

# WYKORZYSTANIE ANALIZY KORESPONDENCJI W BADANIACH MARKETINGOWYCH

Agnieszka Stanimir

*Katedra Ekonometrii, Akademia Ekonomiczna im. O. Langego, Wrocław*

## 1 WPROWADZENIE

Praca jest poświęcona analizie korespondencji, wyspecjalizowanej metodzie eksploracji danych. Główne cele pracy dotyczą aspektów teoretycznych i praktycznych stosowania analizy korespondencji. Celem teoretycznym jest przedstawienie jej założeń metodologicznych i rozwiązań technicznych, co ułatwi stosowanie tej metody w badaniach prowadzonych w Polsce (punkty 2, 3, 4).

Propozycja stosowania analizy korespondencji jako narzędzia wspomagającego podejmowanie decyzji marketingowych jest celem aplikacyjnym pracy. Atrakcyjność analizy korespondencji w badaniach marketingowych wynika z szerokiego obszaru jej zastosowań oraz możliwości graficznej prezentacji wyników. Wyniki otrzymane przy jej wykorzystaniu są czytelne, co bardzo ułatwia sporządzanie raportu z badania (punkt 5, 6).

Analiza korespondencji należy do grupy wielowymiarowych metod badania współwystępowania. Uwaga badacza jest w nich skierowana na relacje zachodzące między zmiennymi lub obiektami. Analiza korespondencji rozwiązuje jedno z najtrudniejszych zadań, a mianowicie umożliwia trafne rozpoznanie współwystępowania kategorii zmiennych lub obiektów, zmierzonych na skali nominalnej. Możliwych do zastosowania metod analizy danych pochodzących ze skali najniższej jest niewiele. Dzięki wskazaniu powiązań kategorii cech analiza korespondencji rozszerza zakres badania cech nominalnych. Zaletą analizy korespondencji jest możliwość czytelnej, graficznej prezentacji współwystępowania kategorii zmiennych. Mimo swoich zalet

analiza korespondencji jest rzadko wykorzystywana w praktyce polskich badań marketingowych (podobnie jak w innych dziedzinach). Przyczyn tej sytuacji należy szukać m.in. w braku kompleksowego opracowania w języku polskim, prezentującego zarówno samą metodologię, jak też sposoby oceny i prezentacji wyników.

Biorąc pod uwagę, że prowadzenie badania marketingowego jest postępowaniem wieloetapowym, należy wskazać, w którym kroku następuje wprowadzenie i wykorzystanie analizy korespondencji. Przyjmując, że etapy badania są następujące (na podst. [7, 15, 18, 20, 22]):

- 1 określenie problemu i sformułowanie hipotez badawczych;
- 2 ustalenie możliwych rozwiązań, przygotowanie do gromadzenia danych;
- 3 zebranie danych;
- 4 redakcja i analiza danych;
- 5 przygotowanie i zaprezentowanie raportu;

to analiza korespondencji pojawia się w etapie 2 (gdyż od znajomości metod analitycznych zależy wykorzystanie i budowa instrumentów pomiarowych), praktycznie wykorzystana jest w etapie 4 oraz wspomaga realizację etapu 5.

W trakcie badania marketingowego gromadzone są dane wtórne oraz pierwotne. Analizę korespondencji pierwszego rodzaju wymienionych danych zaprezentowano w punkcie 5, natomiast przykład badania marketingowego opartego na danych pierwotnych przedstawiono w punkcie 6.

## 2 ANALIZA KORESPONDENCJI. POJĘCIA, DEFINICJE I TECHNIKI

Analiza korespondencji swój rozwój zawdzięcza zaproponowaniu w 1900 roku przez Pearsona statystyki  $\chi^2$  do testowania dopasowania modelu przez porównanie częstości oczekiwanych z zaobserwowanymi [6]. W 1924 roku Fisher wskazał liczbę stopni swobody  $(I-1)(J-1)$  do testowania przy wykorzystaniu statystyki  $\chi^2$  Pearsona niezależności dwóch cech o liczbie kategorii  $I$  oraz  $J$  [9]. Kolejnym etapem pracy Pearsona jest zdefiniowanie w 1904 roku miary *mean square contingency*:

$$\phi^2 = \frac{\chi^2}{n}$$

jako siły zależności między cechami [16]. Do rozwoju analizy korespondencji przyczynili się również Guttman, Burt, Benzécricie oraz Greenacre.

Analiza korespondencji jest techniką eksploracyjną, dzięki której są identyfikowane jednocześnie wystąpienia poszczególnych kategorii rozpatrywanych zmiennych nominalnych. Wyniki analizy korespondencji są przedstawiane graficznie [8, 12]. Jest to możliwe dzięki naniesieniu punktów obrazujących kategorie cech na jedno-, dwu- lub trójwymiarowy układ współrzędnych, przy utracie jak najmniejszej części z informacji o rzeczywistej strukturze powiązań między nimi.

Chcąc przedstawić zróżnicowane formy przeprowadzania analizy korespondencji wielu zmiennych nominalnych, należy zaprezentować podstawy metodologiczne stosowane w podejściu klasycznym, czyli w trakcie analizy dwóch cech nominalnych.

Punktem wyjścia w klasycznej analizie korespondencji (ang. *simple correspondence analysis*) jest zapisanie zaobserwowanych liczebności jednoczesnych wystąpień kategorii cech  $R$  oraz  $C$ , w postaci tablicy kontyngencji:

$$N = [n_{ij}]$$

gdzie  $i=(1, 2, \dots, r), j=(1, 2, \dots, c)$ .

Następnie wyznaczane są liczebności brzegowe wierszy:

$$n_{i\bullet} = \sum_{j=1}^c n_{ij}; (n_{\bullet j} = \sum_{i=1}^r n_{ij}).$$

Na tej podstawie obliczane są częstości brzegowe wierszy (kolumn):

$$p_{i\bullet} = \sum_{j=1}^c p_{ij} = \sum_{j=1}^c \frac{n_{ij}}{n} = \frac{n_{i\bullet}}{n};$$

$$(p_{\bullet j} = \sum_{i=1}^r p_{ij} = \sum_{i=1}^r \frac{n_{ij}}{n} = \frac{n_{\bullet j}}{n}).$$

gdzie  $n$  jest liczbą badanych obiektów.

Częstości brzegowe wierszy zapisuje się w wektorze:

$$\mathbf{r} = [p_{i\bullet}],$$

a kolumn w wektorze:

$$\mathbf{c} = [p_{\bullet j}].$$

Liczebności oczekiwane obliczane są następująco:

$$n \cdot p_{i\bullet} \cdot p_{\bullet j}.$$

Ważnym elementem analizy korespondencji jest obliczenie macierzy profili wierszowych:

$$\mathbf{R} = \begin{bmatrix} \frac{n_{ij}}{n_{i\bullet}} \\ \frac{n_{ij}}{n_{i\bullet}} \end{bmatrix} = \begin{bmatrix} \frac{p_{ij}}{p_{i\bullet}} \\ \frac{p_{ij}}{p_{i\bullet}} \end{bmatrix} = \mathbf{D}_r^{-1} \mathbf{P}$$

i kolumnowych:

$$\mathbf{C} = \begin{bmatrix} \frac{n_{ij}}{n_{\bullet j}} \\ \frac{n_{ij}}{n_{\bullet j}} \end{bmatrix} = \begin{bmatrix} \frac{p_{ij}}{p_{\bullet j}} \\ \frac{p_{ij}}{p_{\bullet j}} \end{bmatrix} = \mathbf{D}_c^{-1} \mathbf{P}.$$

W obydwu przedstawionych wzorach macierz:

$$\mathbf{P} = [n_{ij} / n]$$

to macierz korespondencji zawierająca częstości zaobserwowane. Natomiast  $\mathbf{D}_r$  i  $\mathbf{D}_c$  to macierze diagonalne o elementach, odpowiednio:

$$[p_{i\bullet}] \text{ i } [p_{\bullet j}].$$

W celu wskazania współrzędnych umożliwiających umieszczenie punktów obrazujących kategorie w przestrzeni o wybranym wymiarze należy posłużyć się rozkładem macierzy według wartości osobliwych. Dekompozycja przebiega następująco:

$$\mathbf{A} = \mathbf{D}_r^{-1/2} (\mathbf{P} - \mathbf{r}\mathbf{c}^T) \mathbf{D}_c^{-1/2} = \mathbf{U}\mathbf{V}^T,$$

gdzie  $\Gamma$  jest diagonalną macierzą niezerowych wartości osobliwych macierzy  $\mathbf{A}$ , ułożonych w porządku nierosnącym,  $\mathbf{U}$  ( $\mathbf{V}$ ) jest macierzą

lewych (prawych) wektorów osobliwych macierzy  $\mathbf{A}$ . Wartości osobliwe macierzy  $\mathbf{A}$  odpowiadają pierwiastkom kwartowym wartości własnych macierzy  $\mathbf{A}\mathbf{A}^T$  oraz  $\mathbf{A}^T\mathbf{A}$ . Wartości własne są zapisywane w diagonalnej macierzy  $\mathbf{\Lambda}$ . Rzeczywista przestrzeń prezentacji powiązań między kategoriami dwóch cech może być najwyżej  $K$ -wymiarowa, gdzie

$$K = \min(r-1; c-1).$$

Natomiast gdy liczebności kategorii cech są zapisane w macierzy Burta, to

$$K = \sum_{q=1}^Q (J_q - 1).$$

Liczba niezerowych wartości własnych, jak również osobliwych jest równa  $K$ . Współrzędne kategorii zapisanych w wierszach ( $\mathbf{F}$ ) i kolumnach ( $\mathbf{G}$ ) w przestrzeni  $K$ -wymiarowej oblicza się następująco:

$$\mathbf{F} = \mathbf{D}_r^{-1/2} \mathbf{U} \mathbf{\Gamma}$$

oraz

$$\mathbf{G} = \mathbf{D}_c^{-1/2} \mathbf{V} \mathbf{\Gamma}.$$

W macierzach  $\mathbf{F}$  oraz  $\mathbf{G}$  wiersze odpowiadają kategoriom cech zapisanych odpowiednio w wierszach i kolumnach tablicy kontyngencji, natomiast kolumny – współrzędnym na kolejnych osiach głównych. Suma wszystkich wartości własnych jest równa średniokwadratowej wielodzielczości Pearsona:

$$\phi^2 = \frac{\chi^2}{n} = \sum_{k=1}^K \lambda_k = \lambda,$$

gdzie  $\lambda_k$  to  $k$ -ta wartość własna. Suma wartości własnych jest w analizie korespondencji inercją całkowitą ( $\lambda$ ). Wartości własne ( $\lambda_k$ ) są nazywane inercjami głównymi.

Metodologia klasycznej analizy korespondencji jest szczegółowo omawiana w pracach [4, 5, 10, 11, 13, 14, 17].

### 3 ANALIZA KORESPONDENCJI WIELU ZMIENNYCH NOMINALNYCH

Zastosowanie klasycznej analizy korespondencji jest możliwe w przypadku wystąpienia dwóch zmiennych. Najczęściej jednak w wyniku prowadzenia pomiarów zbioru cech, które wymagają

szczegółowej analizy jest większy. Jeśli kategorie cech są mierzone na skali nominalnej, to dla każdej pary zmiennych można przeprowadzić klasyczną analizę korespondencji i w ten sposób określić współwystępowanie kategorii par zmiennych. Częściej jednak interesujące dla badacza jest jednoczesne zanalizowanie zaobserwowanych wariantów wszystkich cech dla ogółu obiektów.

Wyznaczenie współrzędnych kategorii w analizie korespondencji wielu zmiennych jest przeprowadzane analogicznie do metodologii stosowanej w podejściu klasycznym. Różnice występują w sposobie zapisania informacji o zaobserwowanych liczebnościach wystąpień kategorii badanych cech. W klasycznej analizie korespondencji rozpoczyna się od zapisania tablicy kontyngencji. Tak samo należy postąpić w analizie wielu zmiennych nominalnych, konstruując odpowiednią tabelę bądź macierz.

#### 3.1 Złożona macierz znaczników

Macierz znaczników to macierz  $n \times j$ , gdzie  $n$  jest liczbą obserwacji (przypadków, respondentów, obiektów),  $j$  jest to liczba kategorii cechy; każdy wiersz tej macierzy zawiera tylko zera i jedną jedynkę w miejscu oznaczającym wybór określonej kategorii. Złożona macierz znaczników  $\mathbf{Z}$  składa się z bloków (podmacierzy) odpowiadających kolejnym zmiennym:

$$\mathbf{Z} = [\mathbf{Z}_1 \dots \mathbf{Z}_Q],$$

gdzie:

$\mathbf{Z}_1 \dots \mathbf{Z}_Q$  – macierze znaczników kolejnych cech,

$Q$  – liczba cech.

Liczba wierszy w każdej macierzy znaczników (podmacierzy) jest taka sama i równa liczbie badanych jednostek (respondentów, obiektów). Natomiast liczba kolumn każdego bloku jest inna i równa liczbie kategorii cechy, której odpowiada blok.

Elementy złożonej macierzy znaczników przyjmują tylko wartości 0 i 1. Liczba jedynek w każdej podmacierzy jest równa liczbie obiektów. W wierszu każdego bloku może wystąpić tylko jedna jedynka, w tej kolumnie, która jest przypisana kategorii, odpowiadającej charakterystyce badanego obiektu.

### 3.2 Macierz Burta

Po utworzeniu złożonej macierzy znaczników  $\mathbf{Z}$ , macierz Burta jest wynikiem działania:

$$\mathbf{B} = \mathbf{Z}^T \mathbf{Z}.$$

Powstaje symetryczna macierz blokowa, w której na głównej przekątnej umieszczone są macierze diagonalne, zawierające liczebności wystąpień kategorii cech, a poza przekątną położone są tablice kontyngencji dla każdej pary rozpatrywanych zmiennych.

### 3.3 Wielowymiarowa tablica kontyngencji

Wielowymiarowa tablica kontyngencji jest krzyżową klasyfikacją wielu zmiennych [1, 14]. Rozpatrując trzy cechy, tablicę taką można stworzyć z wierszy, kolumn i warstw. Sposób konstrukcji wielowymiarowej tablicy kontyngencji komplikuje się wraz ze wzrostem liczby analizowanych cech. Warstwy mogą wystąpić również w kolumnach. By zbudować taką tablicę, należy dysponować danymi o liczbie jednoczesnych wystąpień wszystkich kombinacji kategorii różnych cech.

### 3.4 Łączona tablica kontyngencji

Budowanie łączonej tablicy kontyngencji jest możliwe dzięki skorzystaniu z reguły addytywności statystyki  $\chi^2$ , tzn. gdy grupy danych dotyczą zbliżonego, pod względem wielkości i struktury, doświadczenia, to wartości statystyk  $\chi_i^2$  uzyskane z każdego ( $i$ -tego) pomiaru można zsumować i na tej podstawie wyciągnąć wnioski o całym rozpatrywanym zjawisku [21]. Sumowaniu w takim przypadku podlegają również wartości stopni swobody.

Zakładając, że każda z tablic składowych ma w kolumnach tę samą cechę wspólną, to kolumny łączonej tablicy kontyngencji będą odpowiadały kategoriom cechy wspólnej, a wiersze kolejnym kategoriom pozostałych cech.

## 4 TECHNIKI PREZENTACJI, OCENY I INTERPRETACJI WYNIKÓW

Ważnym etapem badania marketingowego jest wykonanie i zaprezentowanie raportu. Jest z tym związanych wiele czynności. Decyzja o wykorzystaniu otrzymanych w trakcie badania wyników jest podejmowana po ich interpretacji,

poprzedzonej pozytywną oceną poprawności przeprowadzonych analiz. Zinterpretowane rezultaty i wnioski z wykonanych analiz należy przedstawić w najbardziej zrozumiałej formie.

### 4.1 Sposoby wyboru wymiaru przestrzeni rzutowania

Niezależnie od przyjętego sposobu przeprowadzenia analizy korespondencji efektem wykonanych obliczeń są macierze zawierające współrzędne kategorii na wszystkich osiach przestrzeni rzeczywistych powiązań. Najczęściej wymiar tej przestrzeni jest bardzo duży. Problematyczne zatem okazuje się zaprezentowanie powiązań występujących między kategoriami w przestrzeni  $K$ -wymiarowej. Ogólnie przyjętym rozwiązaniem w analizie korespondencji jest rzutowanie punktów na przestrzeń o wymiarze niższym. Wyróżnia się trzy kryteria wyboru wymiaru przestrzeni rzutowania.

- 1 Stopień wyjaśniania inercji. Należy sprawdzić, jaki procent inercji całkowitej stanowią w sumie inercje główne wybranego wymiaru ( $K^*$ ) przestrzeni rzutowania. Przyjmuje się, że najlepszym rozwiązaniem jest wybranie takiego  $K^*$ , gdy stopień wyjaśniania inercji jest bliski 1, a po zwiększeniu wymiaru przestrzeni wartość tego wskaźnika nie wzrasta gwałtownie [2, 13, 19].
- 2 Kryterium łokcia. By wyznaczyć najlepszy wymiar rzutowania, należy najpierw narysować wykres wszystkich niezerowych wartości własnych w porządku nierosnącym. Jeżeli na wykresie zaobserwowano spadek wartości, a następnie spłaszczenie krzywej, to właściwą przestrzenią rzutowania jest przestrzeń wskazana przez numer wartości własnej, dla której nastąpiło zagięcie (uskok) [5].
- 3 Kryterium liczby cech. Jako najlepszy jest wybierany ten wymiar rzutowania kategorii zmiennych, w którym wartości własne są większe od odwrotności liczby analizowanych cech [11]. Kryterium to może być stosowane tylko w trakcie analizy korespondencji wykonanej na podstawie danych zapisanych w macierzy Burta lub złożonej macierzy znaczników.

#### 4.2 Ocena jakości odwzorowania współwystępowania kategorii cech

Istotnym elementem w trakcie wykonywania analizy korespondencji jest sprawdzenie, czy otrzymane wyniki odzwierciedlają rzeczywiste powiązania występujące między kategoriami zmiennych. Niejednokrotnie stopień wyjaśniania inercji jest wykorzystywany jako jedyne kryterium oceny przeprowadzonych działań. Takie postępowanie jest jednak zbyt uproszczone.

Ważną metodą oceny jakości odwzorowania jednoczesnego wystąpienia kategorii cech jest porównanie liczebności rzeczywistych z liczebnościami wyznaczonymi za pomocą macierzy odtworzenia. Liczebności tej macierzy są obliczane na podstawie wartości własnych i współrzędnych wybranego wymiaru przestrzeni rzutowania [3, 10, 11, 13].

Jakość odwzorowania rzeczywistych powiązań kategorii cech w przestrzeni  $K^*$ -wymiarowej można określić analizując udział  $k$ -tej osi głównej w zobrazowaniu wystąpień wybranego wariantu zmiennej. Miara ta jest umownie nazwana korelacją punktu z osią [5, 17]. Na podstawie wartości korelacji punktu z osią wyznaczany jest miernik jakości odzwierciedlenia rzeczywistych wystąpień kategorii w wybranej przestrzeni. Jeśli wybrana najlepsza przestrzeń rzutowania jest  $K^*$ -wymiarowa, to jakość odwzorowania  $i$ -tej kategorii można zmierzyć, sumując wartości kwadratów korelacji  $i$ -tej kategorii w przestrzeni  $K^*$ -wymiarowej. Suma kwadratów korelacji punktów ze wszystkimi osiami jest równa 1.

#### 4.3 Interpretacja wyników analizy korespondencji

Najczęściej stosowanym sposobem przedstawienia efektów wykonanej analizy korespondencji jest graficzna prezentacja jednoczesnych wystąpień kategorii zmiennych. Interpretacja otrzymanych wyników polega na ocenie położenia punktów obrazujących kategorie zmiennych na wykresie. Pod uwagę należy wziąć trzy elementy:

- położenie punktu wobec centrum rzutowania;
- położenie punktu względem innych punktów określających kategorie należące do tej samej cechy;
- położenie punktu względem punktu opisującego kategorię innej cechy.

Jeśli wyniki analizy korespondencji są przedstawiane w układzie współrzędnych, to w celu wskazania grup punktów leżących najbliżej siebie, czyli kategorii powiązanych, można wykorzystać metody klasyfikacji. Są one bardzo pomocne w interpretacji wyników analizy korespondencji, gdy w celu wykonania ostatecznego rzutowania punktów należy skorzystać z przeszerzenia co najmniej czterowymiarowej. Klasyfikacja kategorii jest przeprowadzana bez ograniczenia liczby współrzędnych, a prezentacja rezultatów jest czytelna.

Gdy wyniki analizy korespondencji są prezentowane jako rozrzut punktów w układzie współrzędnych, to każdej osi wybranej przeszerzenia rzutowania można przyporządkować nazwy. Nazwanie osi nie jest warunkiem kompletności analizy korespondencji i nie zawsze jest to możliwe. Nazwy (etykiety) są przyporządkowywane osiom na podstawie ich korelacji z punktami i pochodzą od nazw cech lub kategorii [4, 5].

Duże możliwości poszerzenia interpretacji wyników przeprowadzonej analizy korespondencji daje zamieszczenie na wykresie punktów dodatkowych [5, 10, 11, 13]. Są to dodatkowe kategorie cech zapisanych w wierszach bądź kolumnach, dla których nie można wyznaczyć częstości brzegowych, a przez to ich wkład w inercję całkowitą jest równy zeru. Punkty dodatkowe obrazują informację, która nie jest prezentowana w zanalizowanych interakcjach kategorii cech. Dodatkowo analizowane punkty muszą być merytorycznie związane z tymi, dla których przeprowadzono analizę korespondencji. Dodatkowe punkty są też nazywane punktami pasywnymi w odróżnieniu od punktów aktywnych, obrazujących kategorie, dla których przeprowadzono analizę korespondencji. Punkty dodatkowe nie są brane pod uwagę podczas wyznaczania wartości własnych i współrzędnych punktów aktywnych.

## 5 ANALIZA STANU AKTYWNOŚCI ZAWODOWEJ LUDNOŚCI POLSKI

Jako ciekawy problem badawczy oparty na wystąpieniach danych wtórnych wybrano analizę sytuacji ekonomicznej ludności Polski ze względu na jej stan aktywności zawodowej w powiązaniu z cechami demograficznymi (wiek, płeć). W celu wykonania takiej analizy posłużono się danymi zaczerpniętymi z rocznika statystycznego

*Aktywność ekonomiczna ludności Polski, IV kwartał 2001*, wydane przez GUS w 2002 roku. Wymienione cechy przyjmują następujące kategorie:

- stan aktywności zawodowej: A1 – pracujący w pełnym wymiarze czasu, A2 – pracujący w niepełnym wymiarze czasu, A3 – bezrobotni, A4 – bierni zawodowo;
- wiek: 15-17, 18-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65 i więcej lat;
- płeć: K – kobieta, M – mężczyzna.

Przeprowadzone badania wskazały możliwości aplikacji analizy korespondencji.

### 5.1 Analiza cech: stan aktywności zawodowej i wiek

Analizę dwóch cech przeprowadzono jako pierwszą. Do tego badania wybrano cechy: stan aktywności zawodowej i wiek ze względu na naturalne powiązanie ich kategorii, czego potwierdzeniem okazały się wyniki analizy korespondencji. Dane niezbędne do przeprowadzenia analizy zapisano w tabeli 1.

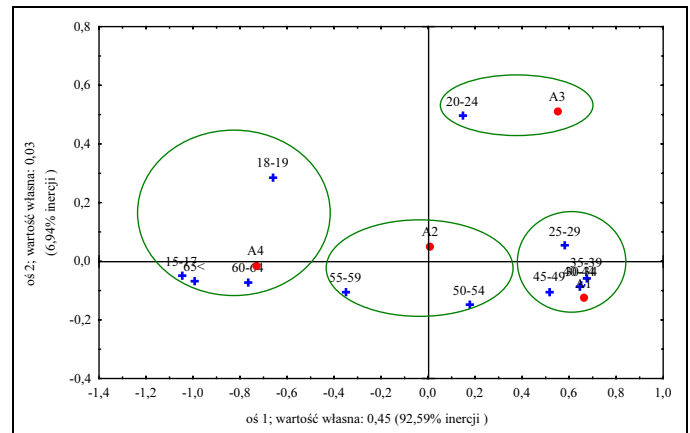
Tabela 1. Tabela kontyngencji cech: stan aktywności zawodowej i wiek oraz liczebności brzegowe wierszy ( $n_{i\cdot}$ ) i kolumn ( $n_{\cdot j}$ ).

		Stan aktywności zawodowej (w tys. osób)					$n_{i\cdot}$
		A1	A2	A3	A4		
Wiek	15-17	6	54	6	1 686	1 752	
	18-19	82	57	168	968	1 275	
	20-24	969	181	767	1 055	2 972	
	25-29	1 798	133	486	461	2 878	
	30-34	1 694	108	332	319	2 453	
	35-39	1 718	102	362	299	2 481	
	40-44	2 053	141	400	399	2 993	
	45-49	2 008	156	367	610	3 141	
	50-54	1 410	137	205	991	2 743	
	55-59	483	123	65	1 031	1 702	
	60-64	177	120	22	1 329	1 648	
	65<	133	200	7	4 514	4 854	
	$n_{\cdot j}$	12 531	1 512	3 187	13 662		

Źródło: opracowanie własne na podstawie danych GUS<sup>44</sup>.

Wyniki zastosowania klasycznej analizy korespondencji przedstawiono na rys. 1. Procent

inercji całkowitej jaki jest wyjaśniany w przestrzeni dwuwymiarowej, przez pierwsze dwie inercje główne jest prawie 100% (92,59%+6,94%=99,53%).



Rys. 1. Podział kategorii cech: stan aktywności zawodowej i wiek na cztery klasy; źródło: opracowanie własne.

Na rys. 1 są zaznaczone klasy, które wyznaczono stosując klasyfikację hierarchiczną metodą Warda na podstawie wartości współrzędnych.

Interpretacje wyznaczonych klas są następujące:

- klasa I; zaliczono do niej kategorie: A1, 25-29, 30-34, 35-39, 40-44, 45-49, co świadczy o tym, że wśród osób zatrudnionych w pełnym wymiarze czasu pracy dominują osoby w średnim wieku produkcyjnym;
- klasa II (A2, 50-54, 55-59); osoby zatrudnione w niepełnym wymiarze czasu pracy mają od 50 do 59 lat;
- klasa III (A3, 20-24); problem bezrobocia jest najbardziej znaczący dla osób między 20 a 24 rokiem życia;
- klasa IV (A4, 15-17, 18-19, 60-64, 65 i więcej); położenie tych punktów wskazuje na naturalną sytuację w społeczeństwie, czyli osoby w wieku przedprodukcyjnym, wczesnym produkcyjnym oraz emerytalnym pozostają bierni zawodowo.

### 5.2 Badanie współwystępowania kategorii cech: wiek, płeć oraz stan aktywności zawodowej

Kolejnym interesującym problemem jest zbadanie współwystępowania kategorii trzech cech. Do dwóch poprzednich dołączono kolejną – płeć. Wykorzystując wielowymiarową tablicę kontyngencji, zbadano kształtowanie się stanu aktywności zawodowej w grupach wiekowych

<sup>44</sup> W tabelach, w których zapisano dane dotyczące aktywności zawodowej ludności Polski, skorzystano z rocznika statystycznego *Aktywność ekonomiczna ludności Polski, IV kwartał 2001*, wydane przez GUS w 2002 roku, tab. 1.3-1.5, s. 4-7, oznaczonych dalej jako źródło GUS.

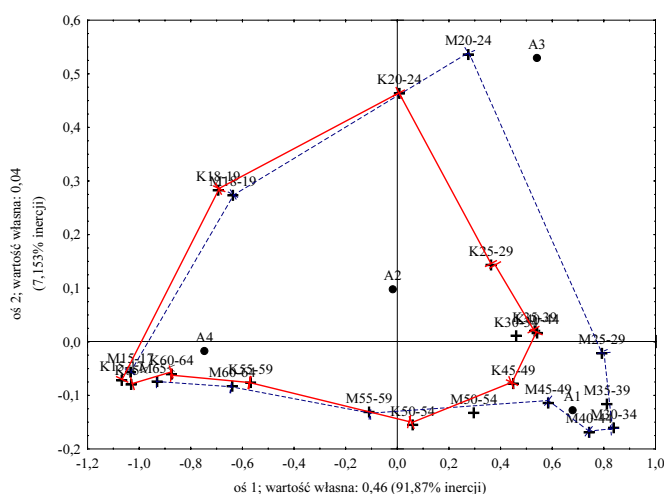
kobiet i grupach wiekowych mężczyzn. Dane niezbędne do wykonania obliczeń zapisano w tabeli 2.

Tabela 2. Wielowymiarowa tablica kontyngencji [PW][A].

Płeć	Wiek	Stan aktywności zawodowej			
		A1	A2	A3	A4
Kobiety	15-17	1	18	2	835
	18-19	33	26	83	488
	20-24	424	106	355	629
	25-29	727	89	244	364
	30-34	706	78	169	255
	35-39	758	65	187	224
	40-44	925	88	224	261
	45-49	967	82	185	359
	50-54	665	70	88	590
	55-59	171	67	25	640
	60-64	58	58	9	779
	65<	47	86	3	2 857
Mężczyźni	15-17	5	36	4	851
	18-19	49	31	85	480
	20-24	545	75	412	426
	25-29	1 071	44	242	97
	30-34	988	30	163	64
	35-39	960	37	175	75
	40-44	1 128	53	176	138
	45-49	1 041	74	182	251
	50-54	745	67	117	401
	55-59	312	56	40	391
	60-64	119	62	13	550
	M65<	86	114	4	1 657

Źródło: opracowanie własne na podstawie danych GUS.

Wyniki analizy korespondencji zaprezentowano w przestrzeni dwuwymiarowej na rys. 2.



Rys. 2. Prezentacja powiązań kategorii z kombinacji cech: płeć i wiek oraz wariantów stanu aktywności zawodowej; źródło: opracowanie własne.

Porównując rys. 1 i 2, można stwierdzić, że punkty obrazujące wiek są ułożone względem stanu aktywności zawodowej w analogiczny sposób. Na rys. 2 połączono punkty *wieku*, według wartości rosnących (dla kobiet odcinkami ciągłymi, dla mężczyzn – przerywanymi). Dzięki temu zauważono, że pomiędzy aktywnością zawodową kobiet i mężczyzn od 20 do 54 roku życia występują różnice. Problem bezrobocia dotyka w większej mierze młodych mężczyzn niż młode kobiety. Zatrudnienie w pełnym wymiarze czasu pracy jest bardziej charakterystyczne dla mężczyzn w wieku od 25 do 49 lat, gdyż punkty dotyczące kobiet od 25 do 49 roku życia są przesunięte w kierunku kategorii A2, czyli zatrudnienia w niepełnym wymiarze czasu pracy. Kobietom w wieku 55-59 bliska jest kategoria A4 (bierne zawodowo), a mężczyźni w tym samym wieku pozostają jeszcze aktywni zawodowo. Sytuacja w aktywności zawodowej kobiet i mężczyzn upodabnia się po skończeniu przez nich 59 lat.

## 6 BADANIE MOTYWÓW PODJĘCIA STUDIÓW WYŻSZYCH W AKADEMII EKONOMICZNEJ WE WROCŁAWIU

Celem przeprowadzanego badania jest poznanie motywów, jakimi kierowali się studenci, wybierając kształcenie w trybie zaocznym na kierunkach studiów wyższych oferowanych przez Akademię Ekonomiczną we Wrocławiu. Kryteria wyboru trybu i rodzaju studiów skonfrontowano z oczekiwaniami, jakie studenci wiążą z uzyskaniem dyplomu ukończenia studiów (możliwość kontynuacji nauki, rozwój kariery zawodowej). Uzyskane wyniki badań mogą wskazać kierunki modernizacji oferty uczelni w zakresie studiów zaocznych w kolejnych semestrach, tak by najlepiej dostosować ją do oczekiwań studentów.

W badaniu skorzystano głównie z danych pierwotnych. W celu ich zgromadzenia przeprowadzono ankietę audytoryjną. Respondentom zadano następujące pytania:

- 1 W jaki sposób dotarłaś/eś do informacji o rodzajach studiów proponowanych przez (wybraną) szkołę (A – oferta była dostępna u pracodawcy; B – opowiadania znajomych; C – ogłoszenie w prasie; D – inne)?
- 2 Co było główną przyczyną rozpoczęcia nauki na studiach wyższych/licencjackich (A – chęć

- nauki; B – uznanie otoczenia; C – szansa na awans; D – wyższa pensja; E – znalezienie pracy)?
- 3 Dlaczego wybrałaś/eś tę szkołę (A – zainteresowanie przedmiotami; B – prestiż uczelni; C – łatwe egzaminy i studia; D – znajomi wybrali tę szkołę; E – inne)?
  - 4 Czy uważasz, że poziom Twojej wiedzy w miarę uczęszczania na kolejne zajęcia wzrasta (A – tak; B – nie)?
  - 5 Czy uważasz, że studia pozwolą Ci osiągnąć wiedzę potrzebną do zdobycia wymarzonej pracy (A – tak; B – nie; C – nie wiem)?
  - 6 Czy już pracujesz (A – tak; B – dorywczo; C – nie)?
  - 7 Czy po skończeniu studiów chcesz kontynuować naukę (A – tak; B – nie; C – nie wiem)?
  - 8 Kto ponosi koszty Twojej nauki (A – rodzice; B – samodzielnie; C – samodzielnie+rodzice; D – pracodawca w części lub w całości, stypendia)?
  - 9 Płeć (A – kobieta; B – mężczyzna).
  - 10 Wykształcenie rodziców (A – przynajmniej jedno z rodziców ma wykształcenie wyższe; B – przynajmniej jedno z rodziców ma wykształcenie średnie, ale drugie nie ma wyższego; C – rodzice mają wykształcenie zawodowe lub podstawowe).

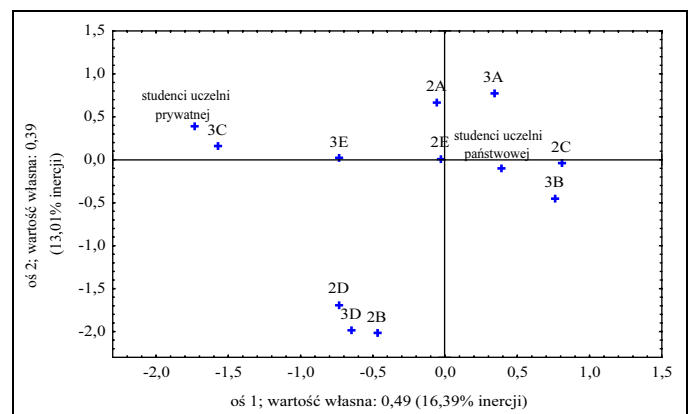
Ankieta przeprowadzona wśród studentów Akademii Ekonomicznej we Wrocławiu, którzy podjęli naukę na Magisterskich Studiach Uzupełniających (MSU) we Wrocławiu (390 osób), Magisterskich Studiach Zaocznych (MSZ) we Wrocławiu (79 osób), MSU w Wałbrzychu (125 osób), MSZ w Głogowie (210 osób) oraz 177 studentów Wyższej Szkoły Zarządzania Edukacją.

Przed przystąpieniem do analizy zbadano, czy wśród wybranych do kwestionariusza pytań występują pytania związane tylko z celem badania, czy też można wskazać pytania pomocnicze. Badanie dotyczy motywów wyboru studiów zaocznych proponowanych przez Akademię Ekonomiczną oraz przyczyn rozpoczęcia nauki. Interpretacja wyników po wykonaniu analizy ma wskazać najlepszą drogę dotarcia do przyszłych studentów z ulepszoną ofertą studiów. Na tej podstawie można stwierdzić, że najbardziej istotne są odpowiedzi na pytania 1, 2, 3, 6, 7, 8. Pozostałe pytania mają charakter pomocniczy.

### 6.1 Wybór szkoły w zależności od motywów podjęcia nauki

Przeprowadzono analizę korespondencji przyczyn rozpoczęcia nauki na studiach wyższych, występujących podczas podejmowania decyzji o wyborze szkoły prywatnej i państwowej (pytanie 2 i 3).

Szkoła prywatna (zob. rys. 3) jest wśród respondentów popularna ze względu na łatwe egzaminy i zaliczenia przedmiotów (3C). Uczelnia państwowa nie jest tak silnie jak szkoła prywatna powiązana z żadną kategorią odpowiedzi analizowanych dwóch pytań. Wynika to z rozłożenia postaw prezentowanych przez studentów na większą liczbę kategorii. Można jednak stwierdzić, że uczelnia państwowa jest wybierana ze względu na prestiż (3B) i przedmioty, które są wykładane w ramach studiów (3A). Motywacją do podjęcia studiów w uczelni państwowej są: chęć nauki (2A), zwiększenie szans na awans (2C) czy znalezienie pracy (2E).



Rys. 3. Prezentacja wyników analizy jednoczesnych wystąpień wariantów przyczyn rozpoczęcia studiów w uczelni prywatnej i państwowej; źródło: opracowanie własne.

### 6.2 Kompleksowa analiza danych pierwotnych

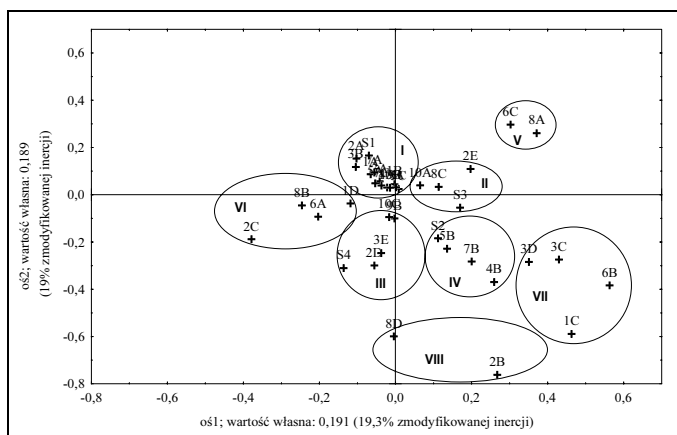
Kolejna przeprowadzona analiza dotyczyła rozpoznania motywów kształcenia i wyboru Akademii Ekonomicznej we Wrocławiu. Analizę wykonano na podstawie wszystkich zawartych w ankiecie pytań.

Tak jak podano, istotnymi pytaniami dla wyjaśnienia problemów badania są: 1, 2, 3, 6, 7, 8. Po zapisaniu w macierzy Burta liczebności wystąpień odpowiedzi na te pytania przeprowadzono analizę korespondencji. Do wymienionych sześciu cech dodano jeszcze jedną, uwzględniając tym samym podział studentów

Akademii Ekonomicznej zgodnie z miejscem prowadzenia zajęć.

Przed interpretacją otrzymanych wyników dołączono do analizy kategorie odpowiedzi pozostałych pytań (4, 5, 9, 10). Odpowiedzi na te pytania uznano za punkty dodatkowe. Graficzna prezentacja zależności jest przedstawiona na rys. 4.

Ponieważ punktów obrazujących kategorie jest bardzo dużo, to w celu wyznaczenia powiązanych wariantów odpowiedzi skorzystano z klasyfikacji hierarchicznej, którą wykonano metodą Warda.



Rys. 4. Prezentacja wyników analizy wszystkich cech (S1 – AE Wrocław MSU, S2 – AE Wrocław MSZ, S3 – AE Głogów MSZ, S4 – AE Wałbrzych MSU); źródło: opracowanie własne.

Pierwsza z wyróżnionych klas charakteryzuje osoby uczące się na magisterskich studiach uzupełniających we Wrocławiu, które dokonały wyboru szkoły ze względu na prestiż uczelni i zainteresowanie przedmiotami kierunkowymi. Głównym powodem, dla którego studenci podjęli naukę, była chęć zdobywania wiedzy. Obecne studia dla części respondentów nie będą ostatnimi, ponieważ po otrzymaniu dyplomu chcą kontynuować naukę. Klasa ta jest odpowiednia dla osób, które uważają, że ukończenie studiów ułatwia znalezienie pracy oraz że poziom ich wiedzy wzrasta w miarę uczęszczania na kolejne zajęcia. Z tą grupą są związane głównie kobiety.

Kolejna grupa opisuje osoby, które rozpoczęły studia po to, by znaleźć pracę, a koszty nauki ponoszą wspólnie z rodzicami. Klasa ta jest charakterystyczna dla studentów z Głogowa.

Klasa III zawiera studentów, którzy studiują z nadzieją, że wpłynie to na wzrost ich pensji, wybrali sale dydaktyczne w Wałbrzychu

z uwagi na bliskość miejsca zamieszkania. Grupa ta jest charakterystyczna dla mężczyzn.

Czwarta klasa obejmuje studentów magisterskich studiów zawodowych we Wrocławiu, którzy nie planują kontynuacji nauki, co może wynikać z negatywnej oceny swojej wiedzy.

Następne klasy nie dotyczą żadnego z oddziałów Akademii Ekonomicznej, czyli nie charakteryzują uczących się tu studentów.

### 6.3 Podsumowanie

Analiza porównawcza uczelni państwowej i prywatnej wskazała, że Akademia Ekonomiczna skupia studentów, którzy chcą zdobyć wiedzę i wierzą, że właśnie ta uczelnia jest do tego najodpowiedniejsza.

Wyodrębnione w trakcie kolejnej analizy klasy pozwoliły na szczegółowe określenie studentów najbardziej zainteresowanych studiami w Akademii Ekonomicznej. Dla badania ogromne znaczenie mają pierwsze dwie z wyróżnionych klas. Opisują osoby, dla których wiedza zdobywana na studiach jest cenna, wybrały uczelnię w sposób świadomy, samodzielnie podejmując decyzję. Ważną informacją jest to, że pewna część tych osób chce kontynuować naukę. Źródłem informacji o studiach były dla tych osób oferty dostępne u pracodawcy, w poprzedniej szkole lub w Internecie.

Oferta magisterskich studiów uzupełniających w Akademii Ekonomicznej powinna być skierowana do osób, które chcą rozwijać swoje zainteresowania zgodnie z profilem „szanowanej” uczelni, wskazując na możliwość kontynuowania nauki. Umiejętnie skonstruowane ogłoszenie o naborze na studia pozwoli przekonać osoby, które się jeszcze wahają, że zachowanie ciągłości studiów zwiększa możliwość znalezienia pracy.

W Akademii Ekonomicznej studiują w większości kobiety. W ofercie studiów można zatem wskazać, że dzięki zdobytym kwalifikacjom wzrastają ich szanse na zatrudnienie w zawodach postrzeganych obecnie jako typowo „męskie”, takich jak: doradca inwestycyjny, makler giełdowy czy projektant rozwiązań informatycznych. Ukończenie odpowiednio dobrego kierunku studiów gwarantuje najlepszą bazę wiedzy do podjęcia specjalistycznych kursów państwowych (biegły rewident, biegły księgowy).

## BIBLIOGRAFIA

- 1) Andersen, E.B. 1997. *Introduction to the Statistical Analysis of Categorical Data*. Berlin: Springer-Verlag.
- 2) Andersen, E.B. 1991. *The Statistical Analysis of Categorical Data*. Berlin, Springer-Verlag.
- 3) Benzécrie, J.-P. 1992. *Correspondence Analysis Handbook*. New York, Marcel Dekker Inc.
- 4) Blasius, J. 2001. *Korrespondenzanalyse*. München, Oldenbourg Verlag.
- 5) Clausen, S.E. 1998. *Applied Correspondence Analysis. An Introduction*. Sage, University Paper 121.
- 6) Cressie, N., Read, T.T.C. 1989. *Pearson's  $\chi^2$  and the Loglikelihood Ratio Statistic  $G^2$ : A Comparative Review*. *International Statistical Review*. 57(1), 19-43.
- 7) Garbarski, L., Rutkowski, I., Wrzosek, W. 2000. *Marketing. Punkt zwrotny nowoczesnej firmy*. Warszawa, PWE.
- 8) Goodman, L.A. 1986. *Some Useful Extensions of the Usual Correspondence Analysis Approach and Usual Log-Linear Models Approach in the Analysis of Contingency Tables*. *International Statistical Review*. 54(3), 243-309.
- 9) Goodman, L.A. 1968. *The Analysis of Cross-Classified Data: Independence, Quasi-Independence, and Interactions in Contingency Tables with or without Missing Entries*. *Journal of the American Statistical Association*. 324(63), 1091-1131.
- 10) Greenacre, M. 1993. *Correspondence Analysis in Practice*. London, Academic Press.
- 11) Greenacre, M. 1984. *Theory and Applications of Correspondence Analysis*. London, Academic Press.
- 12) Greenacre, M., Hastie T. 1987. *The Geometric Interpretation of Correspondence Analysis*. *Journal of the American Statistical Association*. 82(398), 437-447.
- 13) Heijden van der, P.G.M. 1987. *Correspondence Analysis of Logitudinal Categorical Data*. Leiden, DSWO Press.
- 14) Jobson, J.D. 1992. *Applied Multivariate Data Analysis. Vol. II: Categorical and Multivariate Methods*. New York, Springer-Verlag.
- 15) Kotler, Ph. 1999. *Marketing, analiza planowanie, wdrażanie i kontrola*. Warszawa, Gebethner i Ska.
- 16) Lancaster, H.O. 1958. *The structure of Bivariate Distributions*. *The Annals of Mathematical Statistics*. 29(3), 719-736.
- 17) Lebart, L. Morineau, A., Warwick, K. M. 1984. *Multivariate Descriptive Statistical Analysis. Correspondence Analysis and Related Techniques for Large Matrices*. New York, John Wiley & Sons, Inc.
- 18) Malhotra, N.K. 1999. *Marketing Research. An Applied Orientation*. Upper Saddle River, New Jersey, Prentice Hall.
- 19) Sikorki, A. 1987. *Modelowanie zachowań nabywczych – metody analizy wielowymiarowej*. Warszawa, Instytut Rynku Wewnętrznego i konsumentów.
- 20) Walesiak, M. 1996. *Metody analizy danych marketingowych*. Warszawa PWN.
- 21) Yule, G.U, Kendall M.G. 1966. *Wstęp do teorii statystyki*. Warszawa, PWN.
- 22) Zikmund, W.G. 1994. *Exploring Marketing Research*. Philadelphia, The Dryden Press.