zmiennych ukrytych) oraz dodatkowego czynnika specyficznego (swoistego) dla tej zmiennej:

$$y_j = a_{j1}f_1 + a_{j2}f_2 + \dots + a_{jk}f_k + u_j, \quad j=1, \dots, n, \text{ gdzie } f_1, f_2, \dots, f_k \text{ - czynniki wspólne, } u_1, u_2, \dots, u_n \text{ - czynniki specyficzne, } a_{ji} \text{ - ładunki czynnikowe, } j = 1, \dots, n, i = 1, \dots, k.$$

Oznacza to, że między zmiennymi obserwowalnymi  $y_j$  nie istnieje bezpośrednia zależność, zaś przyczynami zjawisk obserwowalnych są pewne ukryte, bezpośrednio nieobserwowalne, zmienne  $f_1, f_2, \dots, f_k$ , występujące w postaci jednego lub więcej czynników,  $k < n$ .

Przyjmuje się następujące założenia:

1. czynniki specyficzne  $u_1, u_2, \dots, u_n$  są nieskorelowane ze sobą,
2. czynniki wspólne  $f_1, f_2, \dots, f_k$  są nieskorelowane ze sobą,
3. czynniki wspólne i specyficzne są ze sobą wzajemnie nieskorelowane.

Ponadto zakłada się, że czynniki wspólne są standaryzowane.

Ponieważ tylko część zmienności wyjaśniana jest przez czynniki wspólne, wartości czynnikowe są jedynie wartościami estymowanymi (w porównaniu ze ścisłymi wartościami składowych głównych).

Analiza czynnikowa pozwala sprowadzić wyjściowy zbiór cech charakteryzujących poszczególne obiekty do mniejszej liczby hipotetycznych cech, zwanych czynnikami. Jej celem jest uzyskanie metacech, syntetyzujących cechy wyjściowe, redukcja zbioru informacji oraz wskazanie czynnika dającego najwięcej informacji o badanym zjawisku. Interpretacja czynników jest najważniejszą częścią procedury analizy czynnikowej i ma decydujące znaczenie dla poznania rzeczywistości. Operacja ta jest jednak bliżej niezidentyfikowana metodologicznie. Próby sformalizowania postępowania w tym zakresie nie dały, jak dotychczas, rezultatów.

Analiza czynnikowa stosowana jest głównie w badaniach psychologicznych, ze względu na charakter problemów (postuluje się występowanie zmiennych niejawnych); jej przydatność w badaniach ekonomiczno-społecznych jest mniejsza.

### 1.2 Metoda składowych głównych

Polega ona na transformacji ortogonalnej zbioru zmiennych wyjściowych  $y_j$  w zbiór, o tej samej liczebności, nowych zmiennych  $v_j$ , nazywanych składowymi głównymi, ( $j=1 \dots n$ ),  $v_j = Yw_j$ , gdzie  $w_j$  - wektor  $n$ -wymiarowy. Metoda składowych głównych stosowana jest do znajdowania kombinacji liniowych z dużą wariancją. Zdecydowana większość wariancji badanych zmiennych jest odwzorowywana wówczas przez pierwszych kilka składowych.

Zagadnienie to sprowadza się do rozwiązania układu równań  $(A - \lambda I)w = 0$  ,

gdzie: A - przyjęta macierz korelacji lub kowariancji macierzy obserwacji Y, I - macierz jednostkowa o wymiarach  $n \times n$ , w - wektor n-wymiarowy. Powyższy układ równań ma rozwiązanie niezerowe wtedy i tylko wtedy, gdy wyznacznik główny tego układu jest równy zero, tzn. spełnione jest równanie  $|A - \lambda I| = 0$ .

Pierwiastki tego równania, czyli wartości własne  $\lambda_j$  macierzy A, porządkujemy wg zasady  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . Każdej wartości własnej  $\lambda_j$  macierzy A odpowiada wektor własny  $w_j$  spełniający układ równań jednorodnych.

Wektory własne  $w_j$  mają następujące własności:

$$w_j'w_j = 1; w_j'w_i = 0; \text{ dla } i \neq j, \quad w_j'Aw_j = \lambda_j; w_j'Aw_i = 0, \text{ dla } i \neq j.$$

Każdy wektor własny  $w_j$  pochłania część zmienności (informacji) zawartej w zbiorze danych wyjściowych Y, przy czym najwięcej zawiera jej wektor  $w_1$  odpowiadający pierwszej wartości własnej  $\lambda_1$ , zaś następne odpowiednio mniej.

Aby zorientować się, jaka część informacji zawartych w zmiennych (cechach) wyjściowych została zawarta w wyznaczonych składowych, należy obliczyć wyrażenie:

$$\left[ \lambda_j / (\lambda_1 + \lambda_2 + \dots + \lambda_n) \right] * 100\%, \quad \text{dla } j=1 \dots n.$$

Dla k pierwszych składowych,  $k \leq n$ , można obliczyć udział skumulowany jako sumę k kolejnych wartości procentowych udziałów.

Składowe główne mogą być wyprowadzone zarówno z macierzy kowariancji, jak i z macierzy korelacji macierzy obserwacji Y. W obu przypadkach otrzymuje się różne rozwiązania.

Miara stopnia, w jakim w danej składowej głównej znalazła odbicie zmienność cech oryginalnych, są współczynniki korelacji tych cech z poszczególnymi składowymi.

Interpretacja wyników - tzn. określenie charakteru składowych głównych - może być przeprowadzona na podstawie analizy :

- wektorów własnych - kwadraty elementów j-tego wektora własnego stanowią miary udziału poszczególnych cech w tworzeniu zmienności j-tej składowej ;
- wartości współczynników korelacji cech z poszczególnymi składowymi oraz
- wartości kwadratów współczynników korelacji cech z poszczególnymi składowymi.

Otrzymane składowe główne mogą być interpretowane jako nowe zmienne o charakterze metacech, jak również mogą posłużyć do klasyfikacji typologicznej badanych obiektów. Zasadniczo, celem stosowania metody składowych głównych nie jest redukcja danych; raczej jest ona przeznaczona do transformacji zmiennych.

Ze względu na fakt, że pierwsze składowe główne zawierają największy procentowy udział ogólnej zmienności wyjściowego zbioru obserwacji, w dalszych badaniach zazwyczaj uwzględnia się tylko kilka pierwszych składowych. Heurystycznym kryterium wyboru liczby rozważanych składowych jest aby łączny udział wyjaśnianej przez nie zmienności wynosił  $75 \div 80$  % całkowitej informacji.

## 2. Zastosowanie metody składowych głównych do zagadnień transportowych

Metoda składowych głównych została zastosowana do zagadnień związanych z określaniem czynników warunkujących zapotrzebowanie na transport. W tym celu dokonano wielocechowej analizy zmiennych charakteryzujących rozwój społeczno-gospodarczy Polski w ujęciu regionalnym, a następnie wykorzystano otrzymane wyniki do przeprowadzenia klasyfikacji typologicznej regionów. Do obliczeń przyjęto zestaw 30 cech odzwierciedlających poziom rozwoju społeczno-gospodarczego i infrastruktury kraju w układzie jednostek terytorialnych, zgodnym z podziałem administracyjnym na województwa. Cechy zostały podzielone na grupy wyznaczające różnorodne charakterystyki regionów, takie jak: wielkość województwa, aktywność zawodowa ludności, poziom rozwoju rolnictwa oraz sposób użytkowania ziemi, poziom handlu i usług, poziom rozwoju gospodarki terenowej, stopień uprzemysłowienia.

Przeprowadzono obliczenia dla pięciu wariantów danych, rozszerzając stopniowo liczbę analizowanych cech od 6 do 30. Do zbioru cech wprowadzono zmienne mogące pośrednio bądź bezpośrednio wywierać wpływ na wielkość zapotrzebowania na przewozy (np. przyjęto zmienne związane z turystyką oraz liczbą szkół ponadpodstawowych jako cechy mogące wpływać na zmianę zapotrzebowania na przewozy pasażerskie) oraz uwzględniono infrastrukturę sieci komunikacji kolejowej i drogowej. Dane dotyczyły roku 1992 i pochodziły z roczników statystycznych GUS.

## 3. Wyniki obliczeń

Jak już wspomniano, obliczenia prowadzono ewolucyjnie rozszerzając stopniowo uwzględnianą liczbę cech. Pozwalało to ustalić, które z cech są istotne dla opisu rozważanego zjawiska, a które można pominąć bez straty informacji. Pierwszy wariant obliczeń obejmował pierwszych 6, zaś ostatni wszystkie 30 przedstawionych w wykazie cech. Składowe główne zostały wyprowadzone z macierzy korelacji macierzy obserwacji Y. Przedstawione wyniki dotyczą wariantu dla 30 cech. Udział zmienności wyjaśnianej przez poszczególne składowe główne był następujący:

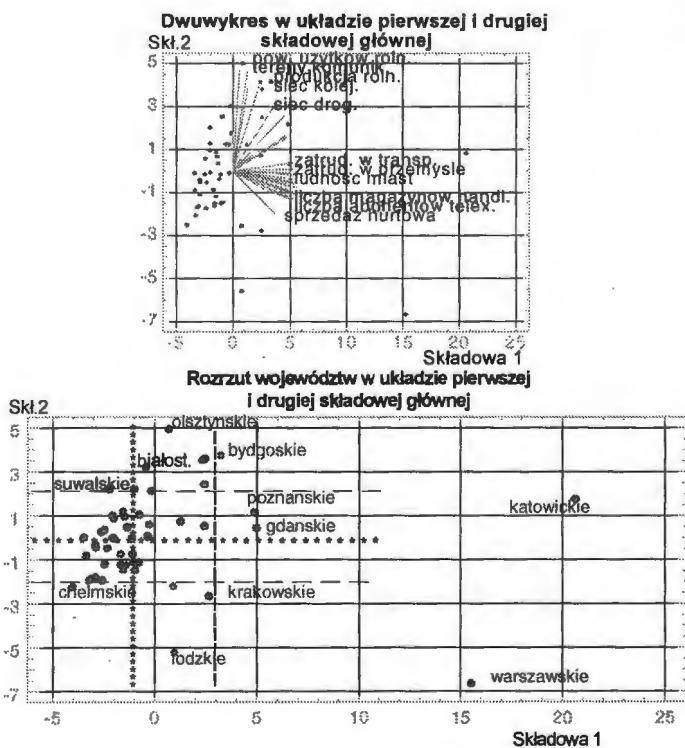
Nr skl.	% wariancji	% skumulowany	Nr skl.	% wariancji	% skumulowany
1	63.04623	63.04623	4	3.67166	90.18422
2	16.89068	79.93691	5	3.07999	93.26421
3	6.57565	86.51256	6	2.33603	95.60023

Charakter składowych głównych określano na podstawie analizy współczynników korelacji cech z poszczególnymi składowymi. Przyjęto, że charakter składowej kształtowany jest przez cechy silnie (dla pierwszych kilku składowych było to na poziomie 0.9) skorelowane z rozważaną składową. Analiza otrzymanych wyników wykazała, że charakter pierwszych dwóch składowych głównych był zachowany we wszystkich wariantach obliczeń. Pierwsza składowa główna kształtowana była przez cechy opisujące poziom rozwoju przemysłu, infrastruktury oraz potencjału demograficznego województw. Drugą składową główną kształtowały przede wszystkim cechy związane z poziomem rozwoju rolnictwa oraz ze sposobem użytkowania ziemi.

W dalszych rozważaniach uwzględniono dwie pierwsze składowe główne z uwagi na przejrzystość interpretacji oraz, jak to widać w zestawieniach dotyczących udziału zmienności wyjaśnianej przez poszczególne składowe główne, z uwagi na fakt, że błąd wynikający z pominięcia pozostałych składowych jest stosunkowo mały (co najwyżej ok. 20%).

Otrzymane wyniki zilustrowano na wykresie (po lewej). Przedstawiono rozrzut województw w układzie pierwszej i drugiej składowej głównej oraz, na tym tle, sposób w jaki poszczególne cechy tworzą omawiane składowe.

Długość wektorów odpowiadających poszczególnym cechom jest proporcjonalna do wartości współczynników korelacji cech z daną składową główną, zaś kąt między wektorami jest odwrotnie proporcjonalny do współczynnika korelacji między cechami.



Na kolejnym wykresie (po prawej) przedstawiono rozrzut województw w układzie I i II składowej głównej dla wariantu 30 cech. Może on służyć jako ilustracja sposobu dokonywania typologii województw według pewnych, arbitralnie przyjmowanych, kryteriów.

Jedną z możliwości jest przyjęcie jako kryterium wartości dwóch pierwszych składowych głównych  $v_1$  i  $v_2$  oraz klasyfikacji typologicznej województw na podstawie podziału wartości składowych wg mediany. W efekcie otrzymuje się 4 klasy

typologiczne, charakteryzowane przez poziom rozwoju przemysłu i infrastruktury (wysoki  $v_1 \geq me_1$  lub niski  $v_1 < me_1$ ) oraz rolnictwa (analogicznie wysoki,  $v_2 \geq me_2$  lub niski  $v_2 < me_2$ ). Wartości mediany były:  $me_1 = -1,217$ ,  $me_2 = -0,096$ .

Podział wg mediany oznaczono linią "\*\*\*\*", zaś wg drugiego, omówionego dalej kryterium, linią przerywaną.

Należy zauważyć, że przedstawiona klasyfikacja ma jedynie charakter pogładowy. Wynika to z faktu, że podział połówkowy (wg mediany) jest zgrubny i nie uwzględnia bardziej szczegółowych kryteriów. Ponadto, przedstawione wyniki otrzymano dla danych z roku 1992, czyli z okresu trwającej transformacji gospodarki, podczas której uległy zaburzeniom dotychczasowe trendy rozwoju województw.

Można również zaproponować inne zasady porządkowania województw, odpowiadające celom, jakim ma służyć klasyfikacja. Jeżeli za podstawowe kryterium przyjąć wartości pierwszej składowej głównej  $v_1$ , wówczas można zauważyć, że pięć silnie rozwiniętych województw - warszawskie, katowickie, gdańskie, poznańskie i bydgoskie - wyraźnie dominuje nad pozostałymi. Jeżeli powyższy podział zostanie następnie uzupełniony kryterium - podziału na klasy według wartości drugiej składowej głównej -  $v_2 \leq -2$ ,  $v_2 \geq 2$ ,  $-2 < v_2 < 2$  - otrzymamy podział pozostałych województw na klasy o wysokim, niskim i średnim poziomie rozwoju rolnictwa oraz średnim poziomie rozwoju przemysłu. W tej ostatniej klasie zawiera się najwięcej województw.

#### 4. Podsumowanie

Klasyfikacja typologiczna województw będąca wynikiem zastosowania metody składowych głównych może być przydatna w wielu dziedzinach, m.in. w badaniach transportowych. Wynika to z faktu, że silnie rozwinięte aglomeracje są ośrodkami generującymi większe zapotrzebowanie na przewozy; zaś charakter składowej dominującej w danym regionie może dodatkowo wskazywać na rodzaj przewozów.

Przedstawione badanie ma charakter wstępny. Zdaniem autorek jego wyniki są zachęcające. Metoda składowych głównych wydaje się być dobrym narzędziem analizy zjawisk społeczno-gospodarczych, w tym również zagadnień transportowych i regionalizacyjnych. Jakość uzyskanych wyników zależy w istotny sposób od jakości i dostępności danych. Przewiduje się rozszerzenie badań i wykorzystanie danych ściśle transportowych, takich jak np. kierunki i wielkość przepływu towarów między jednostkami terytorialnymi oraz na przejściach granicznych.

#### 5. Literatura

- Chojnicki Z., Czyż T. - Analiza zmienności różnicowania przestrzeni społeczno-ekonomicznej Polski, PAN, seria GEOGRAFIA, T.II, PWN Warszawa - Poznań 1978
- Jajuga K. - Statystyczna analiza wielowymiarowa, PWN, Warszawa 1993
- Morrison D.F. - Multivariate statistical methods, McGraw Hill, New York 1975
- Ratajczak W. - Analiza i modele wpływu czynników społeczno-gospodarczych na kształtowanie się sieci transportowej, PWN Warszawa 1980
- Rocznik Statystyczny województw 1993, GUS, Warszawa 1993

**ISBN 83-85847-85-5**

---

**W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy  
prosimy o kontakt  
z Instytutem Badań Systemowych PAN  
ul. Newelska 6, 01-447 Warszawa  
tel. 36-19-01 w. 241 e-mail: kotuszew@ibspan.waw.pl**