

Agnieszka Stanimir

Akademia Ekonomiczna we Wrocławiu

METODY OCENY JAKOŚCI ODWZOROWANIA WSPÓŁWYSTĄPIEŃ KATEGORII CECH W ANALIZIE KORESPONDENCJI

1. Wstęp

Analiza korespondencji umożliwia rozpoznanie jednoczesnych wystąpień kategorii zmiennych nominalnych. W metodzie tej wyodrębnia się dwa podejścia:

- klasyczną analizę korespondencji (dla dwóch zmiennych, których liczebności jednoczesnych wystąpień zapisane są w tablicy kontyngencji; ten sam algorytm wykorzystywany jest również w analizie wielu zmiennych, których liczebności są zapisane w wielowymiarowej tablicy kontyngencji, łączonej tablicy kontyngencji, macierzy znaczników);

- analizę korespondencji wielu zmiennych (liczebności wystąpień kategorii zmiennych zapisane są w macierzy Burta).

Wyniki uzyskane po przeprowadzeniu analizy korespondencji w dowolnym z wymienionych podejść prezentowane są jako rozrzut punktów w przestrzeni o niskim wymiarze. Rzeczywisty wymiar przestrzeni współwystąpień kategorii zmiennych jest najczęściej bardzo wysoki (zob. punkt 2). Powstaje zatem pytanie, w jaki sposób określić ostateczny wymiar przestrzeni rzutowania, tak by utrata informacji o rzeczywistych współwystąpieniach kategorii cech była jak najmniejsza.

Celem artykułu jest zaprezentowanie powszechnie wykorzystywanych technik oceny jakości odwzorowania, które będą podstawą dalszych badań autorki.

2. Podstawowe pojęcia w analizie korespondencji

Krokiem wstępnym do zaprezentowania sposobów oceny jakości odwzorowania współwystąpień kategorii cech w analizie korespondencji jest przedstawienie pod-

stawowego algorytmu metody. Kolejność obliczeń wykonywana w klasycznej analizie korespondencji jest w analizie wielu zmiennych identyczna. Techniki te różnią się jedynie zapisem zmiennych w tabelach. Graficzną prezentację wyników wykonuje się dzięki zastosowaniu rozkładu macierzy według wartości osobliwych i wyznaczeniu na tej podstawie współrzędnych rzutowania kategorii zmiennych. W klasycznym podejściu przeprowadza się rozkład macierzy standaryzowanych różnic:

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

gdzie: $\mathbf{P}$ jest macierzą częstości zaobserwowanych, $\mathbf{r}$ ($\mathbf{c}$) – wektor częstości brzegowych wierszy (kolumn), $\mathbf{D}_r$ ($\mathbf{D}_c$) to diagonalna macierz częstości brzegowych wierszy (kolumn), $\mathbf{V}$ to diagonalna macierz niezerowych wartości osobliwych macierzy $\mathbf{A}$ (ułożonych w porządku nierosnącym), $\mathbf{U}$ to macierz lewych wektorów osobliwych, $\mathbf{V}$ to macierz prawych wektorów osobliwych, wektory zapisane w $\mathbf{U}$ oraz $\mathbf{V}$ są nazywane osiami głównymi rzutowania.

Współrzędne rzutowania kategorii cech zapisanych w wierszach (kolumnach) analizowanej macierzy wyznacza się następująco:

$$\mathbf{F} = \mathbf{D}_r^{-1/2} \mathbf{U}\mathbf{V}^T (\mathbf{G} = \mathbf{D}_c^{-1/2} \mathbf{V}\mathbf{V}^T). \quad (2)$$

W analizie korespondencji jednym z kluczowych pojęć jest inercja całkowita (λ), która zachowuje następujące relacje z macierzą $\mathbf{A}$ i $\mathbf{V}^2 = \mathbf{A}$:

$$\text{tr} \mathbf{A}^T \mathbf{A} = \text{tr} \mathbf{A} \mathbf{A}^T = \text{tr} \mathbf{A} = \frac{\chi^2}{n} = \lambda = \sum_{k=1}^K \lambda_k^2. \quad (3)$$

Inercja całkowita jest interpretowana jako miara zależności między cechami. Inercja całkowita jest sumą inercji głównych, czyli wartości własnych będących kwadratami wartości osobliwych wyznaczonych podczas dekompozycji macierzy $\mathbf{A}$.

Rzeczywisty wymiar przestrzeni rzutowania jest, jak już wspomniano, bardzo wysoki i w podejściu klasycznym jest równy $K = \min(r-1; c-1)$, a w analizie macierzy Burta

$$-K = \sum_{q=1}^Q (J_q - 1) \quad (\text{gdzie } J_q \text{ to liczba kategorii } q\text{-tej cechy, } q = 1, \dots, Q).$$

Korzystając z wartości współrzędnych w rzeczywistej przestrzeni rzutowania, można powrócić do liczebności zaobserwowanych analizowanej tabeli. Dla tablicy kontyngencji powiązanie liczebności zaobserwowanych n_{ij} i współrzędnych wierszy i kolumn jest następujące:

$$n_{ij} = n \cdot p_{i\bullet} \cdot p_{\bullet j} \cdot \left(1 + \sum_{k=1}^K \frac{f_{ik} \cdot g_{jk}}{\sqrt{\lambda_k}} \right), \quad (4)$$

gdzie: $p_{i\bullet}$ ($p_{\bullet j}$) to częstości brzegowe wierszy (kolumn); n jest liczbą badanych obiektów; λ_k to k -ta wartość własna; f_{ik} to współrzędna i -tej kategorii na k -tej osi; g_{jk} to współrzędna j -tej kategorii na k -tej osi.

3. 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.

Stopień wyjaśniania inercji

Popularnym i niewymagającym wielu obliczeń miernikiem oceny jakości odwzorowania rzeczywistych powiązań kategorii cech jest stopień wyjaśniania inercji.

Wskaźnik ten jest głównym sposobem określania wymiaru przestrzeni rzutowania (zob. [Andersen 1991, s. 374; van der Heijden 1987, s. 34; Sikorski 1987, s. 35]) i jest wyznaczany jako udział inercji wybranego wymiaru w inercji całkowitej:

$$\tau_{K^*} = \frac{\sum_{k=1}^{K^*} \lambda_k}{\sum_{k=1}^K \lambda_k} = \frac{\sum_{k=1}^{K^*} \lambda_k}{\lambda}, \quad (5)$$

gdzie: K^* to wybrany wymiar rzutowania ($K^* \leq K$); λ_k są to wartości własne k -tej osi ($k = 1, \dots, K$).

Inercja wybranego wymiaru jest sumą wszystkich wartości własnych przypisanych osiom głównym tego wymiaru: $\sum_{k=1}^{K^*} \lambda_k$.

Sikorski [1987, s. 35] interpretuje τ_{K^*} jako miarę jakości odtworzenia obserwacji. Biorąc pod uwagę powiązania statystyki χ^2 , całkowitej inercji λ oraz wartości własnych (por. wzór (3)), można przyjąć, że τ_{K^*} jest to proporcja χ^2 , która została zdekomponowana w wymiarze K^* :

$$\tau_{K^*} = \frac{\sum_{k=1}^{K^*} \chi_k^2}{\chi^2}. \quad (6)$$

Przyjmuje się, że najlepszym rozwiązaniem jest wybranie takiego K^* , gdy τ_{K^*} jest bliskie 1, a po zwiększeniu wymiaru przestrzeni wartość tego wskaźnika nie wzrasta gwałtownie. Oznacza to, że suma inercji głównych wybranej przestrzeni rzutowania jest bardzo bliska inercji całkowitej. Stawiając dodatkowe ograniczenie, że prezentacja graficzna współwystępowania kategorii może być wykonana jedynie w przestrzeni jedno-, dwu- lub trójwymiarowej, należałoby oczekiwać, że K^* powinno być równe 1, 2 lub 3.

* Punkt 3 zawiera obszerne fragmenty rozprawy doktorskiej autorki pt. „Wykorzystanie analizy korespondencji w badaniach marketingowych”.

Jednoczesne spełnienie obydwu wymagań odnośnie do wartości K^* jest możliwe najczęściej w klasycznej analizie korespondencji. W zwykłej tablicy kontyngencji, łączonej tablicy kontyngencji, macierzy znaczników (dla małej liczby obiektów, zmiennych i kategorii) oraz wielowymiarowej tablicy kontyngencji rzeczywisty wymiar rzutowania nie jest wysoki, a dwie pierwsze wartości własne mają w sumie udział w inercji całkowitej zwykle równy łącznemu udziałowi pozostałych wartości własnych bądź większy od niego.

Problem wyboru przestrzeni rzutowania, tak by wartość τ_{K^*} była bliska jedności, staje się trudny do rozwiązania w analizie macierzy Burta lub bardzo rozbudowanej złożonej macierzy znaczników Z . Rzeczywisty wymiar prezentacji współwystępowania kategorii zmiennych jest w takim przypadku bardzo wysoki. Pierwsza wartość własna nie jest już tak znacznie wyższa od pozostałych, jak w klasycznej analizie korespondencji. Kolejne inercje główne nie różnią się już w tak dużym stopniu. W związku z tym τ_{K^*} będzie bliskie 1, gdy K^* będzie bliskie K .

Macierz odtworzenia

Ważną metodą oceny jakości odwzorowania jednoczesnego wystąpienia kategorii cech jest porównanie liczebności rzeczywistych z liczebnościami wyznaczonymi za pomocą formuły odtworzenia (zob. [Benzécri 1992, s. 126; Greenacre 1984, s. 93; van der Heijden 1987, s. 32]).

Macierz odtworzenia jest macierzą liczebności obliczonych na podstawie wartości własnych i współrzędnych wybranego wymiaru przestrzeni rzutowania.

W klasycznej analizie korespondencji, na podstawie zależności wyrażonej wzorem (4), budowana jest macierz odtworzenia wejściowej macierzy danych, której elementy są wyznaczane w następujący sposób:

$$\tilde{n}_{ij} = n \cdot p_{i\cdot} \cdot p_{\cdot j} \cdot \left(1 + \sum_{k=1}^{K^*} \frac{f_{ik} \cdot g_{jk}}{\sqrt{\lambda_k}} \right). \quad (7)$$

Otrzymane wartości $\tilde{n}_{ij}$ są zaokrąglane do najbliższej wartości całkowitej.

Wynik uzyskany po zastosowaniu wzoru (7) jest, jak pisze Benzécri [1992, s. 128], obrazem wzajemnego przyciągania (*attraction*) lub odpychania (*repulsion*) kategorii różnych cech. Przyciąganiem można określić stan, w którym jest obserwowane współwystępowanie dwóch kategorii różnych cech. Gdy kategorie sporadycznie występują wspólnie, można scharakteryzować zaistniałą sytuację jako odpychanie się wariantów cech. Analizując np. wzór (7), można stwierdzić, że przyciąganie między i -tą oraz j -tą kategorią następuje, gdy $(f_{ik} \cdot g_{jk}) > 0$ (gdzie $k = 1, \dots, K^*$); wtedy liczebność odtworzona jednoczesnego wystąpienia i -tej i j -tej

kategorii wzrasta o $\sum_{k=1}^{K^*} \frac{f_{ik} \cdot g_{jk}}{\sqrt{\lambda_k}}$ ponad liczebność oczekiwaną $(n \cdot p_{i\cdot} \cdot p_{\cdot j})$. Odpy-

chanie kategorii występuje, gdy $(f_{ik} \cdot g_{jk}) < 0$ (gdzie $k = 1, \dots, K^*$), co w rezultacie powoduje obniżenie oczekiwanej liczebności jednoczesnego wystąpienia analizowanych dwóch kategorii. Wyrażenie $(f_{ik} \cdot g_{jk})$ przyjmuje wartości dodatnie, gdy punkty obrazujące i -tą i j -tą kategorię mają współrzędne na k -tej osi o tym samym znaku (podczas projekcji na k -tą oś punkty te są umiejscowione po tej samej stronie centrum rzutowania). Wyrażenie $(f_{ik} \cdot g_{jk})$ przyjmuje wartości ujemne, gdy znaki współrzędnych i -tej i j -tej kategorii na k -tej osi są różne. Gdy odpychanie dwóch kategorii różnych cech jest bardzo duże, może zdarzyć się, że $\sum_{k=1}^{K^*} \frac{f_{ik} \cdot g_{jk}}{\sqrt{\lambda_k}} < -1$, co z kolei spowoduje wystąpienie w macierzy odtworzenia wartości ujemnych. Zaistnienie takiej sytuacji oznacza, że jednoczesne wystąpienie i -tej oraz j -tej kategorii jest w rzeczywistości, w porównaniu ze wspólnymi wystąpieniami innych kategorii, niemożliwe.

W trakcie analizy korespondencji opartej na macierzy Burta formułę odtworzenia konstruuje się w analogiczny sposób do podejścia klasycznego.

Po porównaniu liczebności zaobserwowanych n_{ij} oraz odtworzonych $\tilde{n}_{ij}$ można stwierdzić, czy istnieją między tymi wartościami istotne różnice. Poniżej przedstawiono propozycje oceny jakości odwzorowania rzeczywistej sytuacji współwystępowania cech, wykorzystujące liczebności odtworzone.

Andersen [1991, s. 374] proponuje, by ocenę jakości odwzorowania wykonać na podstawie statystyki największej wiarygodności:

$$L^2 = 2 \cdot \sum_{i=1}^r \sum_{j=1}^c n_{ij} \cdot (\ln n_{ij} - \ln \tilde{n}_{ij}). \quad (8)$$

Statystyka ta jest sprawdzianem prawdziwości hipotezy

$$H_0: \lambda_k = 0, \quad \text{gdzie } k > K^*, \lambda_1 > \lambda_2 > \dots > \lambda_{K^*} > 0$$

oraz ma rozkład χ^2 z $(r - K^* - 1)(c - K^* - 1)$ stopniami swobody. Hipotezą alternatywną wobec H_0 jest:

$$H_1: \lambda_k > 0, \quad \text{gdzie } k > K^*.$$

Jeżeli zatem nie ma podstaw do odrzucenia hipotezy H_0 , to wybrany wymiar rzutowania jest najlepszy, a jakość odtworzenia – wysoka [Gilula, Haberman 1986].

Tak jak wspomniano, w macierzy odtworzenia może się pojawić wartość ujemna liczebności jednoczesnego wystąpienia dwóch kategorii różnych cech. W takim wypadku zastosowanie testu L^2 jest niemożliwe.

Andersen [1991, s. 377; 1997, s. 224] proponuje inny sposób porównania liczebności n_{ij} oraz $\tilde{n}_{ij}$, wykorzystując w tym celu statystykę χ^2 :

$$\chi_{K^*}^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(n_{ij} - \tilde{n}_{ij})^2}{(n_{i\bullet} \cdot n_{\bullet j})/n}. \quad (9)$$

Można jednak wykazać, że:

$$\chi_{K^*}^2 = n \cdot \sum_{k=K^*+1}^K \lambda_k. \quad (10)$$

Andersen [1997, s. 224] stwierdza, że wartość $\chi_{K^*}^2$ jest zbliżona do wartości statystyki χ^2 obliczonej dla wejściowej tablicy kontyngencji wtedy, gdy prezentacja w przestrzeni K^* -wymiarowej w niewielkiej części wyjaśnia powiązania kategorii zmiennych. Wartość $\chi_{K^*}^2$ niższa od χ^2 i dążąca do zera wskazuje na wyjaśnianie współwystępowania kategorii cech w bardzo wysokim stopniu. Miara ta może być bezpośrednio zastosowana do oceny jakości odwzorowania w analizie zwykłych, łączonych lub wielowymiarowych tablic kontyngencji.

Andersen [1991, s. 378; 1997, s. 225] proponuje, by po obliczeniu statystyki $\chi_{K^*}^2$ wyznaczyć następującą miarę stopnia przeniesienia powiązań występujących między kategoriami cech, z tablicy kontyngencji do przestrzeni o wymiarze $K^* < K$:

$$r_{K^*}^2 = \frac{\chi^2 - \chi_{K^*}^2}{\chi^2}. \quad (11)$$

Korzystając z zależności: $\lambda = \frac{\chi^2}{n}$ oraz $\lambda = \sum_{k=1}^K \lambda_k$, jak również ze wzoru (10),

można się przekonać, że:

$$r_{K^*}^2 = \tau_{K^*}. \quad (12)$$

Okazuje się, że proponowany przez Andersena miernik stopnia przeniesienia powiązań kategorii cech z tablicy kontyngencji do przestrzeni K^* -wymiarowej jest równy udziałowi sumy wartości własnych z wybranej przestrzeni rzutowania w inercji całkowitej.

„Korelacje” punktów z osiami

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. Jeżeli optymalna przestrzeń rzutowania jest wymiaru K^* , to rozpatrywane są korelacje punktów (kategorii cech) z K^* pierwszymi osiami. Korelacje punktu z osią wskazują, która oś najlepiej opisuje punkt (kategorię).

Korelacje punktów z osiami mogą być wyznaczone niezależnie od zastosowanego w analizie korespondencji podejścia i są obliczane w następujący sposób:

$$cor_{ik}^2 = \frac{p_{i\bullet} \cdot f_{ik}^2}{\sum_{k=1}^K p_{i\bullet} \cdot f_{ik}^2} = \frac{\lambda_{7k}}{\lambda_i}, \quad (13)$$

$$cor_{jk}^2 = \frac{p_{\bullet j} g_{jk}^2}{\sum_{k=1}^K p_{\bullet j} g_{jk}^2} = \frac{\lambda_{jk}}{\lambda_j}, \quad (14)$$

gdzie: $\lambda_{ik} = p_{i\bullet} \cdot f_{ik}^2$ ($\lambda_{jk} = p_{\bullet j} \cdot g_{jk}^2$) to inercja i -tej (j -tej) kategorii na k -tej osi;

$\lambda_i = \sum_{k=1}^K p_{i\bullet} \cdot f_{ik}^2$ ($\lambda_j = \sum_{k=1}^K p_{\bullet j} \cdot g_{jk}^2$) to inercja i -tej (j -tej) kategorii ($\lambda = \sum_{i=1}^r \lambda_i$ lub

$$\lambda = \sum_{j=1}^c \lambda_j).$$

Można zapisać, że inercja k -tej osi wynosi:

$$\lambda_k = \sum_{i=1}^r \lambda_{ik} = \sum_{j=1}^c \lambda_{jk} = \sum_{i=1}^r p_{i\bullet} \cdot f_{ik}^2 = \sum_{j=1}^c p_{\bullet j} \cdot g_{jk}^2. \quad (15)$$

W pracach Clausena [1998, s. 17] oraz Lebart, Morineau i Warwicka [1984, s. 47] występuje procentowy wkład punktu (kategorii) w tworzenie określonej osi. Udział każdego punktu oblicza się dzieląc λ_{ik} lub λ_{jk} przez λ_k .

Znak korelacji obliczanej jako pierwiastek z wyrażenia (13) i (14) zależy od znaku współrzędnej punktu na określonej osi.

Na podstawie wzorów (13) i (14) można stwierdzić, że wartość cor_{ik}^2 , (cor_{jk}^2) wskazuje, jaka część udziału i -tego (j -tego) punktu w całkowitej inercji zależy od udziału i -tej (j -tej) kategorii w k -tej inercji głównej.

Wartość korelacji punktu z osią wskazuje, jak dobrze punkt jest opisywany przez daną oś. Na tej podstawie wyznaczany jest miernik jakości odzwierciedlenia rzeczywistych wystąpień kategorii w wybranej przestrzeni. Jeśli wybrana najlepsza przestrzeń rzutowania jest K^* -wymiarowa, to jakość odwzorowania można zmierzyć za pomocą następującego miernika:

$$qcor_{iK^*}^2 = \sum_{k=1}^{K^*} cor_{ik}^2, \quad (16)$$

$$qcor_{jK^*}^2 = \sum_{k=1}^{K^*} cor_{jk}^2. \quad (17)$$

Wartość tego miernika wskazuje, jaki procent inercji i -tej kategorii jest zawarty w przestrzeni K^* -wymiarowej. Suma kwadratów korelacji punktów ze wszystkimi osiami jest równa 1.

4. Podsumowanie

W przedstawionej pracy autorka zdecydowała się zaprezentować trzy sposoby oceny jakości odwzorowania współwystąpień kategorii cech.

Pierwszy z nich – stopień wyjaśniania inercji – można nazwać miernikiem uniwersalnym. Jednak nie można go uznać za jedyny i wystarczający. Stopień wyjaśniania inercji jest w głównej mierze sposobem wyboru wymiaru przestrzeni rzutowania, a następnie dopiero może być wykorzystany jako kryterium oceny poprawności przeprowadzonej analizy. Wskaźnik ten, mimo iż jego sformułowanie jest zrozumiałe, nie rozwiązuje problemów nurtujących badacza. Niedoskonałość tego miernika jest szczególnie widoczna w przypadku analizy macierzy Burta, wskazującej bardzo wysoki wymiar przestrzeni rzutowania rzeczywistych powiązań (pierwsze inercje główne nie są już znacznie wyższe od kolejnych).

Dla większości tablic (z wyłączeniem macierzy Burta) możliwe jest ocenianie jakości odwzorowania z wykorzystaniem macierzy odtworzenia. Metoda ta, tak jak poprzednia, wydaje się bardzo trafna. W trakcie jej stosowania można jednak napotkać pewne trudności, które opisano w artykule. Wprowadzona przez Andersena [1991] modyfikacja tej metody prowadzi do miernika, który przedstawiono wcześniej, czyli stopnia wyjaśniania inercji.

Jako miary jakości odwzorowania rzeczywistych powiązań kategorii cech w literaturze są również proponowane korelacje punktów z osiami oraz elementy z nimi związane. Mierniki te w najpełniejszy sposób analizują powiązania występujące między wszystkimi punktami i wszystkimi osiami wybranej przestrzeni rzutowania. Jedynym minusem jest tu konieczność przeprowadzenia wielu skomplikowanych obliczeń i dokładnego rozpatrzenia każdej otrzymanej wartości. Z tego względu w wielu opracowaniach, w których analiza korespondencji jest jedynie narzędziem i nie stanowi centralnego punktu badania, ten rodzaj wskaźników oceniających jakość odwzorowania współwystąpień kategorii cech jest pomijany.

Wszystkie zaprezentowane metody oceny jakości odwzorowania mają swoje wady i zalety. Aby poprawnie i dogłębnie ocenić jakość prezentacji wyników analizy korespondencji, należy zastosować jak najwięcej z nich.

Literatura

- Andersen E.B. (1997), *Introduction to the Statistical Analysis of Categorical Data*, Springer-Verlag, Berlin.
- Andersen E.B. (1991), *The Statistical Analysis of Categorical Data*, Springer-Verlag, Berlin.
- Benzécri J.-P. (1992), *Correspondence Analysis Handbook*, Marcel Dekker Inc, New York.
- Clausen S.E. (1998), *Applied Correspondence Analysis. An Introduction*, Sage, University Paper 121.
- Gilula Z., Haberman S.J. (1986), *Canonical Analysis of Contingency Tables by Maximum Likelihood*, „Journal of the American Statistical Association” 81, 395, s. 780-788.

- Greenacre M. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press, London.
- Heijden van der P.G.M. (1987), *Correspondence Analysis of Logitudinal Categorical Data*, DSWO Press, Leiden.
- Lebart L., Morineau A., Warwick K.M. (1984), *Multivariate Descriptive Statistical Analysis. Correspondence Analysis and Related Techniques for Large Matrices*, John Wiley & Sons, Inc., New York.
- Sikorski A. (1987), *Modelowanie zachowań nabywczych – metody analizy wielowymiarowej*, Instytut Rynku Wewnętrznego i Konsumentów, Warszawa.
- Stanimir A. (2002), *Wykorzystanie analizy korespondencji w badaniach marketingowych*, praca doktorska, AE, Wrocław.

METHODS OF QUALITY MEASUREMENT IN DISPLAYING RELATIONS BETWEEN CATEGORIES OF VARIABLES IN CORRESPONDENCE ANALYSIS

Summary

Article presents methods of quality measurement in displaying relations between categories of variables in correspondence analysis. This problem is very important, because analysis' results are usually presented in less than full dimensional space. Diagnosis of range of information lost while displaying relations between categories of variables plays important supporting role in the interpretation of results of correspondence analysis.