

WIADOMOŚCI STATYSTYCZNE

CZASOPISMO GŁÓWNEGO URZĘDU STATYSTYCZNEGO
I POLSKIEGO TOWARZYSTWA STATYSTYCZNEGO

STUDIA METODOLOGICZNE

Małgorzata MACHOWSKA-SZEWCZYK,
Agnieszka SOMPOLSKA-RZECUŁA

Wielowymiarowa analiza zgodności wyników badania ankietowego

W badaniach ankietowych bardzo często wykorzystuje się analizy brzegowe za pomocą dwuwymiarowych tablic kontyngencji, umożliwiające wykrycie zależności pomiędzy parami cech. Analiza zgodności rozwiązuje jedno z trudniejszych zadań, a mianowicie pozwala na trafne rozpoznanie struktur odpowiedzi, wyrażonych częstościami współwystępowania poszczególnych kategorii cech, mierzonych na skali nominalnej. Celem artykułu jest prezentacja i zastosowanie metody analizy zgodności z wykorzystaniem macierzy Burta do badania ankietowego. Zaprezentowana metoda została zilustrowana przykładem badania studentów Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie pod względem odczuć zmian w funkcjonowaniu nowo powstałej uczelni.

Powołanie 1 stycznia 2009 r. nowej uczelni, utworzonej z połączenia Akademii Rolniczej i Politechniki Szczecińskiej, poza rozszerzeniem oferty dydaktycznej o nowe kierunki i specjalności studiów stwarza perspektywę nowych koncepcji technologicznych, obejmujących zarówno nauki techniczne, jak i przyrodnicze. Zachodniopomorski Uniwersytet Technologiczny w Szczecinie ma szansę skorzystania z efektu synergii, mającego początek w połączeniu dwóch znanych uczelni. W kompleksie kilkunastu obiektów dydaktycznych

studiuje ok. 17 tys. studentów na studiach stacjonarnych i niestacjonarnych, pod opieką ponad 1100 nauczycieli akademickich¹.

Źródło danych w badaniu stanowiła ankieta zamieszczona na stronach internetowych poszczególnych wydziałów (Pastusiak, 2010). Ankieta zawierała pytania dotyczące zarówno charakterystyki danej osoby, jak i jej opinii o połączeniu obu szkół wyższych. Każdy respondent mógł także ocenić odczucie nasilenia zmian w różnych dziedzinach działania uczelni (m.in.: nauczanie, obsługę w dziekanacie, rangę uczelni, ofertę oprogramowania, szansę znalezienia pracy przez absolwenta, atmosferę na uczelni).

OPIS METODY²

Analiza zgodności jest wyspecjalizowaną metodą eksploracji danych, która ułatwia prezentację związków pomiędzy cechami oraz pomiędzy jednostkami statystycznymi, przede wszystkim w postaci graficznej. Umożliwia ona analizę nie tylko danych ilościowych, ale także mierzonych na skalach nominalnych, porządkowych oraz nie stawia wymagań co do liczebności zbioru jednostek statystycznych. Celem tej metody jest wydobywanie wiedzy ze zbiorów danych poprzez analizę częstości współwystępowania poszczególnych wariantów obserwowanych cech. Liczebność wystąpień kategorii cech można zapisać w postaci złożonej macierzy znaczników, wielowymiarowej tablicy kontyngencji lub łączonej tablicy kontyngencji. Złożona macierz znaczników $\mathbf{Z}$ (tabl. 1) dla n jednostek statystycznych oraz Q cech składa się z podmacierzy odpowiadających poszczególnym cechom, czyli $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_Q]$. Elementami złożonej macierzy znaczników są wartości: 1 — gdy u danej jednostki zaobserwowano wyróżnioną kategorię cechy lub 0 — gdy nie zaobserwowano.

TABL. 1. PRZYKŁAD ZŁOŻONEJ MACIERZY ZNACZNIKÓW

Numer obiektu	Numer cechy														$z_{j\bullet}$
	1				2				...	Q					
	1_1	2_1	...	J_1	1_2	2_2	...	J_2	...	1_Q	2_Q	...	J_Q		
1	1	0	...	0	1	0	...	0	...	0	0	...	1	Q	
2	1	0	...	0	0	1	...	0	...	0	1	...	0	Q	
3	0	0	...	1	0	1	...	0	...	0	1	...	0	Q	
4	0	0	...	1	1	0	...	0	...	0	1	...	0	Q	
... ..	...	...	...	...	...	...	...	...	...	...	...	...	...	Q	
n	1	0	...	0	0	0	...	1	...	0	0	...	1	Q	
														nQ	

Źródło: Stanimir (2005).

¹ www.zut.edu.pl/index.php?id=3441.

² Opracowano na podstawie: Panek (2009), Stanimir (2005), Stanisław (2006).

W praktycznym zastosowaniu analizy zgodności wielu cech bardzo często zapisuje się dane w postaci macierzy Burta, która powstaje jako iloczyn $\mathbf{B} = \mathbf{Z}^T \mathbf{Z} = [b_{ij}]_{J \cdot J}$, $J = \sum_{q=1}^Q J_q$, gdzie J_q jest liczbą kategorii cechy X_q .

W wyniku tego działania otrzymuje się symetryczną macierz blokową, w której poza główną przekątną znajdują się tablice kontyngencji, odpowiadające dwóm różnym cechom, zawierające liczby jednostek, u których zaobserwowano ustalone kategorie tych dwóch cech. Na głównej przekątnej umieszczone są macierze diagonalne, w których wartości niezerowe oznaczają liczebność wystąpień danej kategorii cechy. Wizualną formę macierzy Burta dla trzech cech przedstawiono w tabl. 2.

TABL. 2. PRZYKŁAD MACIERZY BURTA DLA CECH: X, Y, Z

Cechy		X		Y			Z		
		X_1	X_2	Y_1	Y_2	Y_3	Z_1	Z_2	Z_3
X	X_1		0						
	X_2	0							
Y	Y_1				0	0			
	Y_2			0		0			
	Y_3			0	0				
Z	Z_1							0	0
	Z_2						0		0
	Z_3						0	0	

Źródło: opracowanie własne na podstawie Stanimir (2005).

W tablicy 2 różne odcienie szarości oznaczają komórki macierzy:

-  — Burta zawierające liczbę wystąpień ustalonych kategorii przyjętych cech,
-  — kontyngencji zawierające liczbę jednoczesnych wystąpień poszczególnych kategorii cechy X z odpowiednimi kategoriami cechy Y,
-  — kontyngencji zawierające liczbę jednoczesnych wystąpień poszczególnych kategorii cechy X z odpowiednimi kategoriami cechy Z,
-  — kontyngencji zawierające liczbę jednoczesnych wystąpień poszczególnych kategorii cechy Y z odpowiednimi kategoriami cechy Z.

Kolejnym krokiem w analizie zgodności jest przekształcenie macierzy Burta w macierz zaobserwowanej częstości względnych $\mathbf{P}$, którą otrzymuje się dzieląc liczebność w poszczególnych komórkach przez sumę liczebności wszystkich cech macierzy Burta, czyli nQ^2 . Na podstawie tej macierzy wyznaczono macierz profili wierszowych, uzyskiwaną poprzez podzielenie częstości względnych w każdym wierszu macierzy $\mathbf{P}$ przez sumę występujących częstości w danym wierszu. Analogicznie dzieląc każdą częstość względną w kolumnie macierzy $\mathbf{P}$ przez ich sumę częstości w danej kolumnie można wyznaczyć macierz profili

kolumnowych. Częstość brzegowa kolumn oraz wierszy w macierzach profili są średnimi profilami zwanymi odpowiednio centrum kolumnowym oraz wierszowym. Dowolny wiersz oraz kolumna macierzy profili może zostać przedstawiona jako punkt w przestrzeni wielowymiarowej, zaś centrum kolumnowe oraz wierszowe to punkt przecięcia głównych osi rzutowania. Aby przeprowadzić jednoczesną analizę profili wierszowych (kolumnowych) należy przekształcić macierz $\mathbf{P}$ w macierz $\mathbf{A}$ ważonych odchyłeń profili od centrum wierszowego (kolumnowego), zwaną macierzą różnic zestandaryzowanych, w postaci:

$$\mathbf{A} = [a_{ij}]$$

gdzie $a_{ij} = \frac{p_{ij} - p_{i\bullet} \cdot p_{\bullet j}}{\sqrt{p_{i\bullet} \cdot p_{\bullet j}}}$, p_{ij} — elementy macierzy $\mathbf{P}$, $p_{i\bullet} = \frac{Q \cdot b_{ii}}{n \cdot Q^2}$; $p_{\bullet j} = \frac{Q \cdot b_{jj}}{n \cdot Q^2}$,
 $j \in \{1, 2, \dots, J\}$.

Głównym celem analizy zgodności jest przedstawienie punktów reprezentujących cechy w przestrzeni o jak najmniejszym wymiarze, tak aby jak najdokładniej odwzorować odległości między punktami reprezentującymi kategorie danej cechy. Wymiary tej przestrzeni zwykle są zbyt duże, dlatego zachodzi konieczność zmniejszenia wymiaru pierwotnej przestrzeni cech. Do tego celu służy metoda rozkładu macierzy według wartości osobliwych. Zgodnie z podejściem Greenacre'a (1993) rzutowanie na przestrzeń o niskim wymiarze odbywa się przez rozkład macierzy $\mathbf{A}$ według wartości własnych, czyli:

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

gdzie:

$\mathbf{P}$ — macierz częstości zaobserwowanych $\mathbf{P} = \frac{1}{nQ^2} \mathbf{B}$,

$\mathbf{r} = [p_{i\bullet}]$ — wektor częstości brzegowych wierszy, przy czym $p_{i\bullet} = \frac{Q \cdot b_{ii}}{n \cdot Q^2}$,

$\mathbf{D}_r$ — macierz diagonalna częstości brzegowych wierszy
 $\mathbf{D}_r = \text{diag}(p_{1\bullet}, p_{2\bullet}, \dots, p_{j\bullet})$,

$\mathbf{U}$ — macierz wektorów własnych odpowiadających wartościom własnym λ_k macierzy $\mathbf{A}$,

$\mathbf{\Gamma}^2 = [\gamma_k^2]$ — macierz diagonalna zawierająca kwadraty wartości osobliwych γ_k^2 , które są równe wartościom własnym macierzy $\mathbf{A}$, czyli $\gamma_k^2 = \lambda_k$, $k \in \{1, 2, \dots, K\}$, $K = \sum_{q=1}^Q (J_q - 1)$, J_q — liczba kategorii cechy X_q .

W analizie zgodności używa się pojęcia bezwładności (inercji), które jest analogiczne do wariancji. Rozważa się tylko jedną wartość zwaną bezwładnością całkowitą lub inercją całkowitą, ponieważ całkowita bezwładność wierszy jest równa całkowitej bezwładności kolumn. Całkowita inercja macierzy określa stopień dyspersji profili względem odpowiednich profili przeciętnych. Inercja całkowita bliska zeru oznacza niewielką różnicę między profilami oraz profilem przeciętnym, a tym samym niewielką szansę wystąpienia istotnego powiązania między kategoriami poszczególnych cech. Całkowita inercja jest sumą wartości własnych, czyli $\lambda = \sum_{k=1}^K \lambda_k$. Wartości $\lambda_k, k \in \{1, \dots, K\}$ są nazywane inercjami głównymi. W praktyce stosuje się tzw. procent bezwładności, czyli λ_k / λ , oznaczający tę część bezwładności pierwotnej, która została wyjaśniona przez k -ty wymiar rzutowania.

Poprzez rozkład macierzy $\mathbf{A}$ według wartości własnych definiowany jest układ współrzędnych, w który rzutowane będą punkty odpowiadające kolejnym kolumnom (wierszom). Wyznacza się tu macierz $\mathbf{F}$ zawierającą współrzędne kategorii na wszystkich osiach powiązań następująco:

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

Wymiar przestrzeni rzeczywistych powiązań, otrzymany na podstawie analizy macierzy Burta, wynosi $K = \sum_{q=1}^Q (J_q - 1)$. W dalszej części badania dokonuje się rzutowania na przestrzeń o niższym wymiarze, tak aby strata informacji była jak najmniejsza. Wybór rozmiaru przestrzeni prezentacji wyników o możliwie najmniejszej stracie informacji opiera się na ocenie poziomu inercji całkowitej, która powinna być możliwie największa. Rozmiar przestrzeni rzutowania można określić na podstawie wartości miernika, który wskazuje na udział inercji ustalonego wymiaru w bezwładności całkowitej (procent skumulowany), wyznaczonego zgodnie ze wzorem:

$$\tau_{K^*} = \frac{\sum_{k=1}^{K^*} \lambda_k}{\sum_{k=1}^K \lambda_k} = \frac{\sum_{k=1}^{K^*} \lambda_k}{\lambda}$$

gdzie $K^* \in \{1, \dots, K\}$ oznacza wybrany wymiar rzutowania.

Przyjmuje się, że najlepszym rozwiązaniem jest wybranie takiego K^* , dla którego τ_{K^*} jest bliskie 1, a po zwiększeniu wymiaru przestrzeni wartość tego wskaźnika gwałtownie nie wzrasta.

W sytuacji gdy macierz Burta jest bardzo rozbudowana, rzeczywisty wymiar przestrzeni współwystępowania kategorii jest bardzo wysoki. Pierwsza wartość własna nie jest wtedy znacząco wyższa od pozostałych (jak w klasycznej analizie zgodności z dwiema cechami). Stosowanie kryterium stopnia wyjaśniania

bezwładności całkowitej przez inercje główne niewiele zmniejsza wymiar przestrzeni rzutowania w porównaniu z wymiarem przestrzeni rzeczywistych powiązań. Należy w takim przypadku sprawdzić, jaki wymiar rzutowania wskażą inne metody.

Jedną z metod określenia wymiaru przestrzeni rzutowania, opartego na analizie macierzy Burta lub złożonej macierzy znaczników, jest kryterium liczby cech zaproponowane przez Greenacre'a. Za optymalny wymiar K^* przestrzeni uznaje się największą liczbę k , dla której wartości własne spełniają warunek:

$$\lambda_k > \frac{1}{Q}, \quad \text{czyli} \quad K^* = \max \left\{ k \in \{1, \dots, K\} : \lambda_k > \frac{1}{Q} \right\}$$

Greenacre podaje także metodę „ulepszenia” współrzędnych, otrzymanych w wyniku przekształcenia macierzy Burta, następująco:

$$\tilde{\mathbf{F}} = \mathbf{F}^* \mathbf{\Gamma}^{-1} \tilde{\mathbf{\Lambda}}$$

gdzie:

$\tilde{\mathbf{\Lambda}}$ — macierz diagonalna pierwszych K^* wartości własnych zmodyfikowanych według wzoru: $\tilde{\lambda}_k = \left(\frac{Q}{Q-1} \right)^2 \left(\sqrt{\lambda_k} - \frac{1}{Q} \right)^2$,

$\mathbf{F}^*$ — macierz pierwszych K^* współrzędnych kategorii z macierzy $\mathbf{F}$.

Efekty przeprowadzonej analizy zgodności najczęściej przedstawia się graficznie i na tej podstawie ocenia położenie punktów wobec centrum rzutowania względem innych punktów określających kategorie tej samej cechy lub względem punktu opisującego kategorię innej cechy. Jeżeli punkt znajduje się blisko układu współrzędnych, to oznacza, że wyznaczony dla niego profil jest zbliżony do profilu przeciętnego. Punkty położone daleko od centrum rzutowania potwierdzają hipotezę o zależności cech. Jeżeli dwa punkty opisujące różne kategorie tej samej cechy leżą blisko siebie, to ich profile są podobne i możliwe jest połączenie liczebności dwóch kategorii w jedną bez istotnego wpływu na wyniki dalszych badań. Analizując położenie punktów opisujących kategorie dwóch różnych cech można stwierdzić, że istnieją powiązania między kategoriami, jeżeli punkty są położone blisko siebie. Jeżeli zaś punkty leżą po przeciwnych stronach centrum rzutowania, to oznacza, że kategorie nie występują wspólnie, a częstość występowania obu kategorii jednocześnie wynosi prawie zero.

Do graficznej prezentacji współwystępowania cech w przestrzeni o wymiarze większym niż trzy można zastosować wybrane metody klasyfikacji. Jako obiekty należy określić kategorie wszystkich analizowanych cech, jako cechy należy przyjąć zaś wartości współrzędnych rzutowania każdej kategorii. Metody klasyfikacji są również użyteczne, gdy liczba wszystkich wariantów cech jest duża

oraz rozrzut punktów na wykresie nie pozwala jednoznacznie wyodrębnić klas. W opracowaniu przeprowadzono klasyfikację wyników analizy zgodności, dla przykładu jedną z bardziej popularnych metod aglomeracyjnych — metodą Warda (Ward, 1963; Gordon, 1999; Ostasiewicz, 1998), która jest uznawana za najbardziej efektywną. Metoda Warda jest prawie o 40% bardziej efektywna³ w porównaniu z drugą, co do efektywności — metodą najdalszego sąsiedztwa. Można również przeanalizować inne metody grupowania czy skalowanie wielowymiarowe i wskazać kryteria wyboru najlepszego sposobu prezentacji oraz porównać otrzymane wyniki, jednak to zadanie wykraczałoby poza ramy objętościowe i zagadnieniowe tego artykułu. Zagadnienie to może zostać podjęte w kolejnych opracowaniach.

MATERIAŁ BADAWCZY

Badanie przeprowadzono na podstawie kwestionariuszy ankiety wypełnianych *on-line* przez 441 studentów od 1 marca 2010 r. do końca czerwca 2010 r. Badanie zostało przeprowadzone techniką CAWI (Mazurek-Łopacińska, 2005). Na podstawie kwestionariusza wyłoniono potencjalne cechy przedstawione w zestawieniu. Dla każdej cechy jakościowej, której liczba możliwych wariantów do wyboru była większa niż 1 (*Pow_Stud*, *Pow_Pol*, *Odcz_Zm*), wprowadzono odpowiednią liczbę nowych cech odzwierciedlających poszczególne kategorie.

ZESTAWIENIE POTENCJALNYCH CECH PRZYJĘTYCH DO BADANIA

Nazwa cechy	Opis cechy i jej kategorie
<i>St</i>	miejsce studiów: PS — dawna Politechnika Szczecińska, AR — dawna Akademia Rolnicza w Szczecinie
<i>Rodz</i>	poziom i forma studiów: S1 — stacjonarne I stopnia, N1 — niestacjonarne I stopnia, MS_S2 — stacjonarne jednolite magisterskie lub stacjonarne II stopnia, N2 — niestacjonarne II stopnia, D — doktoranckie
<i>Wydz</i>	wydział: BHZ — Biotechnologii i Hodowli Zwierząt, BA — Budownictwa i Architektury, Ek — Ekonomiczny, E — Elektryczny, I — Informatyki, IMM — Inżynierii Mechanicznej i Mechatroniki, KSR — Kształtowania Środowiska i Rolnictwa, NZR — Nauk o Żywności i Rybactwa, TICH — Technologii i Inżynierii Chemicznej, TM — Techniki Morskiej
<i>Sat</i>	czy studia są satysfakcjonujące? T — tak, NT — najczęściej tak, NN — najczęściej nie, N — nie
<i>Pres</i>	która uczelnia cieszyła się większym prestiżem przed połączeniem? PS — Politechnika Szczecińska, AR — Akademia Rolnicza w Szczecinie, N — nie mam zdania
<i>Pow_Stud</i>	powody podjęcia studiów w Politechnice Szczecińskiej/Akademii Rolniczej w Szczecinie (można wybrać co najmniej jedną odpowiedź)
<i>NS</i>	zainteresowanie naukami ścisłymi: 1 — tak, 0 — nie
<i>Prac</i>	większe szanse znalezienia pracy jako absolwent tej uczelni: 1 — tak, 0 — nie
<i>Prest</i>	studiowanie na PS/AR cieszy się prestiżem: 1 — tak, 0 — nie
<i>Rodz</i>	namowa rodziców: 1 — tak, 0 — nie

³ Przez efektywność metod taksonomicznych należy rozumieć zdolność prawidłowego rozpoznawania rzeczywistej struktury obiektów w wielowymiarowych przestrzeniach cech, którą bada się na przykładach empirycznych lub na podstawie zbiorów sztucznie wygenerowanych (Malina, 2004).

ZESTAWIENIE POTENCJALNYCH CECH PRZYJĘTYCH DO BADANIA (dok.)

Nazwa cechy	Opis cechy i jej kategorie
<i>Zn</i>	namowa znajomych: 1 — tak, 0 — nie
<i>Pow_Pol</i>	powody połączenia Politechniki Szczecińskiej i Akademii Rolniczej w Szczecinie (można zaznaczyć co najmniej jedną odpowiedź)
<i>PodPrest</i>	podniesienie prestiżu uczelni: 1 — tak, 0 — nie
<i>Dot</i>	zwiększenie dotacji dla uczelni: 1 — tak, 0 — nie
<i>Kadr_fin</i>	problemy kadrowe lub finansowe: 1 — tak, 0 — nie
<i>Nab</i>	problemy z naborem studentów: 1 — tak, 0 — nie
<i>In</i>	inne: 1 — tak, 0 — nie
<i>Odcz_Zm</i>	odczucie nasilenia zmian w dziedzinie (można zaznaczyć co najmniej jedną odpowiedź)
<i>Zm_N</i>	nauczania: poz — odczucia pozytywne, brak — nie odczuwa zmian, neg — odczucia negatywne
<i>Zm_D</i>	obsługi w dziekanacie: poz — odczucia pozytywne, brak — nie odczuwa zmian, neg — odczucia negatywne
<i>Zm_Adm</i>	administracji centralnej: poz — odczucia pozytywne, brak — nie odczuwa zmian, neg — odczucia negatywne
<i>Zm_Ucz</i>	rangi uczelni: poz — odczucia pozytywne, brak — nie odczuwa zmian, neg — odczucia negatywne
<i>Zm_Opr</i>	oferty oprogramowania: poz — odczucia pozytywne, brak — nie odczuwa zmian, neg — odczucia negatywne
<i>Zm_Pr</i>	znalezienia pracy przez absolwenta: poz — odczucia pozytywne, brak — nie odczuwa zmian, neg — odczucia negatywne
<i>Zm_Atm</i>	atmosfery na uczelni: poz — odczucia pozytywne, brak — nie odczuwa zmian, neg — odczucia negatywne
<i>Plec</i>	pleć: K — kobieta, M — mężczyzna
<i>Urodz</i>	miejsce urodzenia: W — wieś, M — miasto inne niż Szczecin, Sz — Szczecin
<i>W</i>	wiek: 19—23, 24—28, pow_28
<i>Mzam</i>	miejsce zamieszkania podczas studiów: A — akademik, St — stancja, DR — dom rodzinny, M — własne mieszkanie
<i>Prac</i>	czy pracujesz? T — tak, N — nie
<i>Kw</i>	miesięczna kwota do dyspozycji w zł: do_800, 801—1500, 1501—2500, pow_2500
<i>Pkt</i>	czy w ostatnim semestrze uzyskałeś/aś minimum punktowe do przejścia na następny semestr? TbP — tak, bez problemów, TzT — tak, z trudem, N — nie, powtarzam semestr
<i>Op</i>	według twojej opinii jesteś studentem: Bdb — bardzo dobrym, Db — dobrym, Sr — średnim, Sl — słabym
<i>Cz_roz</i>	jak często w tygodniu poświęcasz wieczory na rozrywkę? 1 — Przecięcie raz w tygodniu, 2—3 razy w tygodniu, Pow_3—4 i więcej razy w tygodniu.

Ź r ó ł o: opracowanie własne na podstawie kwestionariusza ankiety (Pastusiak, 2010).

Spośród potencjalnych cech przedstawionych w zestawieniu wyodrębniono zależne, dotyczące odczuć zmian studentów w poszczególnych dziedzinach funkcjonowania uczelni:

- *Zm_N* — nauczanie,
- *Zm_D* — obsługa w dziekanacie,
- *Zm_Adm* — administracja centralna,
- *Zm_Ucz* — ranga uczelni,
- *Zm_Opr* — oferta oprogramowania,
- *Zm_Pr* — szansa znalezienia pracy przez absolwenta,
- *Zm_Atm* — atmosfera na uczelni.

Przed zastosowaniem wielowymiarowej analizy zgodności zbadano, czy pomiędzy cechami określającymi odczucia zmian w poszczególnych dziedzinach

działalności uczelni istnieje zależność. Okazało się, że można zaobserwować następujące powiązania:

- w nauczaniu — z miejscem studiów, wydziałem, satysfakcją ze studiów, prestiżem wcześniej istniejącej uczelni, problemami z naborem studentów;
- w funkcjonowaniu dziekanatu — z miejscem studiów, poziomem i formą studiów, wydziałem, prestiżem wcześniej istniejącej uczelni, szansą znalezienia pracy;
- w administracji centralnej — z miejscem studiów, poziomem i formą studiów, wydziałem, prestiżem wcześniej istniejącej uczelni, płcią, czasem poświęcanym na rozrywkę;
- rangi uczelni — z miejscem studiów, wydziałem, prestiżem wcześniej istniejącej uczelni, zainteresowaniem naukami ścisłymi, szansą znalezienia pracy, prestiżem studiowania na wcześniej istniejącej uczelni, podniesieniem prestiżu nowej uczelni, wiekiem studenta;
- w ofercie oprogramowania — z miejscem studiów, wydziałem, namową znajomych, miesięczną kwotą do dyspozycji;
- w szansach znalezienia pracy — z miejscem studiów, wydziałem, satysfakcją ze studiów, prestiżem wcześniej istniejącej uczelni, zainteresowaniem naukami ścisłymi, prestiżem studiowania na wcześniej istniejącej uczelni, namową znajomych, podniesieniem prestiżu nowej uczelni, problemami z naborem studentów, innymi powodami połączenia uczelni, wiekiem;
- atmosfery na uczelni — z miejscem studiów, wydziałem, szansą znalezienia pracy, podniesieniem prestiżu nowej uczelni, problemami z naborem studentów.

W przypadku każdej z siedmiu cech zależnych przeprowadzono wielowymiarową analizę zgodności, aby wyłonić te kategorie cech niezależnych, które współwystępują z kategoriami odpowiednich cech zależnych. Dla przykładu została przedstawiona interpretacja jednego z siedmiu analizowanych modeli, w którym obserwowano powiązania między strukturami odpowiedzi, wyrażonymi częstościami występowania kategorii cechy zależnej, opisującej odczucia zmian w nauczaniu (Zm_N) z kategoriami cech niezależnych.

WYNIKI BADANIA

Do kwantyfikacji opinii studentów po utworzeniu Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie za pomocą metody wielowymiarowej analizy zgodności wykorzystano program komputerowy *Statistica* oraz *Excel*. W przypadku badania zależności pomiędzy kategoriami pierwszej cechy zależnej (Zm_N) a kategoriami cech z nią powiązanych otrzymano macierz Burtę o wymiarach 23×23 . Wymiar rzeczywistej przestrzeni współwystępowania odpowiedzi na pytania wynosił 17.

W następnym kroku sprawdzono, jak wartości własne przestrzeni o niższym wymiarze wyjaśniają całkowitą bezwładność. Wyniki zamieszczono w tabl. 3, w której zawarto wartości własne λ_k , wartości osobliwe γ_k , udział inercji

głównych w całkowitej inercji (procent bezwładności λ_k / λ) oraz udział wartości własnych z wymiaru K w inercji całkowitej (procent skumulowany τ_K).

TABL. 3. WARTOŚCI OSOBLIWE I WARTOŚCI WŁASNE ORAZ STOPIEŃ WYJAŚNIENIA CAŁKOWITEJ BEZWŁADNOŚCI

Liczba wymiarów K	Wartości		Procent	
	własne λ_k	osobliwe γ_k	bezwładności λ_k / λ	skumulowany τ_K
1	0,4091	0,6396	14,4384	14,4384
2	0,2398	0,4897	8,4637	22,9020
3	0,2090	0,4571	7,3751	30,2771
4	0,1870	0,4324	6,5989	36,8760
5	0,1818	0,4264	6,4165	43,2925
6	0,1772	0,4210	6,2541	49,5467
7	0,1703	0,4127	6,0106	55,5573
8	0,1676	0,4093	5,9139	61,4712
9	0,1663	0,4078	5,8704	67,3415
10	0,1573	0,3966	5,5514	72,8930
11	0,1504	0,3878	5,3092	78,2022
12	0,1426	0,3776	5,0328	83,2350
13	0,1332	0,3650	4,7009	87,9358
14	0,1273	0,3568	4,4942	92,4300
15	0,1191	0,3451	4,2042	96,6342
16	0,0788	0,2807	2,7812	99,4153
17	0,0166	0,1287	0,5847	100,0000
	2,8333			

Źródło: opracowanie własne.

Według kryterium Greenacre'a jako najlepszy wybierany jest ten wymiar rzutowania kategorii cech, w którym wartości własne spełniają warunek: $\lambda_k > \frac{1}{Q}$. W analizowanym przypadku wartość ta wynosi 0,1667, dla $Q=6$.

Można stwierdzić, że jest to bezwładność przestrzeni R^8 , a łączna bezwładność w analizowanym przypadku wynosi 2,8333. Dokonano też modyfikacji wartości własnych według kryterium Greenacre'a. W tabl. 4 zawarto wartości własne, osobliwe i stopień wyjaśnienia całkowitej bezwładności po modyfikacji.

TABL. 4. ZMODYFIKOWANE WARTOŚCI OSOBLIWE I WARTOŚCI WŁASNE ORAZ STOPIEŃ WYJAŚNIENIA CAŁKOWITEJ BEZWŁADNOŚCI

Liczba wymiarów K	Wartości		Procent	
	własne $\tilde{\lambda}_k$	osobliwe $\tilde{\gamma}_k$	bezwładności $\tilde{\lambda}_k / \tilde{\lambda}$	skumulowany $\tilde{\tau}_K$
1	0,3459	0,1196	28,0331	28,0331
2	0,1737	0,0302	14,0799	42,1130
3	0,1441	0,0208	11,6769	53,7899
4	0,1234	0,0152	10,0032	63,7931
5	0,1186	0,0141	9,6154	73,4085
6	0,1144	0,0131	9,2720	82,6805
7	0,1081	0,0117	8,7607	91,4412
8	0,1056	0,0112	8,5588	100,0000
	1,2338			

Źródło: jak przy tabl. 3.

Dodatkowo sporządzono wykres wartości własnych (wykr. 1).

Jednym ze sposobów określenia liczby wartości własnych, które wskazują współrzędne istotne dla rzutowania na przestrzeń o niskim wymiarze, jest kryterium „łokcia”. Na wykresie wszystkich niezerowych wartości własnych w porządku nierosnącym poszukuje się miejsca, w którym następuje łagodny spadek wartości własnych (wskazuje się numer wartości własnej, dla której zaobserwowano „zagięcie”) (Stanimir, 2005). Na podstawie wykr. 1 stwierdzono, że tzw. „łokieć” występuje w przypadku $k = 4$. Zatem analiza powiązań między kategoriami cech będzie odbywać się w przestrzeni czterowymiarowej, która wyjaśnia prawie 64% łącznej bezwładności.

Powiązania pomiędzy strukturami odpowiedzi, wyrażonymi częstościami współwystępowania poszczególnych kategorii cech, zostały pokazane na dendrogramie, który otrzymano za pomocą metody Warda (wykr. 2).

Analizując dendrogram otrzymano pięć grup kategorii cech. W pierwszej grupie znalazło się najwięcej kategorii cech. Osoby, które nie odczuwały zmian w nauczaniu po połączeniu uczelni deklarowały satysfakcję ze studiów, nie widziały problemów z naborem na studia, były studentami dawnej Politechniki Szczecińskiej (Wydziału Informatyki oraz Inżynierii Mechanicznej i Mechatroniki) oraz studentami Akademii Rolniczej. Twierdziły też, że studia dają im satysfakcję, a przed połączeniem Politechnika Szczecińska cieszyła się większym prestiżem w porównaniu z Akademią Rolniczą. Studenci odczuwający pozytywne zmiany w nauczaniu po połączeniu uczelni byli studentami Wydziału Nauk o Zdrowiu i Rybactwa oraz Wydziału Ekonomicznego Akademii Rolniczej. Natomiast osoby odczuwające negatywne zmiany w nauczaniu nie czuły satysfakcji ze studiowania, przewidywały problemy z naborem i były studentami Wydziału Technologii i Inżynierii Chemicznej. Pozostałe kategorie cech nie wykazały powiązań częstościowych między kategoriami cechy zależnej. Można zauważyć, że wśród nich wystąpiły następujące grupy powiązań:

- studenci Wydziału Biotechnologii i Hodowli Zwierząt nie miały zdania na temat prestiżu uczelni przed połączeniem,
- studenci Wydziałów Elektrycznego, Kształtowania Środowiska i Rolnictwa oraz Budownictwa i Architektury najczęściej nie odczuwały satysfakcji ze studiowania i sądziły, że Akademia Rolnicza cieszyła się większym prestiżem w porównaniu z Politechniką Szczecińską.

Wyniki poszczególnych etapów analizy zgodności na podstawie cech wykazujących powiązania z cechami zależnymi przedstawiono w tabl. 5.

W większości przypadków ostatecznym wymiarem przestrzeni obserwacji struktur odpowiedzi wyrażonych częstościami współwystępowania kategorii cech było $k = 4$. Jedynie powiązania kategorii cech Zm_D i Zm_Adm z kategoriami cech niezależnych można rozpatrywać w przestrzeni trójwymiarowej. Na wykresie 3 przedstawiono wyniki analizy zgodności w przestrzeni trójwymiarowej, kolorem czerwonym oznaczono kategorie cechy zależnej.

TABL. 5. WYNIKI ANALIZY ZGODNOŚCI

Cecha zależna	Liczba cech niezależnych Q	Wymiar macierzy Burta	Kryterium Greenacre'a	Wymiar przestrzeni współwystępowania kategorii cech			Stopień wyjaśnienia inercji w %
				rzeczywisty	po analizie wartości własnych	na podstawie kryterium łockia	
<i>Zm_N</i>	6	23x23	0,1667	17	8	4	63,8
<i>Zm_D</i>	6	24x24	0,1667	18	9	3	48,1
<i>Zm_Adm</i>	7	27x27	0,1429	20	10	3	53,8
<i>Zm_Ucz</i>	9	28x28	0,1111	19	8	4	63,8
<i>Zm_Opr</i>	5	20x20	0,2000	15	7	4	61,8
<i>Zm_Pr</i>	12	36x36	0,0833	24	11	4	43,8
<i>Zm_Atm</i>	6	20x20	0,1667	14	6	4	78,7

Ź r ó d ł o: jak przy tabl. 3.

W celu wskazania powiązań między kategoriami cech zastosowano także metodę Warda i otrzymano cztery klasy kategorii. W tabl. 6 przedstawiono trzy klasy zawierające kategorie cechy zależnej *Zm_Adm* współwystępującej z kategoriami cech niezależnych, natomiast klasa czwarta zawiera te kategorie cech niezależnych, które nie wykazują żadnych powiązań z kategoriami cechy zależnej *Zm_Adm*. Należą do nich: Wydz: BA, Wydz: IMM, Rodz: MS_S2, Wydz: NZR, Wydz: KSR, Rodz: N1.

TABL. 6. KLASY POWIĄZAŃ MIĘDZY STRUKTURAMI ODPOWIEDZI, WYRAŻONYMI CZĘSTOŚCIAMI WYSTĘPOWANIA, KATEGORIAMI CECHY ZALEŻNEJ *ZM_ADM* I CECAMI NIEZALEŻNYMI

Kategorie cechy zależnej <i>Zm_Adm</i>	Kategorie cech niezależnych
Pozytywne odczucie zmian w administracji	Wydz: TICH, Cz_roz: 1, Plec: K, Pres: PS, ST:PS
Negatywne odczucie zmian w administracji	Wydz: BHZ, Rodz: D
Brak odczuć zmian w administracji	Wydz: E, Rodz: N2, Cz_Roz: Pow_3, Wydz: I, Plec:M, Rodz:S1, Pres:AR, Wydz:EK, ST:AR

Ź r ó d ł o: jak przy tabl. 3.

Podsumowanie

W artykule przedstawiono metodę analizy zgodności z wykorzystaniem macierzy Burta oraz przykładowe zastosowanie w bardzo złożonym badaniu ankietowym. Metoda zgodności jest wyspecjalizowaną metodą eksploracji danych, która umożliwia analizę nie tylko danych ilościowych, ale także mierzonych na skalach nominalnych, porządkowych oraz nie stawia żadnych wymagań co do liczebności zbioru obiektów. Zaletą tej metody jest redukcja wymiaru obserwacji struktur odpowiedzi wyrażonych częstościami występowania (w tym współwystępowania) poszczególnych kategorii badanych cech bez istotnej straty posiadanego zasobu informacyjnego. Umożliwia to wychwycenie związków pomiędzy tymi strukturami oraz znacznie ułatwia prezentację takich struktur w postaci graficznej.

W opracowaniu przedstawiono zastosowanie wielowymiarowej analizy zgodności do badania powiązań między kategoriami cech, charakteryzujących odczucia zmian studentów w funkcjonowaniu nowo powstałego Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie. Źródło informacji stanowiła ankieta zamieszczona na stronach internetowych poszczególnych wydziałów, wypełniona przez 441 studentów. Wyodrębniono siedem cech zależnych, które przedstawiały odczucia w funkcjonowaniu uczelni w wielu aspektach: nauczaniu, obsłudze w dziekanacie i administracji centralnej uczelni, randze uczelni, ofercie oprogramowania, szansie znalezienia pracy przez absolwenta oraz atmosferze na uczelni. Powiązania częstościowe między kategoriami cech badano w przestrzeni trój- i czterowymiarowej.

Pozytywne zmiany w nauczaniu po połączeniu uczelni odczuwali studenci dawnej Akademii Rolniczej w Szczecinie (wydziałów Nauk o Zdrowiu i Rybac-

stwa oraz Ekonomicznego). Negatywne zmiany w nauczaniu odczuwali jedynie studenci Wydziału Technologii i Inżynierii Chemicznej byłej Politechniki Szczecińskiej, którzy nie czuli satysfakcji ze studiowania oraz dostrzegali problemy z naborem. Osoby, które nie odczuwały zmian w nauczaniu po połączeniu uczelni deklarowały satysfakcję ze studiów, nie widziały problemów z naborem na studia, były studentami dawnej Politechniki Szczecińskiej (Wydziału Informatyki oraz Inżynierii Mechanicznej i Mechatroniki) oraz studentami Akademii Rolniczej w Szczecinie. Osobom tym studia przynosiły najczęściej satysfakcję oraz uznały one, że przed połączeniem Politechnika Szczecińska cieszyła się większym prestiżem w porównaniu z Akademią Rolniczą.

Wielowymiarowa analiza zgodności umożliwiła graficzną prezentację struktur odpowiedzi wyrażonych częstościami występowania poszczególnych kategorii cech, którą wykorzystano w prezentacji powiązań w przestrzeni trójwymiarowej, co świadczy o przydatności wykorzystania tej metody w złożonym badaniu ankietowym.

dr Małgorzata Machowska-Szewczyk, dr Agnieszka Sompolska-Rzechuła — *Zachodniopomorski Uniwersytet Technologiczny w Szczecinie*

LITERATURA

- Gordon A. D. (1999), *Classification*, Chapman & Hall/CRC, Boca Raton
- Greenacre M. (1993), *Correspondence analysis in practice*, Academic Press, London
- Mazurek-Lopacińska K. (2005), *Badania marketingowe. Teoria i praktyka*, Wydawnictwo Naukowe PWN, Warszawa
- Ostasiewicz W. (red.) (1998), *Statystyczne metody analizy danych*, Akademia Ekonomiczna we Wrocławiu, Wrocław
- Panek T. (2009), *Statystyczne metody wielowymiarowej analizy porównawczej*, Wydawnictwo SGH w Warszawie, Warszawa
- Pastusiak M. (2010), *Analiza statystyczna połączenia Politechniki Szczecińskiej oraz Akademii Rolniczej w Szczecinie*, Praca inżynierska, ZUT Szczecin
- Stanimir A. (2005), *Analiza korespondencji jako narzędzie do badania zjawisk ekonomicznych*, Akademia Ekonomiczna we Wrocławiu, Wrocław
- Stanisz A. (2006), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, tom 3: Analizy wielowymiarowe, StatSoft, Kraków
- Ward J. H. (1963), *Hierarchical grouping to optimize an objective function*, „Journal of the American Statistical Association”, No. 58

SUMMARY

The article presents a multidimensional analysis of compliance, using Burt matrix, which allows to detect relationships between categories of the dependent and independent characteristics. The surveys use often boundary analysis using

two-dimensional contingency tables to detect relationships between pairs of characteristics. The analysis of compatibility addresses one of the most difficult tasks — allows to diagnose the accurate structure of responses, expressed by occurrence frequency of characteristics categories measured on a nominal scale. The presented method is illustrated on the survey example of students at the Zachodniopomorski University of Technology in Szczecin on the perception of changes in the functioning of the university (the merged Technical University and the Agricultural Academy in Szczecin). Source of data in the study was a survey, published on the websites of each university departments. It included questions about both the characteristics of a person, as well as its assessment of the institution, in areas such as: teaching, support for the Dean's Office and central university administration, universities rank, software offer, the chances of finding work for graduates, and the atmosphere at the university. The value of this paper is to demonstrate the usefulness of the analysis of compliance in a complex survey.

РЕЗЮМЕ

В статье представляется многомерный анализ соответствия, использующий матрицу Барта, которая позволяет обнаружить связи между категориями зависимого признака и категориями независимых признаков.

В анкетных обследованиях часто употребляется предельный анализ с использованием двухмерных таблиц контингенции, позволяющих обнаружить зависимости между парами признаков. Анализ соответствия решает одну из труднейших задач — позволяет точно определить структуру ответов, выраженных частотой одновременного наличия категории признаков, измеряемых на номинальной шкале.

Представленный метод был изображен на примере обследования мнений студентов Западнопоморского технологического университета в Щецине касающихся восприятия изменений в области функционирования школы образованной в результате соединения Щецинского политехнического института и Университета сельского хозяйства в Щецине. Источником данных в обследовании была анкета помещена на веб-сайтах отдельных факультетов. В анкете находились вопросы касающиеся как характеристики данного студента, так и его мнения по функционированию школы в области обучения, работы деканата и центральной администрации школы, значения школы, программного обеспечения, шансов найти себе работу выпускниками, или атмосферы в школе. Достоинством разработки является представление пригодности анализа соответствия в сложном анкетном обследовании.