

Kazimierz Zając

Akademia Ekonomiczna w Krakowie
Wyższa Szkoła Zarządzania i Nauk Społecznych w Tychach

Daniel Kosiorowski

Akademia Ekonomiczna w Krakowie

PRZYCZYNEK DO DYSKUSJI NA TEMAT PROWADZENIA BADAŃ NA PODSTAWIE OSOBLIWEJ MACIERZY KOWARIANCJI

1. Wstęp

Możemy mieć nadzieję, że kiedyś dzięki metodom matematycznym decyzje nieco bardziej będą opierały się o wiedzę, a nieco mniej o zgadywanie i że świat, w którym żyjemy, będzie nieco lepiej funkcjonował, będąc w mniejszym stopniu zdany na łaskę nieprzewidzianych wydarzeń

(R. Stone)

Matematyzacja nauki – pisze C.V. Newson – doprowadziła, zwłaszcza w ostatnich latach, do rewolucyjnych zmian w naszej cywilizacji. Nadto wnioski, uzyskane przez uczonych, zmierzających do pogłębienia rozumienia istoty matematyki, ukazały w nowym świetle wiele aspektów ludzkiej wiedzy.

Galileusz pierwszy zrozumiał doświadczenia, które prowadzą do ilościowych wyników. U podstaw każdego doświadczenia leży pomiar. Działalność nie dająca się zmierzyć podlega zmianom według kaprysu lub mody – stwierdza Russel. Reguły pomiaru są „pomostem” między czysto formalnymi rachunkami a empiryczną rzeczywistością. Te trzy elementy: formalny rachunek matematyczny, wielkości obserwowalne, baza empiryczna oraz reguły łączące matematykę z empirią, wnioskowanie statystyczne jest istotnym składnikiem naukowej teorii. Teoria zaś tylko wtedy zasługuje na miano naukowej, gdy wynikają z niej jakieś wnioski nadające się do eksperymentalnego sprawdzenia.

Należy podkreślić, że znajomość teorii jest niezbędna w badaniach naukowych. Jest ona potrzebna nie tylko po to, aby określić, jakie hipotezy badawcze należy postawić i jakimi metodami je weryfikować. Dlatego też teoria musi istnieć zawsze *a priori* w stosunku do empirycznej obserwacji faktów. Nabierają one znaczenia wówczas, gdy zostaną ustalone i usystematyzowane w ramach teorii. Inaczej mówiąc, fakty w gruncie rzeczy mają sens tylko w konfrontacji z teorią.

U podstaw każdej usystematyzowanej próby poszukiwania prawdy o społeczeństwie zatem leży zawsze teoria, gdyż wytycza kierunek tych badań. Można więc ją określić jako hipotetyczny zarys opisu faktów oraz istniejących między nimi powiązań przyczynowych.

Nauki statystyczne, do których duży wkład wniósł znakomity uczony prof. dr hab. Zdzisław Hellwig, są tym szczególnym aspektem postępu ludzkości, który nadał XXI wiekowi jego specjalny charakter – to do statystyka zwraca się obecny wiek w poszukiwaniu tego, co jest najistotniejsze we wszystkich przedsięwzięciach (R. Fisher).

Popperowska krytyka radykalnego empiryzmu w kwestii stosunków między faktami i teoriami wydaje się trafna.

H. Poincare [1908, s. 119-120] pisał: „Mówi się często, że trzeba eksperymentować bez myśli z góry powziętej. Jest to niemożliwe, nie tylko bowiem sprawiłoby, że wszelkie doświadczenie stałoby się jałowym, ale nawet przy najlepszej chęci byłoby niemożliwe. Każdy nosi swoje pojmowanie świata, od którego niełatwo się wyzwolić. Trzeba na przykład posługiwać się językiem, a język nasz cały ulepiony jest z myśli z góry powziętych i inaczej być nie może...”.

Nadanie poprawnej formy statystycznej i matematycznej pewnemu zjawisku pozwala zorientować się, w jakich warunkach i jakie czynniki wpływają na jego przebieg, jak wpływają (w jakim kierunku i z jakim nasileniem). Dlatego też autorzy, wychodząc z hipotezy, że rozwój demograficzny warunkuje rozwój społeczno-ekonomiczny, starali się ją zweryfikować na podstawie przeprowadzonej analizy opartej na danych empirycznych z wykorzystaniem metod ilościowych. Pozwalają one weryfikować pewne hipotezy ogólne, lepiej charakteryzować opisywaną rzeczywistość. Wprowadzenie metod ilościowych jest zatem, z metodologicznego punktu widzenia, usprawnieniem bardzo doniosłym, z wielokrotnia bowiem siłą poznawczą danej dyscypliny naukowej i chroni wnioski przed subiektywizmem badacza.

Wydaje się, że na wstępie warto podać kilka spostrzeżeń na temat związków procesów demograficznych i społeczno-ekonomicznych. Zdania na ten temat są podzielone. Wydaje się, że problem jest zbyt złożony, aby można było w sposób jednoznaczny stwierdzić silne związki między rozwojem demograficznych a ekonomicznym danego kraju. Różne są teorie i doktryny dotyczące związków między liczbą ludności a jej dobrobytem.

Oceniając wzrost gospodarczy danego kraju, badania koncentruje się zwykle na dwóch podstawowych problemach, a mianowicie rozważany jest rozwój ekonomiczno-społeczny oraz kształtowanie się zjawisk demograficznych.

W pracy, przyjmując na początku sformułowaną hipotezę, zmierzamy do tego, aby przy odpowiednio skonstruowanych metodach na podstawie badań empirycznych uzyskać podstawy przyjęcia lub odrzucenia podanej hipotezy.

2. Wprowadzenie do części metodycznej

Macierz kowariancji wielowymiarowej zmiennej losowej, będącej modelem interesującej nas populacji, bardzo często stanowi podstawę rozumowań dotyczących własności populacji. Macierz ta pojawia się w zagadnieniach regresji, klasyfikacji, dyskryminacji, skalowania wielowymiarowego, konstruowania portfela inwestycyjnego.

Właściwe oszacowanie macierzy kowariancji na podstawie małej próby należy do zadań niełatwych, i szczególnie istotnych, gdyż przekłada się bezpośrednio na wnioski, jakie formułuje się w badaniach. Zagadnienie wiąże się zarówno z typem rozkładu w populacji (prace prof. Zdzisława Hellwiga dotyczące wielowymiarowych testów normalności), jak i z wpływem obserwacji odstających (prace prof. Krzysztofa Jajugi dotyczące przycinania wielowymiarowych wektorów losowych, redukcji wymiaru problemu statystycznego).

W niniejszej pracy zwracamy uwagę na zagadnienia związane osobliwą macierzą kowariancji z próby. Macierzy rozpinającej podprzestrzeń o mniejszym wymiarze aniżeli macierz kowariancji w populacji, nie posiadającej odwrotności ważnej np. przy klasyfikacji obiektów z wykorzystaniem odległości Mahalanobisa, teście T^2 Hotellinga. Zagadnienie to nie jest w pełni rozwiązane, było podejmowane w odniesieniu do uogólnionych modeli liniowych m.in. przez J. Neymana, F.N. Davida, E. Parzena, C.R. Rao. Zagadnienie pojawia się w naturalny sposób, gdy liczba obserwacji przekracza liczbę rozpatrywanych w badaniach zmiennych. Z racji swej doniosłości praktycznej (koszt badań) stanowi obecnie przedmiot intensywnych studiów na uniwersytetach amerykańskich, jak zaznaczył prof. Vyacheslav Girko na konferencji WAS'2004.

Jednym z kierunków badań są studia nad wykorzystaniem uogólnionych odwrotności macierzy kowariancji z próby \mathbf{S}^- , tzn. takich macierzy, które czynią zadość warunkowi $\mathbf{S}^- \mathbf{S} \mathbf{S}^- = \mathbf{S}^-$. Przykładowo wykorzystuje się uogólnione odwrotności Moore'a-Penrose'a macierzy kowariancji z próby:

$$\mathbf{S}^- = \sum_{j=1}^p \lambda_j^{-1} \mathbf{P}_j \mathbf{P}_j^T, \quad (1)$$

gdzie \mathbf{P}_j są wektorami własnymi \mathbf{S} odpowiadającymi p – niezerowym wartościom własnym λ_j , $j = 1, \dots, p$.

Dla punktów \mathbf{x} i \mathbf{m} populacji rozpatrywanej ze względu na wielowymiarową cechę dystans Mahalanobisa przyjmuje postać:

$$(\mathbf{x} - \mathbf{m})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}), \quad (2)$$

Innym kierunkiem badań zaznaczanym m.in. przez Eddy'ego na łamach „The Annals of Statistics” jest takie przeformułowanie przedmiotu badań, aby możliwa była zamiana miejsc zmiennych i obserwacji, tzn. transpozycja macierzy obserwacji. W niniejszej pracy zastosowano zabieg polegający na transpozycji macierzy obserwacji 16 województw rozpatrywanych ze względu na 34 standaryzowane zmienne społeczno-ekonomiczne. Uzyskano tym samym macierz obserwacji dla 16 zmiennych reprezentujących relatywne pozycje województw, macierz zawierającą 34 obserwacje aspektów relatywnej pozycji województw RP.

3. Rezultaty badań

Badania oparto na danych dotyczących województw Polski w roku 2002. W wyniku przeprowadzonej analizy na 34 zmiennych demograficznych i społeczno-ekonomicznych, stosując analizę głównych składowych wyodrębniono 2 zmienne demograficzne i 3 zmienne społeczno-gospodarcze. Wyniki przedstawiono w tab. 1 i 2.

Tabela 1. Procent wyjaśnionej wariancji przez kolejne składowe

| Zmienne demograficzne | | | | |
|-----------------------|----------------|---------|---------|---------|
| | Wartość własna | % ogółu | Skumul. | Skumul. |
| 1 | 3,67 | 30,15 | 3,67 | 30,15 |
| 2 | 3,31 | 27,12 | 6,98 | 57,27 |
| 3 | 2,04 | 16,75 | 9,02 | 74,03 |
| 4 | 1,27 | 10,44 | 10,29 | 84,47 |
| 5 | 0,60 | 4,90 | 10,89 | 89,37 |
| 6 | 0,54 | 4,47 | 11,44 | 93,83 |
| 7 | 0,29 | 2,39 | 11,73 | 96,23 |
| 8 | 0,26 | 2,10 | 11,98 | 98,33 |
| 9 | 0,10 | 0,82 | 12,08 | 99,15 |
| 10 | 0,06 | 0,47 | 12,14 | 99,62 |
| 11 | 0,04 | 0,37 | 12,19 | 99,98 |
| 12 | 0,00 | 0,02 | 12,19 | 100,00 |

Źródło: obliczenia własne na podstawie danych GUS.

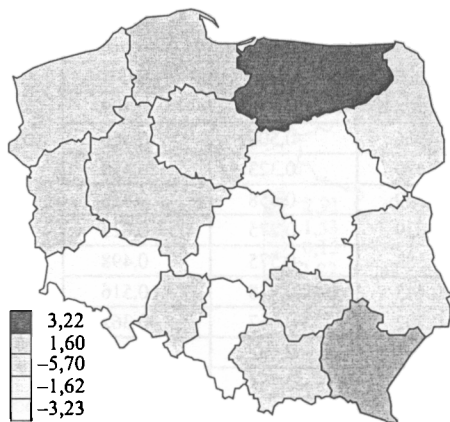
Tabela 1 wykazuje, że 57,27% zmienności 12 cech demograficznych wyjaśniają dwie wyodrębnione składowe demograficzne. Natomiast tab. 2 przedstawia współrzędne wyodrębnionych składowych. W wyniku otrzymanych wartości łatwo zauważyć, że 12 cech demograficznych w odmienny sposób wpływa na główne składowe. Na przykład zgony na 1000 ludności, saldo migracji, liczba lekarzy ujemnie wpływają na składową pierwszą, a dodatnio – na składową drugą.

Tabela 2. Współrzędne składowych demograficznych

| Zmienne demograficzne | Czynn. D1 | Czynn. D2 |
|--|-----------|-----------|
| Urodzenia żywe na 1000 ludności | 0,761 | 0,513 |
| Zgony na 1000 ludności | -0,729 | 0,217 |
| Przyrost naturalny na 1000 ludności | 0,871 | 0,125 |
| Małżeństwa na 1000 ludności | 0,356 | 0,797 |
| Saldo migracji | -0,034 | 0,806 |
| Zgony niemowląt na 1000 urodzeń żywych | -0,390 | -0,127 |
| Ludność miejska (w %) | -0,420 | -0,160 |
| Stopa bezrobocia | 0,422 | -0,525 |
| Przeciętna liczba osób na 1 izbę | 0,557 | 0,473 |
| Liczba mieszkań oddanych do użytku na 100 ludności | -0,096 | 0,583 |
| Utrzymywane szkoły podstawowe | -0,399 | 0,673 |
| Liczba lekarzy | -0,552 | 0,572 |
| Liczba pielęgniarok | -0,624 | 0,185 |

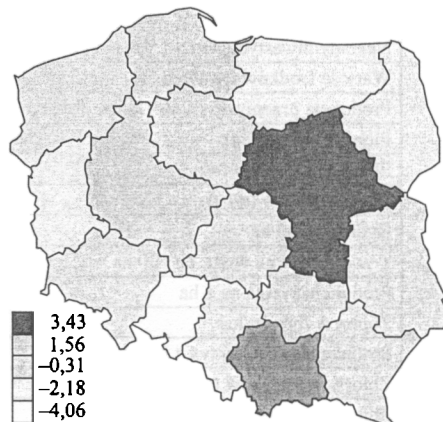
Źródło: obliczenia własne na podstawie danych GUS.

Bardziej szczegółowe informacje otrzymano, przedstawiając wartości składowych w poszczególnych województwach, co przedstawiono na rys. 1 i na rys. 2. Pierwsza składowa przyjmuje wartość największą w województwie warmińsko-mazurskim, a najmniejszą – w województwie łódzkim. Natomiast druga składowa wartość największą przyjmuje w województwie mazowieckim, a najmniejszą – w województwie polskim.



Rys. 1. Wartości pierwszej składowej demograficznej

Źródło: obliczenia własne, dane GUS.



Rys. 2. Wartości drugiej składowej demograficznej

Źródło: obliczenia własne, dane GUS.

Przechodząc z kolei do omówienia głównych składowych ekonomiczno-gospodarczych zawartych w tab. 3 i 4, stwierdzono, że 3 składowe ekonomiczne wyjaśniają 73,25% wariacji 19 zmiennych ekonomicznych (tab. 3). Współrzędne trzech wyodrębnionych składowych przedstawia tab. 4.

Tabela 3. Procent wyjaśnionej wariacji przez kolejne składowe

| Zmienne gospodarcze | | | | |
|---------------------|-----------|---------|---------|---------|
| | W. własna | % ogółu | Skumul. | Skumul. |
| 1 | 6,15 | 34,50 | 6,15 | 34,50 |
| 2 | 3,65 | 20,50 | 9,80 | 55,00 |
| 3 | 3,25 | 18,25 | 13,05 | 73,25 |
| 4 | 2,05 | 11,52 | 15,10 | 84,78 |
| 5 | 0,84 | 4,70 | 15,94 | 89,47 |
| 6 | 0,70 | 3,95 | 16,64 | 93,43 |
| 7 | 0,44 | 2,49 | 17,09 | 95,92 |
| 8 | 0,23 | 1,32 | 17,32 | 97,24 |
| 9 | 0,18 | 1,03 | 17,50 | 98,27 |
| 10 | 0,14 | 0,77 | 17,64 | 99,04 |
| 11 | 0,08 | 0,45 | 17,72 | 99,49 |
| 12 | 0,06 | 0,34 | 17,78 | 99,83 |
| 13 | 0,02 | 0,09 | 17,80 | 99,92 |
| 14 | 0,01 | 0,05 | 17,81 | 99,97 |
| 15 | 0,00 | 0,03 | 17,81 | 100,00 |

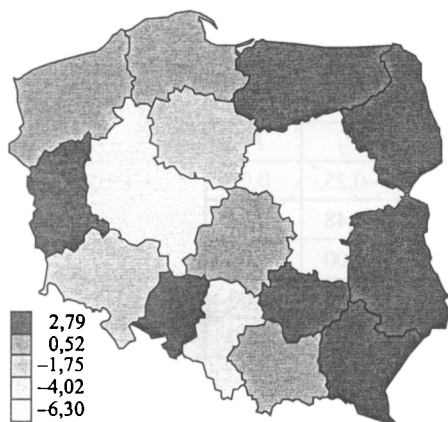
Źródło: obliczenia własne, dane GUS.

Tabela 4. Współrzędne składowych ekonomicznych

| Zmienne gospodarcze | Czynn. P ₁ | Czynn. P ₂ | Czynn. P ₃ |
|--|-----------------------|-----------------------|-----------------------|
| Zatrudnienie w przemyśle na 1000 ludności | -0,753 | 0,089 | -0,191 |
| Zatrudnienie w budownictwie na 1000 ludności | -0,762 | 0,336 | -0,269 |
| Nakłady inwestycyjne na 1 mieszkańca | -0,823 | 0,208 | -0,015 |
| Wartość środków trwałych na 1 mieszkańca | -0,768 | 0,183 | -0,091 |
| Produkcja przemysłowa na 1 mieszkańca | -0,921 | 0,168 | -0,109 |
| Plony 4 zbóż (q/ha) | -0,566 | -0,506 | 0,466 |
| Plony ziemniaków (q/ha) | -0,490 | 0,323 | 0,559 |
| Plony buraków (q/ha) | -0,369 | -0,658 | 0,416 |
| Bydło na 100 ha | -0,029 | 0,275 | 0,823 |
| Trzoda chlewna w szt. na 100 ha | -0,388 | -0,575 | 0,498 |
| Produkcja żywca na 1 ha | -0,443 | -0,418 | 0,516 |
| Zbiory 4 zbóż (q/ha) | -0,302 | -0,667 | -0,362 |
| Produkcja mleka (w tys. t na 1 ha) | 0,007 | 0,432 | 0,798 |
| Zbiory ziemniaków (w tys. t) | -0,184 | -0,782 | -0,160 |
| Zbiory buraków (w tys. t) | -0,228 | -0,717 | -0,003 |
| Usługi rynkowe (%) | -0,625 | -0,120 | -0,575 |
| Sprzedaż detaliczna towarów na 1 mieszkańca | -0,766 | 0,324 | 0,147 |
| Emisja gazów (%) | -0,606 | 0,169 | -0,270 |
| Emisja pyłów (w tys. t) | -0,565 | 0,384 | -0,087 |

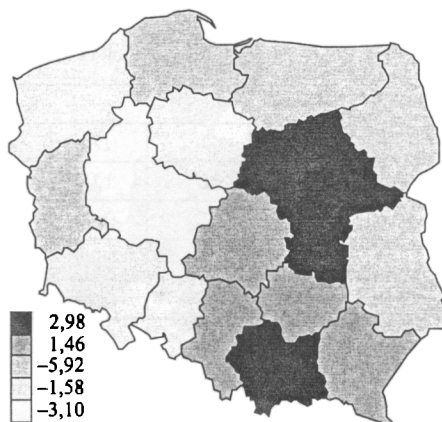
Źródło: obliczenia własne, dane GUS.

Porównując wartości współrzędnych w tab. 4, dochodzimy do wniosku, że składowe wyrażają wpływ różnie działających czynników na rozwój ekonomiczny. Składowe ekonomiczne przedstawiono na rys. 3, 4, 5.



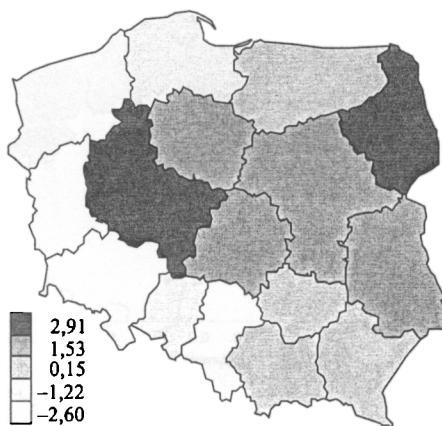
Rys. 3. Wartości pierwszej składowej ekonomicznej

Źródło: obliczenia własne, dane GUS.



Rys. 4. Wartości drugiej składowej ekonomicznej

Źródło: obliczenia własne, dane GUS.



Rys. 5. Wartości trzeciej składowej ekonomicznej

Źródło: obliczenia własne, dane GUS.

Jeśli chodzi o składową pierwszą, wartość maksymalną reprezentuje województwo podkarpackie, najmniejszą – województwo mazowieckie. Odnośnie do składowej drugiej wartość maksymalna przypada na województwo mazowieckie i małopolskie, minimalną zaś zawiera województwo wielkopolskie. Jeśli chodzi

o składową trzecią, wartość maksymalną wykazuje województwo wielkopolskie, a minimalną – dolnośląskie.

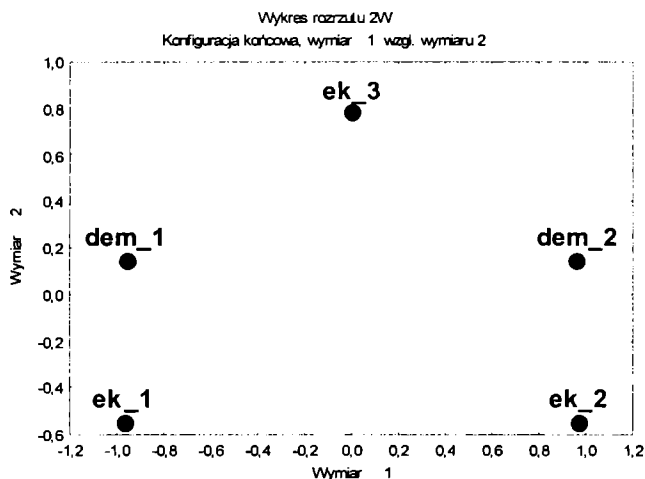
Chcąc wyrazić zależności pomiędzy składowymi demograficznymi i ekonomicznymi, obliczono macierz korelacji zawartą w tab. 5.

Tabela 5. Macierz korelacji dla wyodrębnionych składowych

| Korelacje $N = 16$ | | | | | |
|--------------------|-------|-------|-------|-------|-------|
| | D_1 | D_2 | P_1 | P_2 | P_3 |
| D_1 | 1,00 | -0,00 | 0,44 | -0,25 | 0,02 |
| D_2 | -0,00 | 1,00 | -0,34 | 0,48 | 0,38 |
| P_1 | 0,44 | -0,34 | 1,00 | -0,00 | 0,00 |
| P_2 | -0,25 | 0,48 | -0,00 | 1,00 | 0,00 |
| P_3 | 0,02 | 0,38 | 0,00 | 0,00 | 1,00 |

Źródło: obliczenia własne, dane GUS.

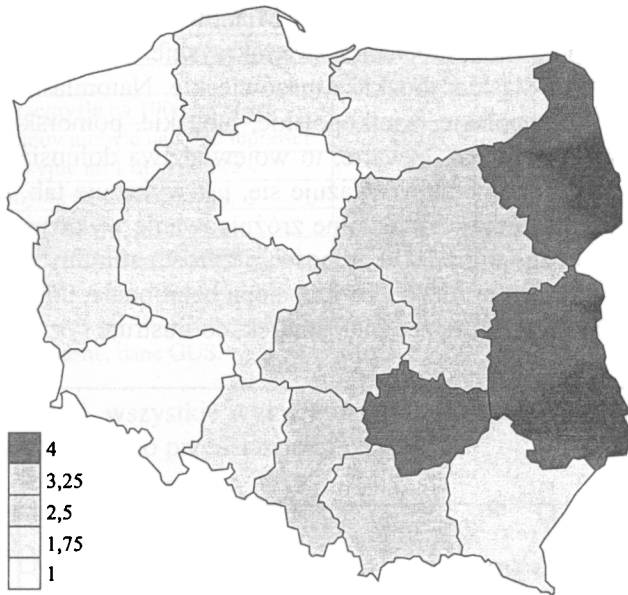
Bardziej wyrazisty obraz prezentowanej relacji przedstawia rys. 6.



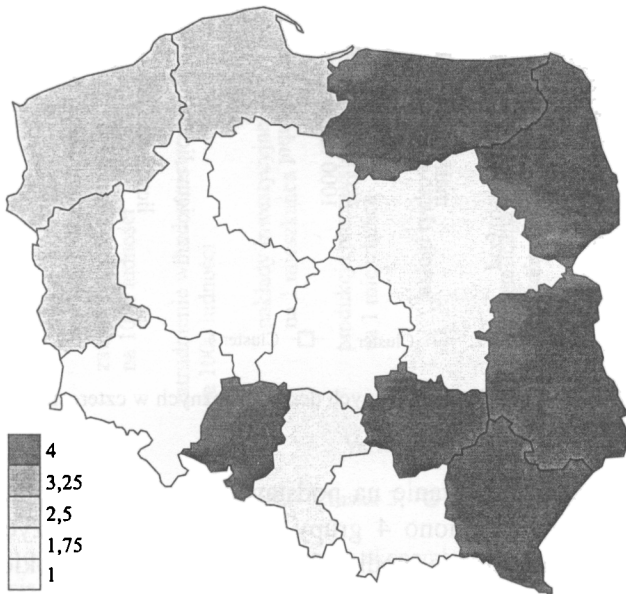
Rys. 6. Rezultaty skalowania dla składowych na podstawie macierzy korelacji
Źródło: obliczenia własne, dane GUS.

Na rysunku 6 wyraźnie widać, że pierwsza składowa demograficzna jest dodatnio skorelowana z pierwszą składową gospodarczą, druga zaś jest skorelowana z drugą składową ekonomiczną. Można stwierdzić, że trzecia składowa gospodarcza jest niezależna od pozostałych składowych.

Dalszą analizę oparto na metodzie EM (maksymalizacja wartości oczekiwanej), w wyniku której otrzymano rys. 7 i 8.

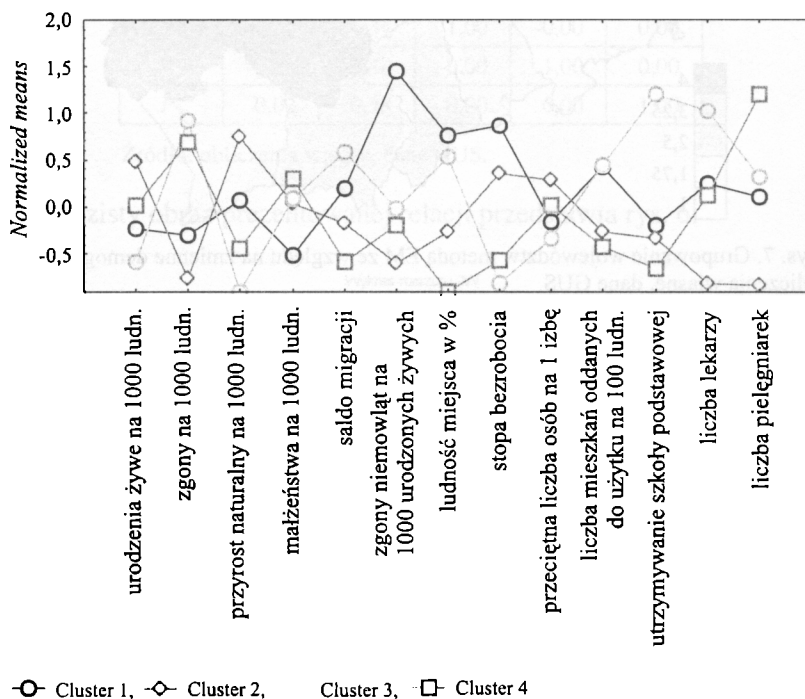


Rys. 7. Grupowanie województw metodą EM ze względu na zmienne demograficzne
 Źródło: obliczenia własne, dane GUS.



Rys. 8. Grupowanie województw metodą EM ze względu na zmienne przemysłowe
 Źródło: obliczenia własne, dane GUS.

Interesujące wyniki przedstawia rys. 7. Mianowicie pierwsze skupisko stanowią województwa: podlaskie, lubelskie i świętokrzyskie. Drugie skupisko to województwa małopolskie, śląskie, łódzkie i mazowieckie. Natomiast trzecie skupisko tworzą województwa: opolskie, wielkopolskie, lubuskie, pomorskie i warmińsko-mazurskie. Ostatnie skupisko, czwarte, to województwa dolnośląskie, kujawsko-pomorskie i zachodniopomorskie. Okazuje się, jak wykazuje tab. 6, że wyodrębnione 4 skupiska województw szczególne zróżnicowanie wykazują ze względu na następujące zmienne: zgony na 1000 ludności, przyrost naturalny na 1000 ludności, zgony niemowląt na 1000 urodzeń żywych, stopa bezrobocia, utrzymywane szkoły podstawowe, liczba lekarzy i liczba pielęgniarek, co ilustruje rys. 9.



Rys. 9. Wartości przeciętne zmiennych demograficznych w czterech skupiskach
Źródło: obliczenia własne, dane GUS.

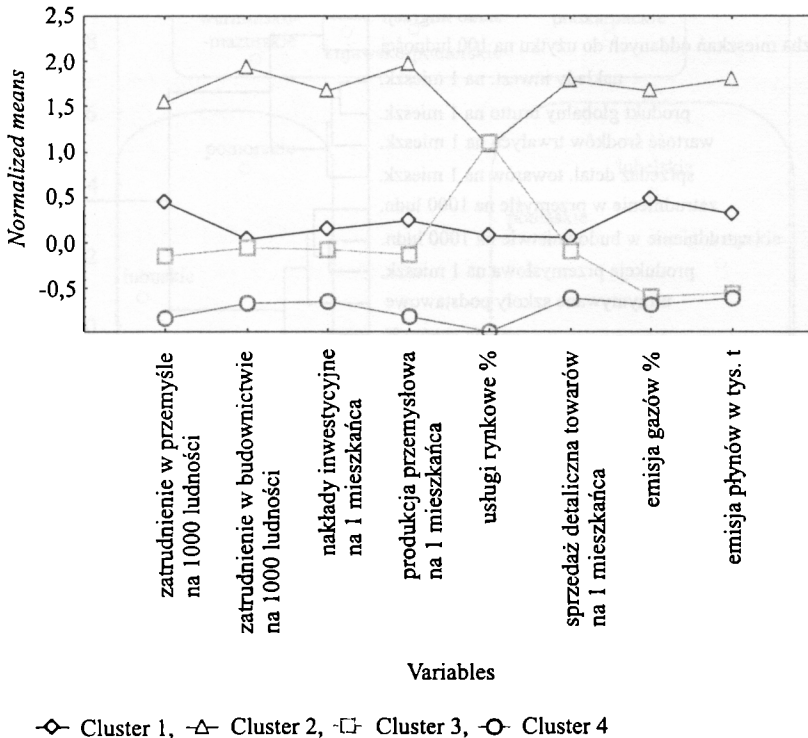
Biorąc pod uwagę grupowanie na podstawie metody EM zmiennych ekonomicznych (rys. 8), wyodrębniono 4 grupy. Do pierwszej zaliczyć można województwa: warmińsko-mazurskie, podlaskie, lubelskie, podkarpackie, świętokrzyskie i opolskie. Drugą grupę stanowią województwa: lubuskie, zachodniopomorskie i pomorskie. Trzecia grupa to województwa śląskie i mazowieckie. Ostatnią grupę stanowią pozostałe województwa.

Tabela 6. Testy ANOVA – dobroć dopasowania czterech skupisk

| ANOVA for continuous variables – Number of clusters: 4 | | | | | | |
|--|------------|----|-----------|----|-------|---------|
| | Between SS | df | Within SS | df | F | p value |
| Zatrudnieni w przemyśle na 1000 ludności | 9,87 | 3 | 5,13 | 12 | 7,70 | 0,00 |
| Zatrudnieni w budownictwie na 1000 ludności | 10,09 | 3 | 4,91 | 12 | 8,23 | 0,00 |
| Nakłady inwestycyjne na 1 mieszkańca | 8,21 | 3 | 6,79 | 12 | 4,84 | 0,02 |
| Produkcja przemysłu na 1 mieszkańca | 11,93 | 3 | 3,07 | 12 | 15,53 | 0,00 |
| Usługi rynkowe (w %) | 11,54 | 3 | 3,46 | 12 | 13,34 | 0,00 |
| Sprzedaż detal. towarów na 1 mieszkańca | 8,60 | 3 | 6,40 | 12 | 5,37 | 0,01 |
| Emisja gazów (w %) | 10,56 | 3 | 4,44 | 12 | 9,50 | 0,00 |
| Emisja pyłów (w tys. t) | 10,10 | 3 | 4,90 | 12 | 8,25 | 0,00 |

Źródło: obliczenia własne, dane GUS.

Jak wynika z tab. 7, wszystkie wymienione tam zmienne, różnicują wyodrębnione skupiska województw, co przedstawiono na rys. 10.

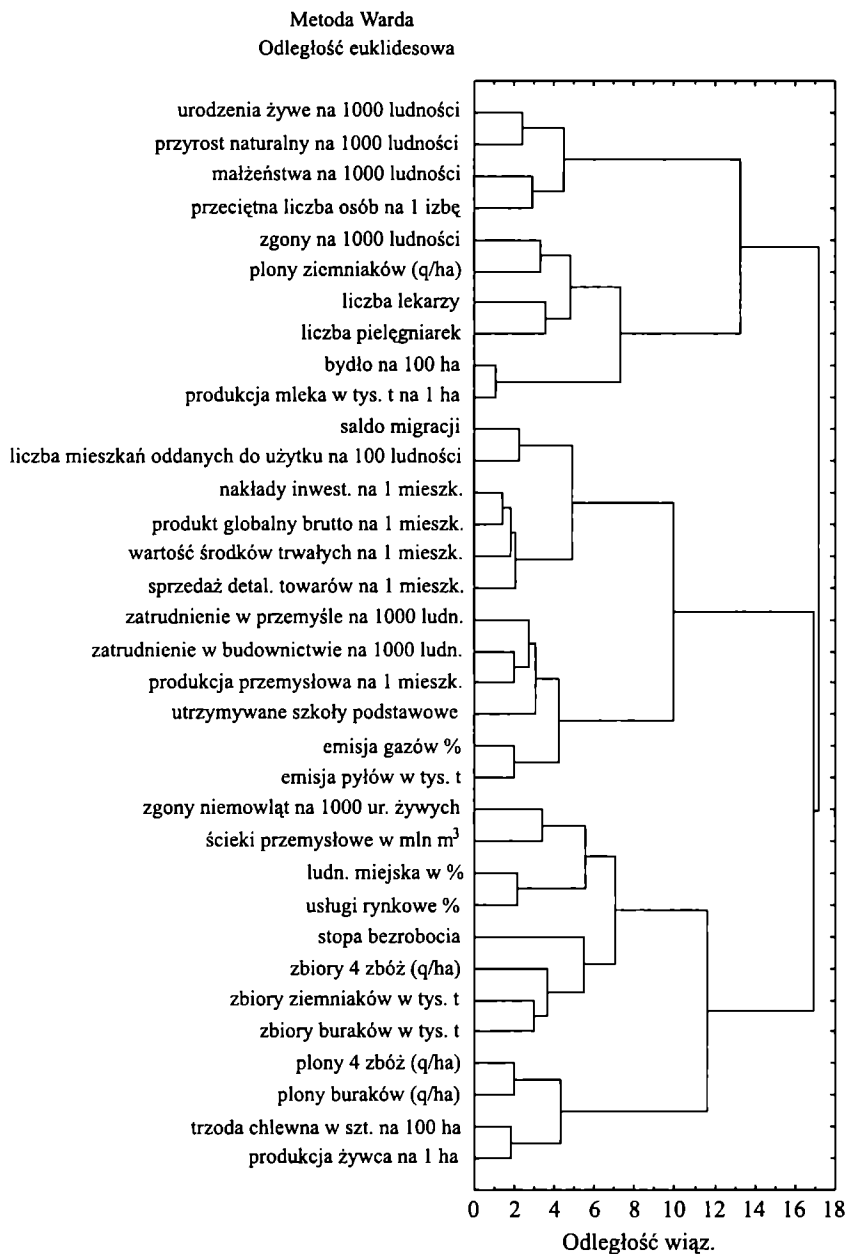


Rys. 10. Wartości przeciętne zmiennych ekonomicznych w czterech skupiskach

Źródło: obliczenia własne, dane GUS.

Pewne *novum* w przeprowadzonej analizie stanowi pogrupowanie 34 zmiennych łącznie wziętych demograficznych i ekonomicznych potraktowanych jako przypadki na 16 województw traktowanych jako zmienne.

Po przeprowadzeniu stosownej standaryzacji uzyskano ilustrację zależności pomiędzy relatywnymi pozycjami województw.

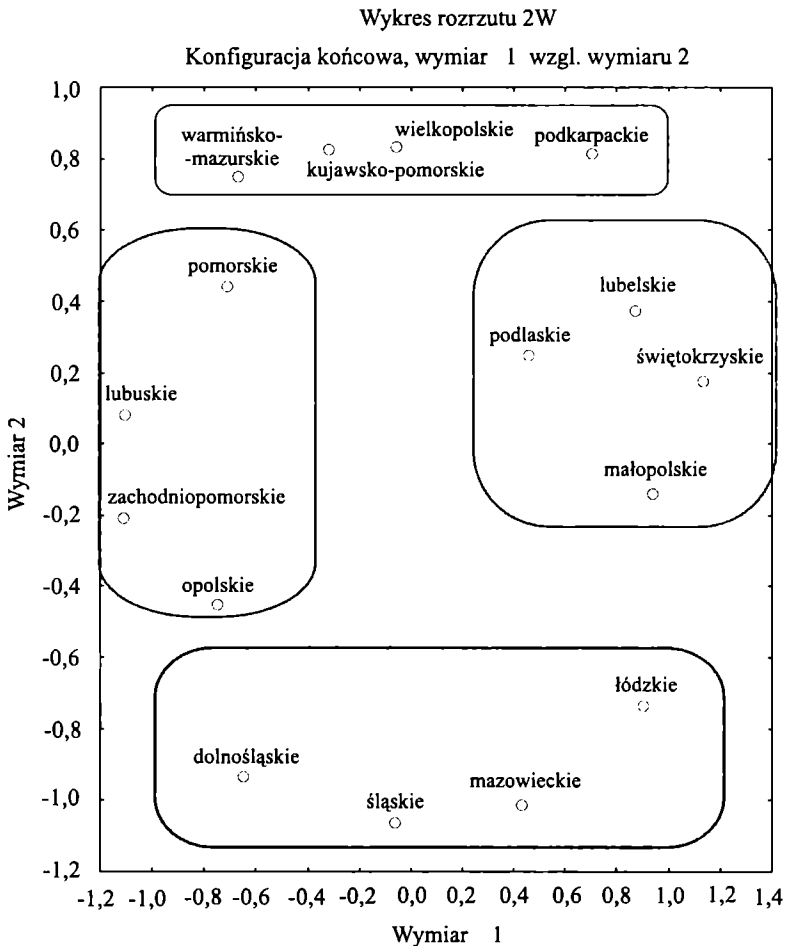


Rys. 11. Rezultaty grupowania zmiennych traktowanych jako przypadki dla 16 województw traktowanych jako zmienne

Źródło: obliczenia własne, dane GUS.

W wyniku grupowania wszystkich zmiennych (demograficznych i ekonomicznych) otrzymano trzy następujące grupy zmiennych: grupę pierwszą, obejmującą zmienne demograficzne i związane z nimi ekonomiczne, drugą grupę zawierającą zmienne związane z rozwojem cywilizacyjnym.

Należy zauważyć, że w każdej z wymienionych grup występuje zmienna demograficzna, co można uznać za interesującą prawidłowość: zmienne demograficzne są ściśle związane z rozwojem gospodarczym, cywilizacyjnym i kulturalnym. Dzięki temu autorzy otrzymali potwierdzenie podanej na wstępie hipotezy badawczej. W wyniku tego podejścia uzyskano wzajemne relacje pomiędzy województwami, co ilustruje rys. 12.



Rys. 12. Relatywne pozycje województw w procesie rozwoju – skalowanie w odniesieniu do transportowanej macierzy obserwacji

Źródło: opracowanie własne, dane GUS.

Analizując rys. 12, można zauważyć grupy województw o charakterystycznych silnych wzajemnych związkach. Innymi słowy, można powiedzieć, że ze względu na rozpatrywane uprzednio zmienne poziomy rozwoju województw wchodzących w skład wyodrębnionych grup ciążą ku sobie.

Charakterystyczną grupę tworzą województwa: lubelskie, podlaskie, świętokrzyskie, małopolskie i podkarpackie. Do drugiej grupy można zaliczyć województwa kujawsko-pomorskie, wielkopolskie, warmińsko-mazurskie i pomorskie. Trzecia grupa to województwa: lubuskie, zachodniopomorskie i opolskie. Następną grupę tworzą województwa: dolnośląskie, śląskie, mazowieckie i łódzkie.

Na uwagę zasługuje to, że ponieważ podstawą skalowania jest macierz korelacji (macierz kowariancji zmiennych standaryzowanych), rys. 12 ilustruje związki pomiędzy relatywnymi pozycjami poszczególnych województw. Przez relatywne pozycje rozumie się miejsce, jakie województwo przyjmuje ze względu na ranking 34 zmiennych standaryzowanych. Przy tym ranking taki można określić jako rozwój społeczno-gospodarczy.

Z powyższych rozważań wynika także, że jeśli jedna grupa województw wykazuje rozwój, to inna grupa może wykazywać spadek rozwoju. Jest to więc związane z ograniczonymi środkami finansowymi przeznaczonymi na rozwój gospodarki polskiej.

4. Zakończenie

Jak można zauważyć, autorzy zastosowali dwa podejścia metodyczne, które, uzupełniając się wzajemnie, pozwalają na szerszą interpretację merytoryczną omawianego zagadnienia.

Pierwsze podejście umożliwiło wyodrębnienie czynników rozwoju demograficznego i społeczno-ekonomicznego. To pozwoliło w konsekwencji na pogrupowanie województw na bardziej jednorodne grupy. Wynika z tego, że pierwszy czynnik demograficzny skorelowany jest z poziomem rozwoju gospodarczego. Drugi czynnik wyraża związek rozwoju demograficznego z rozwojem raczej społeczno-kulturalnym.

Z przeprowadzonej analizy można zauważyć wyróżniające się trzy centra rozwoju gospodarczego Polski – są to: województwa mazowieckie, wielkopolskie i śląskie.

Z kolei z drugiego podejścia otrzymano dodatkowe informacje o wyodrębnionych grupach województw charakteryzujących się rozwojem rolniczym, przemysłowym i społeczno-kulturalnym. Została też wyodrębniona grupa województw, które trudno zakwalifikować do jakiegokolwiek typu rozwoju.

Literatura

- Borg I., Groenen P., *Modern Multidimensional Scaling*, Springer-Verlag, New York 2002.
- Cieślak M., *Zmienne wiodące w analizach i prognozach demograficznych w badaniach koniunktury demograficznej*, red. J. Paradysz, materiały z konferencji, Baranowo, 7-9 grudnia 1994.
- Grabiński T., Zając K., *Taksonomiczne metody określania faz rozwojowych procesów demograficznych*, „Studia Demograficzne” 1977 nr 31, Warszawa.
- Hicks I.R., *The Social Framework, an Introduction to Economics*, Oxford 1948.
- Kalecki M., *Teoria dynamiki gospodarczej*, Warszawa 1958.
- Kosiorowski D., Zając K., *Czynniki kulturowe w rozwoju społeczno-gospodarczym*, materiały z Konferencji Naukowej pt. „Klasyfikacja i analiza danych – teoria i zastosowania”, Białowieża, 15-17 września 2004.
- Kuznetz S., *Wzrost gospodarczy narodowy. Produkt i struktura produkcji*, Warszawa 1976.
- Liu R.Y., Parelius J.M., Singh K., *Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference*. „The Annals of Statistics” 1999, vol. 27, nr 3, s. 783-858.
- Poincare H., *Nauka i hipoteza*, Warszawa 1908.
- Rao Radhakrishna C., Toutenburg Helge, *Linear Models – Least Squares and Alternatives*, Springer-Verlag, New York 1999.
- Sokołowski A., Zając K., *Rozwój demograficzny a rozwój gospodarczy*, PWE, Warszawa 1987.
- Zając K., *Porównanie poziomu rozwoju demograficznego i społeczno-gospodarczego na przykładzie województw w latach 1975-1981*, „Studia Demograficzne” 1984 nr 4/78, Warszawa.

A CONTRIBUTION TO THE DISCUSSION ABOUT STUDIES ON THE BASE OF SINGULAR COVARIANCE MATRIX

Summary

The objective of the paper is to give into a consideration for the researchers some simple and an intuition appealing method of dealing with singular covariance matrix in economic surveys (e.g. we have more variables than observations). In the empirical studies conducted for Polish territorial units the presented method (among others) is used to verify the hypothesis stating that there is a relation between socio-economic and demographic development.