

Iwona Staniec
Politechnika Łódzka

ANALIZA KORESPONDENCJI W OCENIE WIARYGODNOŚCI KREDYTOBIORCY

1. Analiza korespondencji

Celem referatu jest ocena wiarygodności kredytowej kredytobiorcy za pomocą analizy korespondencji na podstawie danych dotyczących kredytobiorców ubiegających się o kredyt konsumpcyjny w firmie Gwarant.

Metoda analizy korespondencji rozwinęła się w latach 60. i 70. we Francji. Należy ona do metod skalowania wielowymiarowego. Analiza korespondencji to eksploracyjna technika analizy tablic kontyngencji, która zmierza do odtworzenia odległości pomiędzy punktami reprezentującymi wiersze i/lub kolumny w przestrzeni o mniejszej liczbie wymiarów przy jednoczesnym zachowaniu jak największej liczby pierwotnych informacji¹. W analizie korespondencji stosuje się zasadę redukcji wielowymiarowej przestrzeni niemetrycznej do przestrzeni metrycznej o mniejszej liczbie wymiarów tak, aby można było umieścić obiekty na płaszczyźnie. Zatem celem metody jest wyznaczenie mapy percepcji, czyli wykresu rozproszenia przedstawiającego rozmieszczenie obiektów w dwuwymiarowej przestrzeni. Obiektami na wykresie są punkty reprezentujące wiersze i/lub kolumny tablicy kontyngencji. Otrzymane mapy percepcji pozwalają na analizę [Greenacre 1993, s. 54]:

- podobieństw i różnic wariantów wiersza ze względu na daną kategorię kolumny,
- podobieństw i różnic kategorii kolumn ze względu na dany wariant wiersza,
- zależności między wierszami i kolumnami.

Za wadę analizy korespondencji uważa się trudności w interpretacji odległości pomiędzy kategoriami przedstawianymi na mapie percepcji, gdyż interpretacja ta zależy od użytej metody normalizacji.

¹ Definicję przytoczono za [Gatnar, Walesiak 2004, s. 284].

Do zalet analizy korespondencji możemy zaliczyć:

- możliwość jej wykorzystania w badaniu danych jakościowych,
- dążenie do identyfikacji nieznanymi cech obiektów poprzez takie ich uporządkowanie, aby obiekty o cechach podobnych znalazły się blisko siebie na mapie percepcji,
- brak założeń, które musi spełniać np. analiza czynnikowa [Górnik 2001, s. 134].

W analizie korespondencji analiza wykresu połączona z analizą miar charakteryzujących zależności między zmiennymi oraz miar jakości dopasowania prowadzi do nadania interpretacji merytorycznej uzyskanym wymiarom oraz do wnioskowania o związkach zachodzących między zmiennymi.

Punktem wyjścia do analizy korespondencji jest wyznaczenie macierzy:

- korespondencji – tablicy kontyngencji, w której liczebności są przekształcone w częstości względne,
- profili wierszowych – częstość względną w macierzy korespondencji dzielimy przez sumę wszystkich częstości w odpowiadającym wierszu,
- profili kolumnowych – częstość względną w macierzy korespondencji dzielimy przez sumę wszystkich częstości w odpowiadającej kolumnie.

Po wyznaczeniu macierzy profili wylicza się odległości między wierszami/kolumnami, stosując do tego metrykę chi-kwadrat. Metryka ta jest ważoną metryką euklidesową, w której wagami są odwrotności odpowiednich profili. Skutkiem tego kategorie z mniejszą liczbą elementów wywierają większy wpływ na odległości między punktami niż kategorie z większą liczbą obserwacji.

W analizie korespondencji inercja (bezwładność) jest miarą wariancji, czyli pozwala zmierzyć, jak bardzo poszczególne profile wierszowe/kolumnowe różnią się od odpowiedniego profilu przeciętnego. Jeżeli różnica między profilami a profilem przeciętnym jest niewielka, to wartość inercji jest bliska zeru.

2. Dane wykorzystane w eksperymentach numerycznych

W przypadku analizy wiarygodności kredytobiorców interesujące są problemy zależności między kategoriami, takimi jak: wiarygodność kredytobiorcy a jego płeć, wysokość wynagrodzenia, wiek, liczba osób w rodzinie klienta i liczba rat, na które był rozłożony kredyt. Do analiz empirycznych dane zaczerpnięto z bazy firmy Gwarant. Dane te dotyczyły 4999 klientów, w tym tylko 477 (10,5%) było niewiarygodnych. Cechom charakteryzującym klientów przypisano następujące kategorie:

- wiarygodność kredytobiorcy: 1 – wiarygodny (tzn. spłacił kredyt), 0 – niewiarygodny, nie spłacił kredytu w całości;
- płeć: M – mężczyzna, K – kobieta;
- wysokość wynagrodzenia – dane pogrupowano w szereg rozdzielczy (5 klas),
- wiek – dane pogrupowano w szereg rozdzielczy (8 klas),
- liczba osób w rodzinie klienta – wśród danych występowały rodziny 1-, 2-, 3-, 4- i 5-osobowe;
- liczba rat – dane pogrupowano w szereg rozdzielczy (5 klas).

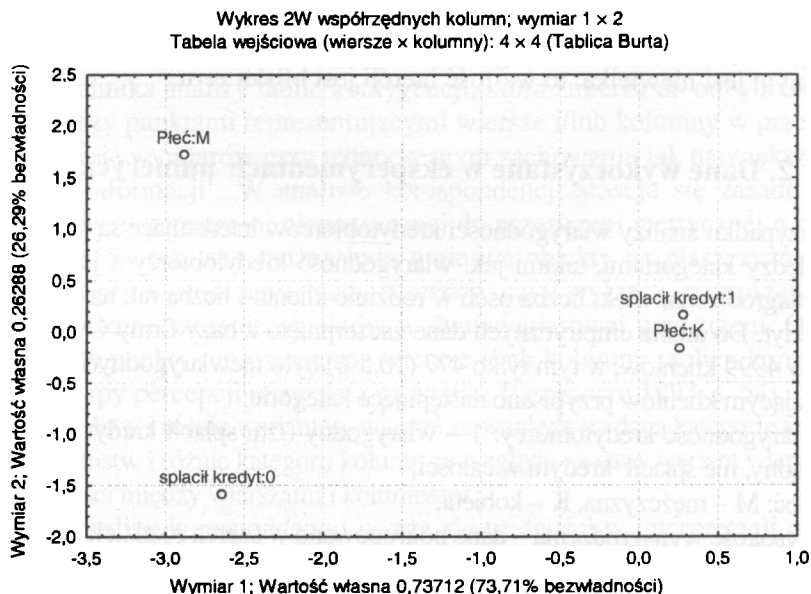
Analizując powiązania między większą niż dwie liczbą zmiennych, należy sprawdzić, czy pary cech wskazują zależność. W tym celu dla każdej pary cech wyznaczono wartości statystyki χ^2 oraz średniokwadratową wielodzielczość ϕ^2 i przedstawiono je w tabeli.

Tabela 1. Wartość statystyki χ^2 i średniokwadratowa wielodzielczość ϕ^2

χ^2 / ϕ^2	L_rat	Wiek	Zarobki	Płeć	Liczba osób	Wiarygodność
L_rat		49,4	21,57	48	18,8	37,37
Wiek	0,0099		31,51	11,18	18,49	10,12
Zarobki	0,0043	0,0063		247	111	3,94
Płeć	0,0096	0,0022	0,0494		10,1	1124
Liczba osób	0,0038	0,0037	0,0222	0,0020		6,66
Wiarygodność	0,0075	0,0020	0,0008	0,2248	0,0013	

Źródło: opracowanie własne.

Wartości średniokwadratowej wielodzielczości dla kombinacji cech pokazują, że cechy mogą być niezależne. Zależność występuje jedynie między płcią i zarobkami oraz płcią a wiarygodnością. Należy jednak pamiętać, że statystyka chi-kwadrat jest bardzo czuła na niewielkie liczebności. Mężczyzn w badanej próbie było tylko 406, w tym 229 uznano za niewiarygodnych. Kobiet zaś było 4593, z czego tylko 248 uznano za niewiarygodne.



Rys. 1. Mapa percepcji dla płci i wiarygodności kredytobiorcy

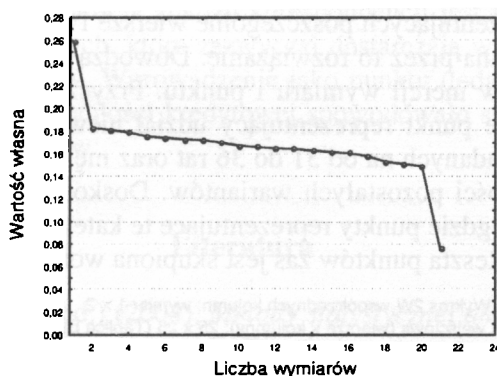
Źródło: opracowanie własne z użyciem pakietu Statistica 6.0.

Bezwładność w przypadku wiarygodności i płci wynosi 73,71%. Analizując przedstawioną mapę percepcji, można zauważyć, że kobiety są powiązane bardzo silnie z klientami wiarygodnymi.

3. Analizy numeryczne²

Analiza korespondencji pozwala wykryć zarówno wystąpienie, jak i brak powiązań między kategoriami cech. Zatem można bez utraty wiarygodności uzyskanych wyników pozostawić wszystkie cechy w gronie analizowanych zmiennych. Brak powiązań kategorii cech będzie widoczny po sporządzeniu mapy percepcji, ponieważ punkty obrazujące jedną cechę nie będą korespondować z żadną kategorią określającą drugą cechę.

W celu określenia wymiaru przestrzeni dla zaprezentowanych cech wykorzystano test osypiska. Na rysunku 2 przedstawiono krzywą łączącą kolejne wartości własne. Po analizie przebiegu krzywej można stwierdzić, że wyrażany uskok występuje dla $k = 2$.



Rys. 2. Test osypiska

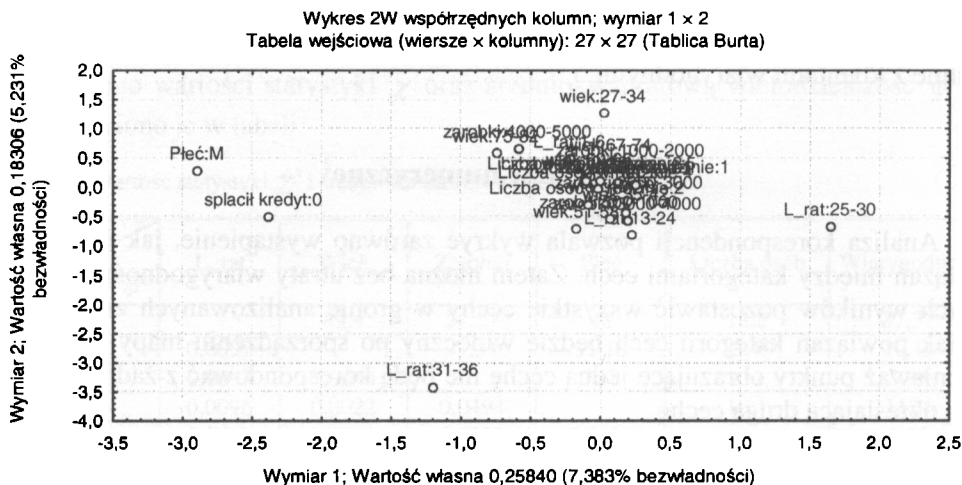
Źródło: opracowanie własne z użyciem pakietu Statistica 6.0.

Analiza przedstawionego wykresu pozwala stwierdzić, że mapa percepcji powinna być dwuwymiarowa, gdyż dla dwójki zbocze krzywej jest najbardziej strome.

Na rysunku 3 zaprezentowano położenie punktów obrazujących kategorie analizowanych cech w dwuwymiarowym układzie współrzędnych. Biorąc pod uwagę odległość punktów od centrum rzutowania, można uznać, że punkty obrazujące wszystkie kategorie cech położone są bardzo blisko siebie oraz punktu przecięcia osi głównych, co oznacza, że nie ma znaczącej różnicy w kategoriach przedstawionych cech. Jedynie mało liczne warianty cech znalazły się poza centroidą.

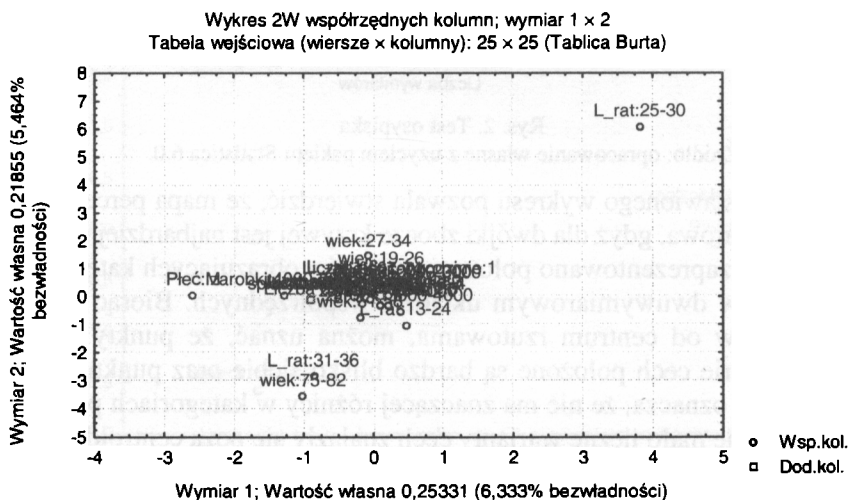
Przeprowadzona analiza korespondencji na całym zestawie danych nie dała zadowalających wyników, gdyż procent wyjaśnionej inercji wynosi ok. 13%. Więk-

² Analizy numeryczne przeprowadzono z użyciem pakietu Statistica.



Rys. 3. Mapa percepcji dla wszystkich cech
Źródło: opracowanie własne z użyciem pakietu Statistica 6.0.

szość punktów reprezentujących poszczególne wiersze i kolumny macierzy danych nie jest reprezentowana przez to rozwiązanie. Dowodzą tego niskie wartości jakości i niskie udziały w inercji wymiaru i punktu. Przyczyny takiego stanu należy upatrywać w tym, że punkt reprezentujący udział niewiarygodnych kredytobiorców, kredytów rozkładanych na od 31 do 36 rat oraz mężczyzn jest bardzo niski w stosunku do liczebności pozostałych wariantów. Doskonale to widać na przykładzie mapy percepcji, gdzie punkty reprezentujące te kategorie są w sporym oddaleniu od pozostałych. Reszta punktów zaś jest skupiona wokół centroidu.



Rys. 4. Mapa percepcji z punktem dodatkowym „wiarygodność kredytowa”
Źródło: opracowanie własne z użyciem pakietu Statistica 6.0.

Potraktowanie wiarygodności kredytowej jako punktu dodatkowego spowodowało zmniejszenie odsetka wyjaśnianej wariancji aż do 12%. To postępowanie nie doprowadziło do poprawy, a wręcz do spadku miary dobroci oraz jakości reprezentacji punktów.

4. Podsumowanie

Mimo wielu zalet i wad analiza korespondencji może być wykorzystywana w ocenie wiarygodności kredytobiorców. Niestety w przedstawionym przykładzie wyraźnie pokazała, iż wybrane cechy, tzn. płeć, wysokość wynagrodzenia, wiek, liczba osób w rodzinie klienta oraz liczba rat nie są cechami pozwalającymi na jednoznaczną identyfikację wiarygodności kredytobiorcy. Interpretując zależności w macierzy danych, możemy stwierdzić, że wszystkie profile są bliskie profilom średnim, gdyż wszystkie punkty leżą w bliskim sąsiedztwie początku układu współrzędnych. Zatem nie ma istotnych różnic w wariantach przedstawionych cech odpowiadającym wiarygodnym i niewiarygodnym kredytobiorcom.

Bardzo użyteczną własnością analizy korespondencji jest możliwość wprowadzenia punktów dodatkowych, które zazwyczaj dostarczają wskazówek ułatwiających interpretację wyników. Wprowadzenie jako punktu dodatkowego w analizowanym przypadku wiarygodności kredytowej spowodowało spadek jakości reprezentacji pozostałych punktów.

Literatura

- Gatnar E., Walesiak M. (red.) (2004), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE, Wrocław.
- Górnjak J. (2001), *Zastosowanie analizy korespondencji w badaniach społecznych i marketingowych*, [w:] Materiały SPSS dla kursu S2: Wprowadzenie do analizy danych w marketingu, SPSS Polska.
- Greenacre M.J. (1993), *Correspondence Analysis in Practice*, Academic Press, London.
- Lasek M. (2002), *Data mining. Zastosowania w analizach i ocenach klientów bankowych*, Biblioteka Menedżera i Bankowca, Warszawa.
- Stanimir A. (2005), *Analiza korespondencji jako narzędzie do badania zjawisk ekonomicznych*, AE, Wrocław.

CORRESPONDENCE ANALYSIS IN EVALUATION OF THE INBORROWER'S CREDITWORTHINESS

Summary

The evaluation by use correspondence analysis of the credit risk has been presented in this article. It is based on data gained from "Gwarant" company for individual credits on purchase set. The test contained 4999 credit contracts. Among them 477 contracts (9.54 %) were not repaid and 4522 contracts (90.46 %) were paid off. Therefore a mischievousness was only 9.54 % in analysing test.