

An application of nonlinear canonical correlation analysis on medical data

Ayşe Canan YAZICI¹, Ersin ÖĞÜŞ¹, Handan ANKARALI², Fikret GÜRBÜZ³

Aim: In studies in the field of medicine, necessity to examine the relations between data sets composed of categorical variable groups is widely encountered. In this study, it was aimed to examine nonlinear canonical correlation analysis (OVERALS) method, which allows examination of relations among K number of categorical variable sets and structural similarities of the data set, and to discuss usefulness of the method in more comprehensive data sets obtained from studies carried out in the field of medicine in terms of practice and interpretation.

Materials and methods: OVERALS method was applied to a part of data set obtained from a study carried out with diarrhea patients. In the study, 10 variables were divided into 3 groups, namely anamnesis, symptoms, and laboratory tests. In order to examine similarities and relationships among these 3 variable groups, OVERALS method was used and results were expressed with graphical presentations.

Results: It was observed that OVERALS analysis allows more detailed presentation of data structure and relations among variable sets.

Conclusions: OVERALS analysis proved to be a quite useful method in graphical expression and interpretation of data structure, revealing similarities and relational structures among multi-dimensional categorical variable sets, which are used often in the field of medicine and their comprehensive interpretation.

Key words: Alternating least squares algorithm (ALS), homogeneity analysis (HOMALS), multivariate analysis, nonlinear canonical correlation analysis (OVERALS)

Doğrusal olmayan kanonik korelasyon analizinin tıp alanında kullanımı

Amaç: Tıp alanında yapılan çalışmalarda kategorik değişken kümelerinden oluşan veri setleri arasındaki ilişkilerin incelenmesi gerekliliği ile yaygın bir biçimde karşılaşılmaktadır. Bu çalışmada K adet kategorik değişken kümesi arasındaki ilişkileri ve veri kümelerinin yapısal benzerliklerini inceleme olanağı sağlayan doğrusal olmayan kanonik korelasyon analizi (OVERALS) yönteminin incelenmesi, tıp alanında yapılan çalışmalardan elde edilen geniş kapsamlı veri setlerinin analizinde yöntemin uygulama ve yorum bakımından tartışılması amaçlanmıştır.

Yöntem ve gereç: OVERALS yöntemi, ishal hastaları ile yapılan bir çalışmadan elde edilen veri setinin bir bölümüne uygulanmıştır. Çalışmada, hastanın öyküsü, belirtiler ve laboratuvar testleri olmak üzere üç gruba ayrılan on adet değişken ele alınmıştır. Bu üç değişken grubu arasındaki benzerlik ve ilişkileri incelemek üzere OVERALS yöntemi kullanılmış ve sonuçlar grafiksel gösterimlerle de ifade edilmiştir.

Bulgular: OVERALS çözümlenmesi ile veri yapısının ve değişken kümeleri arasındaki ilişkilerin daha ayrıntılı bir biçimde ortaya konarak yorumlanabildiği görülmüştür.

Sonuç: OVERALS çözümlenmesinin, veri seti yapısının grafiksel olarak ifade edilebilmesi ve yorumlanabilmesi, tıp alanında sıklıkla karşılaşılan çok boyutlu kategorik değişken kümeleri arasındaki benzerlik ve ilişki yapılarını ortaya çıkartabilmek ve kapsamlı bir biçimde yorumlayabilmek bakımından oldukça kullanışlı bir yöntem olduğu görülmüştür.

Anahtar sözcükler: En küçük kareler algoritması, homojenite analizi, çok değişkenli analiz, doğrusal olmayan kanonik korelasyon analizi (OVERALS)

Received: 27.03.2008 – Accepted: 09.09.2009

¹ Department of Biostatistics, Faculty of Medicine, Başkent University, Ankara - TURKEY

² Department of Biostatistics, Faculty of Medicine, Düzce University, Bolu - TURKEY

³ Department of Biometry and Genetics, Faculty of Agriculture, Ankara University, Ankara - TURKEY

Correspondence: Ayşe Canan YAZICI, Department of Biostatistics, Faculty of Medicine, Başkent University, Ankara - TURKEY

E-mail: acyazici@baskent.edu.tr

Introduction

It is well known that in certain medical studies, particularly in detailed extensive surveys, the number of examined variables is usually high. Variables obtained from such studies are occasionally categorical or categorized later. There are certain limitations in analyzing categorical data using the classical statistical techniques. For instance, they require some assumptions about the data and obtained information. A major difference between OVERALS and most other techniques for categorical data analysis lies in the use of models. For example in log-linear analysis, a distribution is assumed about the data, after that a model for the data is hypothesized and estimations are made under the assumption that this model is true. Then, for evaluating the model, these estimations are compared with the observed frequencies (1). In OVERALS, there is no need for an assumption about the underlying distribution of the data and no model has to be hypothesized. In addition, a composition of the data is obtained to study the structure in the observed data in some optimal way. It is able to integrate the analysis of more than K sets of categorical variables while retaining all facets of the data by OVERALS.

In standard canonical correlation analysis, the goal is to explain as much variance in the relationships among 2 sets of numerical variables as possible in a low dimensional space. The optimal scaling and consequently OVERALS approach expand the standard analysis in 3 crucial ways. Firstly, OVERALS allows treating more than 2 sets of variables; more than 1 independent set and 1 dependent set. Thus, nonlinear relationships among variable sets can be analyzed. However, variables can be scaled as either nominal, ordinal, or numerical. Moreover, instead of maximizing correlations between the variable sets, the sets are compared to an unknown compromise set that is defined by the object scores (2).

The OVERALS technique analyzes relationships among K sets of variables and searches for what is common among sets of variables measured on the same objects. The method involves comparing K sets of variables to one another in the same graph, after having removed linear dependencies within each of the sets. It is an interesting technique because it subsumes a number of existing techniques for multivariate data analysis as special cases (3-7).

The purpose of OVERALS is to determine how similar sets of categorical variables are to each other. Analogously to the situation in multiple regression and canonical correlation analyses, OVERALS focuses on the relationships among sets; all particular variables contribute to the result as long as it provides information that is independent of the other variables in the same set. The goal is to account for as much variance in the relationships among the sets as possible in a low-dimensional space (3-6, 8).

The aim of this study was the application of OVERALS using a real medical data and to discuss the usefulness of the application and interpretation of the method in medical studies.

Materials and methods

An application sample

Data used in this study were taken from a specialty thesis with permission (9). It was part of a large survey data collected at the Medical Faculty of Başkent University, Department of Clinical Microbiology and Infectious Disease in Ankara, Turkey, between November 2002 and November 2003. The study was performed following the approval of the ethics committee of Başkent University. The survey was about patients with diarrhea. Data were obtained from 85 patients, consisting of 38 women and 47 men. Mean age and standard deviation of the patients were 33.65 ± 11.99 and median was 29 years.

We used 10 variables from the survey and classified them in 3 sets: anamnesis, symptoms, and laboratory tests. The labels of the sets, variables and categories in data, and the symbols that represent the categories in graphics are given in Table 1.

Anamnesis set was formed of variables of suspicious food intake and transplantation history. Symptoms set was composed of abdominal pain, nausea/vomiting history, the type of diarrhea, and fever variables. Leukocyte count, erythrocyte count, and parasites under microscope with trichrome stain were the laboratory test variables.

Statistical analysis

The OVERALS model is a form of homogeneity analysis with restrictions. Homogeneity analysis determines transformations of the categories of

variables to maximize homogeneity and aims at the representation of the structure of non-numerical multivariate data. A certain criterion is optimized by assigning scores to objects and categories of variables. These scores can be used to construct a geometrical representation of the dependencies in the data in a low dimensional Euclidean space (3,5,10-15).

Homogeneity analysis is the basic technique in the Gifi system of descriptive nonlinear multivariate analysis. The Gifi system is characterized by the optimal scaling of categorical variables, which is implemented through alternating least squares algorithms (5).

Assume that J denotes the categorical variables collected for N objects or individuals, where variable $j \in J = \{1,2,\dots,J\}$ has c_j categories. The aim is to construct a low-dimensional joint map of objects and categories in Euclidean space R^p . The interest is representing these objects in a p -dimensional space ($p < J$). X is the $N \times p$ matrix whose elements are also known as the object scores called object scores matrix and containing the coordinates of the object vertices in R^p and the resulting p optimal scales. $Y_j(j \in J)$ is a matrix called the $c_j \times p$ category quantification matrix containing the coordinates of the c_j category vertices of variable j and multiple category quantifications of variable $j \in J$. G_j is the indicator matrix for variable j of order $N \times c_j$. It is a binary matrix with entries $G_j(i, t) = 1, i = 1,\dots, N, t = 1,\dots, c_j$ if object i belongs to category t , and $G_j(i, t) = 0$ if it belongs to some other category. According to the homogeneity principles, we would like to quantify (transform) the variables to achieve maximum homogeneity. The matrix $G = (G_1, \dots, G_j)$ is simply the adjacency matrix of the bipartite graph. If edges are used to connect each category, the loss function is the average squared edge length (over all variables) and given by

$$\sigma(X; Y_1, \dots, Y_J) = J^{-1} \sum_{j=1}^J SSQ(X - G_j Y_j = J^{-1} \text{tr}(X - G_j Y_j)'(Y - G_j Y_j) \tag{1}$$

Symbol $SSQ(\cdot)$ is used for the sum of squares of the elements of a vector or matrix.

The loss function (1) is called the Gifi loss function.

Alternating Least Squares algorithm (ALS) was used to minimize the loss function. The minimization is subject to the condition that

$$X'X = NI_p \tag{2}$$

for avoiding the trivial solution corresponding to $X = 0$, and $Y_j = 0$ for every $j \in J$ and

$$u'X = 0 \tag{3}$$

where, u is a column with n elements equal to one.. The condition $u'X = 0$ guarantees that X is in deviations from the column means, while $X'X = NI_p$ makes the columns of X uncorrelated, with variances equal to 1. Elements of X are called object scores. The ALS algorithm cycles through the following steps until it converges.

In the first step, (1) is minimized with respect to Y_j for fixed X . The set of normal equation is given by

$$D_j Y_j = G_j' X, \tag{4}$$

where $D_j = G_j' G_j$ is the $c_j \times c_j$ diagonal matrix containing the univariate marginal of variable j . Hence, the solution of equation (4) is given by

$$\hat{Y}_j = D_j^{-1} G_j' X, \quad j \in J \tag{5}$$

In the second step of the algorithm, the loss function is minimized with respect to X for fixed Y_j 's and the result is given by

$$\hat{X} = J^{-1} \sum_{j=1}^J G_j \hat{Y}_j \tag{6}$$

In the third step of the algorithm, the X matrix is column-centered and then orthonormalized by the modified Gram-Schmidt procedure,

$$X = \sqrt{N} \text{GRAM}(W) \tag{7}$$

where $W = \hat{X} - u(u'\hat{X}/N)$

These steps are repeated until the algorithm converges to the global minimum. This solution is also known in the literature as the HOMALS solution (homogeneity analysis by means of alternating least squares).

HOMALS is primarily a data-descriptive technique of primarily categorical data. The main objective is to scale the categories so that a particular criterion is optimized (the edge length loss function (1)). The Gifi groups by considering generalizations of the loss function (1) and by placing restrictions on the category quantifications, attempts to incorporate other popular multivariate techniques in the system, while retaining the focus on the graphical representations of the data and the exploratory nature of the techniques (4,5,16,17-24).

Generalizations of the homogeneity analysis leading to OVERALS

In Hotelling’s canonical correlation analysis, one studies the relationship between 2 sets of variables after having removed linear dependencies within each of these sets (5). OVERALS involves comparing K sets of variables after having removed linear dependencies within each set. Various approaches suggested generalizing Hotelling’s canonical correlation procedure to K sets of variables. In a K set problem, there are $K(K-1)/2$ canonical correlations among the optimal set of canonical variables that can be collected in a $K \times K$ correlation matrix R. The generalizations deal with different criteria that can be formulated as functions of the matrix R. In the Gifi system, the criteria that maximize the largest eigenvalue of R that is equivalent to maximize the sum of correlations between each canonical variable and an unknown coordinate vector x is considered (5). The set J of the j variables is classified into K subsets $J(1), \dots, J(k), \dots, J(K)$. So, the generalization of the Gifi loss function is given in Eq. (8).

$$\sigma(X; Y_1, \dots, Y_j) = K^{-1} \sum_{k=1}^K \text{SSQ}(X - \sum_{j \in J(k)} G_j Y_j) \quad (8)$$

This function is minimal with the normalization, as in homogeneity analysis, of $u'X = 0$ and $X'X = NI_p$. According to Eq. (8), all variables within each set $J(k)$, $k = 1, \dots, K$ are treated as additive and optimal transformations of a variable j within a set $J(k)$

depends on the optimal transformations of the remaining variables of set $J(k)$. So, a correction is employed for the contribution of the other variables and is reflected in the ALS algorithm given below.

In the first step, the optimal Y_j for given X is:

$$Y_j = D_j^{-1} G_j' (X - V_{kj}), j \in J, \quad (9)$$

where $V_{kj} = \sum_{j \in J(k)} G_j Y_j - G_j Y_j, k = 1, \dots, K, j \in J$

In the second step, the optimal X for given Y_j 's is:

$$X = K^{-1} \sum_{k=1}^K \sum_{j \in J(k)} G_j Y_j \quad (10)$$

In the third step of the algorithm, the X matrix is column centered and then orthonormalized in order to satisfy the normalization constraints.

This ALS algorithm is known in the literature as the OVERALS algorithm (3,5,8).

In this study, data analysis was performed using SPSS version 13.0 (SPSS Inc. Chicago, IL, USA).

Results

Relationships and similarities among and within these 3 sets of variables, which were analyzed using OVERALS, are presented in Table 1.

Loss values, eigenvalues, and fit values to show the similarities among sets are presented in Table 2. Fit and loss values displayed how well the OVERALS solution fits the optimally quantified data with respect to the association among sets (4).

Eigenvalue indicates the level of relationship shown by each dimension. Maximum value of an eigenvalue is 1 with a minimum of 0. Eigenvalues obtained from the study were quite high (0.698 and 0.648). An actual fit value of 1.346, which is the sum of eigenvalues, was calculated for variation. We used 2-dimensional solutions, so $1.346 / 2 = 67.3\%$ of the variation was calculated in the analysis. $0.698 / 1.34652\%$ of the actual fit was calculated by the first dimension and $0.648 / 1.34648\%$ by the second dimension. Loss represents the proportion of variation in object scores for each dimension and set in Table 2.

Table 1. The investigated variable sets, categories, and their coding.

Sets	Variables	Categories	Category symbols
Anamnesis	Suspicious food intake history	Yes No	SY SN
	Transplantation history	Yes No	TY TN
Symptoms	Abdominal pain	Yes No	AY AN
	Nausea/vomiting history	Yes No	NVY NVN
	Diarrhea type	Mucous Watery	DM DW
		Hemorrhagic mucous Foamy, malodorous, excessive	DHM DFME
	Fever	Yes No	FY FN
Laboratory tests	Leukocyte count	No Rare Moderate Abundant	LN LR LM LA
	Erythrocyte count	None Rare Moderate Abundant	EN ER EM EA
	Parasites under microscope	None Giardia Entamoeba Blastocyst	PN PG PE PB

Table 2. Two dimensional solution results.

Sets	Dimensions		Sum	
	Dimension 1	Dimension 2		
Loss	Anamnesis	0.602	0.836	1.437
	Symptoms	0.183	0.127	0.310
	Laboratory tests	0.122	0.092	0.214
	Mean of sets	0.302	0.352	0.654
Eigenvalue	0.698	0.648		
Fit			1.346	

Mean of sets is the average loss in sets and gives us the difference between the maximum and actual fits; it was $2 - 1.346 = 0.654$ in our study and not necessarily high. Summation of average loss and fit is equal to the number of dimensions ($0.654 + 1.346 = 2$). Therefore, small loss values indicate large multiple correlations between weighted sums of optimally scaled variables and dimensions (4-6).

Component loadings presented in Table 3 give the correlations between object scores and optimal scaled variables.

In the case of no data loss, component loadings are equal to Pearson correlations between quantified variables and object scores. Multiple nominal variables have 2 component loadings. Because of the quantifications of this kind of variable, they can differ for each dimension (4, 25). Component loadings are presented in Table 3 and they were called dimension 1 and dimension 2. These columns were coordinates of the variable points on the graph given below in Figure 1.

Component loadings can be interpreted easily by this graphical display. Parasites and diarrhea type

variables were multiple nominal, so there are 2 points plotted for them in Figure 1. Each of the quantifications was interpreted as a single variable. The distance from the origin to each variable point approximates the importance of that variable (13, 15). In this example, component loadings indicate that diarrhea type, parasites, and leukocyte and erythrocyte variables were the most effective variables in relationships among variable sets, because they were positioned far away from the origin. This means that, transplantation and fever variables were moderately effective. Other variables have no intense effect on relationships, because they were positioned close to the origin, which denotes the “mean”. Overall, it was observed that symptoms set and laboratory test set were more related with each other compared to the anamnesis set.

The plot of centroids was labeled by categories presented in Figure 2. This plot shows how well variables separate groups of objects. Centroids were in the center of gravity of the objects. In order to understand the relationships between variables, matching clusters of categories in centroid plots need to be identified (4).

Table 3. OVERALS component loadings.

Sets	Variables	Dimension 1	Dimension 2	
Anamnesis	Suspicious food intake history	-0.274	-0.350	
	Transplantation history	-0.496	0.275	
Symptoms	Abdominal pain	0.044	-0.471	
	Nausea/vomiting history	0.009	-0.600	
	Diarrhea type	Dimension 1	0.301	0.023
		Dimension 2	0.874	-0.049
	Fever	0.028	0.907	
Laboratory tests	Leukocyte count	0.434	-0.672	
	Erythrocyte count	0.598	-0.677	
	Parasites under microscope	Dimension 1	0.590	0.743
		Dimension 2	0.604	0.732

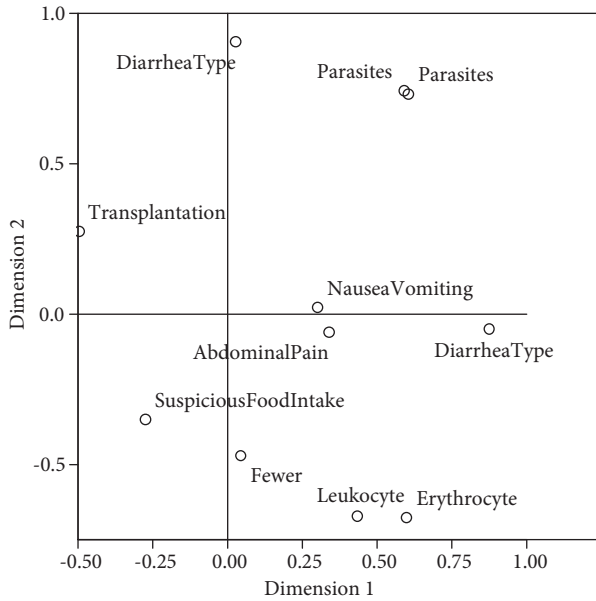


Figure 1. Component loadings.

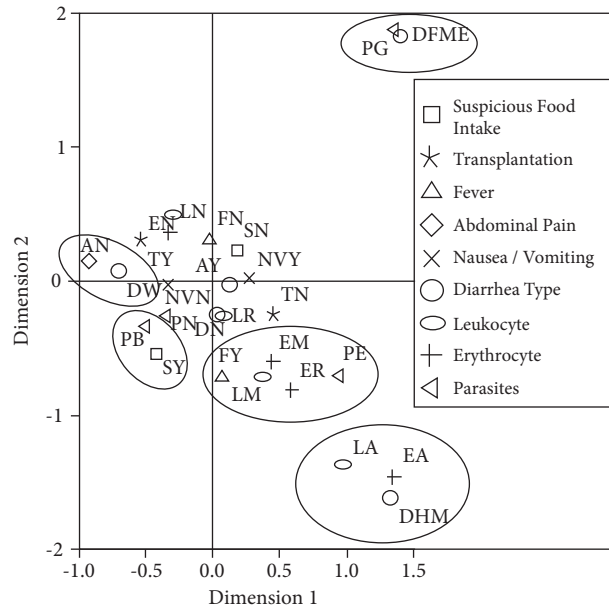


Figure 2. Centroids plot.

Relationships between symptoms set and laboratory tests set were observed clearly at the upper right and lower right corners of the plot. Relationships between diarrhea type variable of symptoms set and parasites under microscope variable of laboratory tests set were observed clearly at the upper right corner of the plot. Foamy, malodorous, excessive (DFME) category of diarrhea type variable and Giardia (PG) category of parasites under microscope variable were positioned together at the upper right and far from the origin. This means that Giardia parasites may cause foamy, malodorous, and excessive type of diarrhea. Looking at the lower right corner of the plot, it was seen that there was hemorrhagic mucous symptom (DHM) if the patient's leukocyte and erythrocyte counts were abundant. No abdominal pain (AN) and watery diarrhea type (DW) categories were placed nearby, so it can be said that abdominal pain may not always be seen with watery diarrhea. If there were fever (FY) and low or moderate leukocyte and erythrocyte counts (EM, LM, ER), the reason could be Entamoeba parasite (PE). If patients have suspicious food intake histories (SY), this may be considered as a possible suspicious situation about the blastocyte parasites (PB).

Overall, it was seen that symptoms set and laboratory tests set influenced each other especially through some variables. However, anamnesis set did not have any strong relation with others as seen in the graph where their categories were placed close to the origin.

Discussion

We applied and discussed OVERALS analysis by means of a medical data set. We defined anamnesis, symptoms, laboratory test characteristics, and then looked for relationships and similarities among these 3 sets. OVERALS has given us detailed information about the structure of the data, relationships among variable sets, relationships within variable sets, and categories of the variables. We had the chance to observe which parasites cause to which diarrhea type or which laboratory test result is related with which symptoms on the graphs. Relations among anamnesis, symptoms, and laboratory test variable set could be detected, so deterministic information about medical diagnosis and treatment could be obtained.

It was possible to investigate the data set from many aspects using the OVERALS solution. We visualized the associations and similarities between

the examined parameters set. It was seen that graphs could give better and easier understanding of the data structure. Graphical interpretation of the data could be a useful tool in an exploratory medical data or complex, large data, as obtained from an epidemiological research. Interpreting the results from a researcher's perspective, they could find inherent relationships among the examined variables,

and consequently design their policies in a more efficacious way.

Acknowledgement

We would like to extend our appreciation to Dr. Emine Kuru İnci who provided us with the data set.

References

1. Van der Heijden PGM, de Falguerolles A, de Leeuw J. A Combined Approach to Contingency Table Analysis Using Correspondence Analysis and Loglinear Analysis. *Applied Statistics* 1989; 38: 249-292.
2. Schiltz L, Schiltz J. Analysing the Relationship Between Traumatic Biographical Events and the Current Structural Functioning of Personality. *Bull. Soc. Sci. Med.* 2008; 1: 175-188.
3. Gifi A. *Nonlinear Multivariate Analysis*. New York: John Wiley & Sons; 1991.
4. Meulman JJ, Heiser WJ. SPSS Inc. SPSS Categories 11.0, 2001. Available from: URL: <http://www.siue.edu/IUR/SPSS/SPSS%20Categories%2011.0.pdf>
5. Michailidis G, De Leeuw J. The gif system of descriptive multivariate analysis. *Statistical Science* 1998; 13(4): 307-336.
6. SPSS: Statistical Package for the Social Sciences, Version 13.0, 2004. SPSS Inc. Chicago, IL.
7. Takane Y, Takane YO. Nonlinear Generalized Canonical Correlation Analysis by Neural Network Models. In: Nishisato S, Baba Y, Bozdogan H, & Kanefuji K, editors. *Measurement and Multivariate Analysis*. Tokyo: Springer Verlag; 2002. p. 183-190.
8. Bond J, Michailidis J. Homogeneity analysis in LISP-STAT. A brief introduction to homogeneity analysis, 1996. Available from: URL: <http://www.jstatsoft.org/v01/i02/PAPER/paper.html>
9. Kuru İE. Solid Organ Transplantasyon Hastalarında İshal Etkenleri. Specialization thesis. Ankara: Başkent University; 2005.
10. Andersen EB. *The statistical analysis of categorical data*. Germany: Springer Verlag; 1990.
11. Greenacre MJ. Practical Correspondence Analysis. In Barnett V, editors, *Interpreting Multivariate Data*. UK: John Wiley and Sons; 1981. p. 119-146.
12. Greenacre MJ, Underhill LG. Scaling a Data Matrix in a Low Dimensional Euclidean Space. In: Hawkins DM, editors. *Topics in Applied Multivariate Analysis*. USA: Cambridge University Press; 1982. p. 183-267.
13. Greenacre M, Hastie T. The geometric interpretation of correspondence analysis. *Journal of the American Statistical Association* 1987; 82: 437-447.
14. Greenacre, M. (1993). *Correspondence Analysis in Practice*. London: Academic Press.
15. Greenacre M. Correspondence Analysis and its Interpretation. In Greenacre M, Blasius J, editors. *Correspondence Analysis in the Social Sciences*. London: Academic Press, Harcourt Brace&Company Publishers; 1994. p. 3-22.
16. Blasius J, Greenacre M. Computation of correspondence analysis. In Greenacre M, Blasius J, editors. *Correspondence analysis in the social sciences* London: Academic Press 1994. p. 53-78.
17. Carroll JD, Green PE. An INDSCAL-Based approach to multiple correspondence analyses. *Journal of Marketing Research* 1988; 25: 193-203.
18. Cui Y, Zhou M, Wong WH. Integrated analysis of microarray data and gene function information. *OMICS* 2004; 8(2): 106-17.
19. De Leeuw J, Van der Heijden P. Correspondence analysis of incomplete contingency tables. *Psychometrika* 1988; 53(2): 223-233.
20. Escofier B, Pages J. Presentation of correspondence analysis and multiple correspondence analyses with the help of examples. In Devillers J, Karcher W, editors. *Applied Multivariate Analysis in SAR and Environmental Studies*. Netherlands: Kluwer Academic Pub; 1991. p. 1-31.
21. Gower JC. Fisher's optimal scores and multiple correspondence analysis. *Biometrics* 1990; 46: 947-961.
22. Greenacre MJ. Correspondence analysis of multivariate categorical data by weighted least-squares. *Biometrika* 1988; 75(3): 457-467.
23. Van der Heijden PGM, Mooijaart A, Takane Y. Correspondence Analysis and Contingency Table Models. In: Greenacre M, Blasius J, editors. *Correspondence Analysis in the Social Sciences* London: Academic Press, Harcourt Brace&Company Publishers; 1994. p. 79-111.
24. Van de Velden M. Dual Scaling and Correspondence Analysis of Rank Order Data. In: Heijmans RDH, Pollock DSG, Satorra A, editors. *Innovations in Multivariate Statistical Analysis*. Boston: Kluwer Academic Publishers; 2000. p. 87-97.
25. Van de Geer JP. *Algebra and Geometry of OVERALS*. University of Leiden: Department of Data Theory, Research Reports 1987; Available from: URL: http://www.datatheory.nl/pdfs/87/87_13.