Uludağ Üniversitesi
İktisadi ve İdari Bilimler Fakültesi Dergisi
Cilt XXI, Sayı 2, 2002, s. 279-294

QUANTITATIVE METHODS TO ANALYSE THE FACTORS ON THOUGHTS OF EMPLOYEES IN TOURISM SECTOR ABOUT THEIR SALARY

Durdu SERTKAYA [*]
Cem KADILAR [*]

Abstract

This paper aims to reveal the opinions of the employees about their earnings in the tourism sector of Turkey according to sex, education level, type of learning the career and age. For this reason, we review how to use the quantitative methods like correspondence, homogeneity, nonlinear principal components, and nonlinear canonical correlation analysis to analyse the survey data.

Keywords: *Correspondence, homogeneity, nonlinear principal components, nonlinear canonical correlation.*

Özet

Bu makale, Türkiye'nin turizm sektöründeki işçilerin, cinsiyete, eğitim düzeyine, işi öğrenme biçimine ve yaşa göre kazançları hakkındaki görüşlerini ortaya çıkarmayı amaçlamaktadır. Bu nedenle, anket verileri analizi için, karşılık getirme, homojenite, doğrusal olmayan temel bileşenler ve doğrusal olmayan kanonik korelasyon analizleri gibi nicel yöntemlerin nasıl kullanıldığı incelenmiştir.

---

[*] *Hacettepe University, Department of Statistics, Beytepe, Ankara-Turkey*

1. INTRODUCTION

Turkey, a country that sits on two continents, serves as a literal and figurative bridge between the east and the west. The history of Turkey dates back centuries. Many historians and archaeologists have flocked to this country to study the numerous civilizations of the world whose roots are found in Turkey. Turkey is surrounded by three seas: The Mediterranean in the south, the Aegean in the west and the Black Sea in the north. The two continents, Europe and Asia, are separated by the straits of Dardanelles and Bosporus. Turkey's climate provides four seasons. Moreover, the hospitality and friendliness of its present-day inhabitants complement the history and nature. Therefore, with all these characteristics, it is not surprising that millions of foreigners visit the country each year (Erdemgil, 2002). However, tourism sector has occasionally had economic ups and downs as a result of a series of events, and the employees in tourism sector unfortunately can never be happy about their earnings. This paper aims to reveal this fact by using the quantitative methods, such as correspondence, homogeneity, nonlinear principal components analysis and nonlinear canonical correlation.

Tourism sector has been a great importance in Turkish government policy for recent years. By this policy, the tourism revenue of Turkey has increased significantly. However, it is unfortunately true that the employees in tourism sector of Turkey generally have a poor living standard and we think that this will probably cause some serious problems for Turkish tourism in the near future. Therefore, in our opinion, to reveal this unhappiness of the employees scientifically is vital for the prospective development of tourism. In this respect, in the second section of this study we examine the quantitative analysis and in the third section, by this analysis, we clarify the conditions of the employees in tourism sector in detail.

2. QUANTITATIVE METHODS

2.1. Correspondence Analysis

Correspondence Analysis (CA) is a technique with which it is possible to find a multidimensional representation of the dependence between the row and column variable of a two-way contingency table (van der Heijden and de Leeuw, 1985: 50; de Leeuw and van der Heijden, 1988: 53; Pack and Jolliffe, 1992: 41). Besides, CA is an exploratory data analysis technique for the graphical display of contingency tables and multivariate categorical data.

In CA, numerical scores are assigned to the rows and columns of a data matrix so as to maximize their interrelationship. The scores are in corresponding units, allowing all the variables to be plotted in the same space for ease of interpretation. This representation then can be used to reveal the structure and patterns inherit in the data. The graphical display obtained from a CA can help in detecting structural relationships among the variable categories (Hoffman and Franke, 1986: 23).

Generally, the focus of interest in an analysis of the data is on determining whether the grouping categories (the rows of the table) can be sufficiently distinguished from each other on the basis of the observed characteristics (the columns of the table). Simple χ^2-analysis determines whether there is significant departure from independence between rows and columns (i.e. whether the rows can be distinguished from each other). In addition, CA provides a graphical display of the rows that highlights the differences among them by partitioning the χ^2-statistic appropriately.

In order to examine CA, consider the two-way table of counts $\mathbf{N} = (n_{ij})$ for $i = 1,\ldots, r$ and $j = 1,\ldots,c$, as the (rxc) matrix of the frequencies. Suppose that $n = \sum_i \sum_j n_{ij}$, $p_{ij} = n_{ij} / n$, $p_{i.} = \sum_j p_{ij}$ and $p_{.j} = \sum_i p_{ij}$. Lebart et al. (1984) defined the matrix of relative frequencies as $\mathbf{F} = \frac{1}{n}\mathbf{N}$. To test for independence between rows and columns of the table Krzanowski (1993) calculated the χ^2 test statistic

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - np_{i.}p_{.j})^2}{np_{i.}p_{.j}}.$$

Thus

$$\frac{\chi^2}{n} = \sum_i \sum_j \frac{(p_{ij} - p_{i.}p_{.j})^2}{p_{i.}p_{.j}} = \sum_i p_{i.} \left\{ \sum_j \frac{1}{p_{.j}} \left(\frac{p_{ij}}{p_{i.}} - p_{.j} \right)^2 \right\}$$

can be writen. Shortly,

$$\frac{\chi^2}{n} = \sum_i p_{i.} D_{i.}^2$$

where $D_{i.}^{2}$ is the so-called "χ^2-distance" between the ith row profile ($p_{ij}/p_{i.}$, j=1,…,c) and the average row profile ($p_{.j}$, j=1,..,c). Thus χ^2/n is a weighted sum of these χ^2-distances, the weights $p_{i.}$ reflecting the relative frequencies in the rows of the table.

The distance between two rows i and i' is given by

$$D^2(i,i') = \sum_{j=1}^{c} \frac{1}{p_{.j}} \left(\frac{p_{ij}}{p_{i.}} - \frac{p_{i'j}}{p_{i'.}} \right)^2 .$$

In symmetric fashion, the distance between j and j' is also written as

$$D^2(j,j') = \sum_{i=1}^{r} \frac{1}{p_{i.}} \left(\frac{p_{ij}}{p_{.j}} - \frac{p_{ij'}}{p_{.j'}} \right)^2 .$$

The distance thus defined is called the chi-square distance (Lebart et al., 1984).

Clouds of points can be interpreted using chi-squared distances: When two row points (or two columns points) are near each other, their profiles are similar. When profiles differ considerably, the distance between the points is large. The profiles of the marginal frequencies of X are projected into the orijin. When the distance of a category point to the orijin is small, the profile of this category point does not differ much from the mean profile (van der Heijden and de Leeuw, 1985:50).

2.2. Homogeneity Analysis

Tenenhaus and Young (1985) analysed a variety of methods for quantifying categorical multivariate data, and showed that they all led to the same equations for analyzing the same data. Because of the widely varying countries and languages in which these methods have been proposed, there are a wide variety of names for homogeneity analysis (HA) like Optimal Scaling, Optimal Scoring and Appropriate Scoring methods, Dual Scaling, Multiple Correpondence Analysis, Scalogram Analysis, and Quantification Method.

HA can be seen as a way of analyzing a subject by variable matrix with categorical variables; or a subject by item matrix of multiple-choice data; or a multi-way contingency table. In all cases HA scales the subjects and categories (items; levels of each way of the table). The scaling is multidimensional, since several scale values are obtained for each subject and category.

HA is a method to maximize the homogeneity of a number of variables. HA determines quantifications or transformations of the categories of each of the variables such that homogeneity is maximized. Using the vector $\mathbf{q_s}$, with t_s elements, the expression $\mathbf{Z}_s\mathbf{q}_s$ represents a single quantification or transformation of the n objects induced by variable i. Note that objects in the same category get the same quantification. In HA, there are d-dimensional quantifications for each variable. Therefore, $\mathbf{Q_s}$ matrix, with $t_s$xd elements, is called multiple nominal quantifications of variable i. Then the matrices $\mathbf{Z}_s\mathbf{Q}_s$ induce k multiple quantifications of the objects. Perfect homogeneity is defined if all multiple quantifications of the objects are the same: $\mathbf{X} = \mathbf{Z}_1\mathbf{Q}_1 = \mathbf{Z}_2\mathbf{Q}_2 = \ldots = \mathbf{Z}_S\mathbf{Q}_S$. HA minimizes the loss of homogeneity by

$$\min \sigma(X,Q) = \sum_{i=1}^{k} SSQ\left(X - Z_i Q_i\right) \qquad (2.1)$$

where $\mathbf{X'X} = n\mathbf{I}$ and $\mathbf{u'X} = \mathbf{0}$. Here $\mathbf{u}$ is a column with n elements equal to 1; SSQ(.) is used for the sum of squares of the elements of a matrix. Elements of $\mathbf{X}$ are called object scores (van der Burg and de Leeuw, 1988:53).

For example, in a survey application the questions are the variables. A single question s consists of a set t_s of response categories. The total number of response categories, t, contained in the questionnaire is

$$t = \sum_{s=1}^{S} t_s$$

On the other hand, the number of individuals who responded to questions are objects for HA.

Lebart, Morineau and Warwick (1984) denoted by **Z** the matrix with r rows and c columns describing the response of the n individuals with binary coding.

Matrix **Z** is the juxtaposition of S submatrices:

$$\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_s, \ldots, \mathbf{Z}_S]$$

Submatrix $\mathbf{Z}_s$ (with r rows and t_s columns) is such that its ith row contains t_s-1 times the value zero, and once the value 1, in the column corresponding to the category of question s chosen by subject i.

The square matrix

$$\mathbf{B} = \mathbf{Z'Z} \tag{2.2}$$

is called Burt's contingency table associated with **Z**, the matrix of responses (Lebart et al., 1984; van der Burg and Leeuw, 1988:53; Greenacre, 1988:75). Matrix **B** is made up of S^2 blocks.

In HA we know that the dimensionality d is bigger than 2. Each dimension adds another quantification of the categories of each variable, and the different quantifications of the same variable have no simple relation to each other. This makes interpretation complicated so rank-one restrictions are used in homogeneity analysis. Mathematically this restriction for variable s can be expressed as

$$Q_s = h_s a_s^{'} \tag{2.3}$$

where $\mathbf{h}_s$ is the t_s-vector of single category quantifications, and $\mathbf{a}_s$ is the d-vector of weights (van der Burg and de Leeuw, 1988:53).

As a result a possible algorithm for multiple quantification is as follows:

1. Start with some initial guess of category quantifications (but take care that they have zero mean)
2. Calculate object scores as the average of quantifications of categories that apply to the object $\left(\tilde{X} \leftarrow Z_s \tilde{Q}_s / t_s\right)$
3. Standardize these object scores $\left(X \leftarrow \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1/2}\right)$
4. Calculate updated category quantifications by taking the average of objects within the same category $\left(Q_s \leftarrow Z_s^{'} X\right)$
5. Go back to Step 2 and repeat the procedure until it converges (Go back to Step 2, setting $\tilde{Q}_s \leftarrow Q_s$, as long as the values of **X** and $\mathbf{Q}_s$ are not sufficiently stabilized according to some preselected criterion of accuracy) (van de Geer,1993; Gifi,1990).

In the end of homogeneity analysis, category quantification computed by (2.3) is demonstrated as a point for each category on the plot. There are some issues to follow up while interpreting this plot:

1. The length of the distance between the point and the origin indicates the importance of the category.
2. The direction of a point with respect to the other points indicates the sign of the correlation between that category and the others. The opposite direction means negative correlation, otherwise there is a positive correlation.
3. When lines are drawn from each point to origin, the angles between the paired lines indicate correlations between related two categories. When the angle gets bigger (smaller), the correlation will be lower (higher) (Palmer, 1993:74).

2.3. Nonlinear Principal Components Analysis

In the field of computer vision, principle component analysis is often used to provide statistical models of shape, deformation or appearance. This simple statistical model provides a constrained, compact approach to model based vision. However, as larger problems are considered, high dimensionalty and nonlinearity make linear PCA an unsuitable and unreliable approach (Bowden et al., 1997:33).

The principal components analysis of an nxs data matrix **H** can be formulated in terms of the loss function

$$\sigma_j(\mathbf{X}, \mathbf{A}) \equiv \mathrm{SSQ}(\mathbf{H} - \mathbf{XA}') \tag{2.4}$$

where **X** is nxd of rank d, **A** is sxd and the elements of **H** comprise the sum-total of possible responses by the subjects. Gifi (1990) used the symbol σ_j(join loss) because σ_j(**X**, **A**)=0 for some **X** and **A** means, that the join rank of **H** is less than or equal to d.

The linear model to a rectangular data matrix **H** is defined by

$$q_{ij} = \sum_{l} x_{il} a_{jl} ,$$

$$h_{ij} > h_{kj} \rightarrow q_{ij} \geq q_{kj},$$

where i, k=1,…,n and j=1,…,S and l=1,…,d.

The loss function is,

$$\sigma_j(\mathbf{Q},\mathbf{X},\mathbf{A}) \equiv S^{-1}\sum_j SSQ(\mathbf{q}_j - \mathbf{X}\mathbf{a}_j) \tag{2.5}$$

which is minimized over all **X**, **A** and over all $\mathbf{q}_j \in C_j$ (j=1,…,S). The vector $\mathbf{a}_j$ has d elements, and there are s such vectors. Note that $\mathbf{a}_j$ is row j of **A**, written as a column (Gifi, 1990).

Gifi (1990) assumed that data are categorical and that the number of categories of a variable is usually much less than the number of observations. In the option of discrete ordinal data, it is more convenient to write σ_j as a function of the category quantifications $\mathbf{y}_j$. Thus the formula (2.5) becomes

$$\sigma_j(\mathbf{Y},\mathbf{X},\mathbf{A}) \equiv S^{-1}\sum_j SSQ(\mathbf{G}_j\mathbf{y}_j - \mathbf{X}\mathbf{a}_j) \tag{2.6}$$

which must be minimized under the conditions

$\mathbf{u'}\mathbf{G}_j\mathbf{y}_j=0$

$\mathbf{y}_j'\mathbf{D}_j\mathbf{y}_j=1$

$\mathbf{y}_j \in C_j$

where C_j is now a cone in t_j-dimensional space, usually t_j<<n.

The conditions for minimum loss for each variable have interesting geometrical interpretations. Object scores are represented as points in d-dimensional space. If a variable is multiple nominal, its loss contribution is

$$tr(\mathbf{X}\text{-}\mathbf{G}j\mathbf{Y}j)'\mathbf{M}j\,(\mathbf{X}\text{-}\mathbf{G}_j\mathbf{Y}_j)$$

which vanishes if and only if $\mathbf{M}_j\mathbf{X}_j=\mathbf{G}_j\mathbf{Y}_j$ which is then of course equal to the corresponding category qualification. We expect object in a category to be close together.

If a variable is multiple ordinal its loss contribution is

$$tr(\mathbf{X}\text{-}\mathbf{G}_j\widetilde{\mathbf{Y}}_j)'\mathbf{M}_j\,(\mathbf{X}\text{-}\mathbf{G}_j\widetilde{\mathbf{Y}}_j)+tr\,(\mathbf{Y}_j\text{-}\widetilde{\mathbf{Y}}_j)'\mathbf{D}_j\,(\mathbf{Y}_j\text{-}\widetilde{\mathbf{Y}}_j)$$

where $\widetilde{\mathbf{Y}}_j=\mathbf{D}_j^{-1}\mathbf{G}_j'\mathbf{X}$.

Observe that for multiple ordinal data there is generally no complete freedom of rotation of the axes. For single variables there are at least two loss components for each variable. For single ordinal and numerical variables we want the category quantifications on a line through the orijin and we want them to be on this line in the correct way (Gifi, 1990).

2.4. Nonlinear Canonical Correlation Analysis

Nonlinear canonical correlation analysis (CCA) is used to measure of association between two random variables that are symmetric nondecreasing functions of the canonical coefficients of the random variables.

Gifi (1990) studied the situation where two sets of variables, having a symmetric role are involved. Consider

$$\sigma_S(\mathbf{X},\mathbf{Y}) = t^{-1}\sum_t \mathrm{SSQ}(\mathbf{X} - \mathbf{G}^t\mathbf{Y}^t) \qquad (2.7)$$

which must be minimized under the condition $\mathbf{X'X}$ = I. The restrictions on Y^t are such that Y^tT_t is feasible whenever Y^t is feasible, no matter how T_k is chosen.

Dauxois and Nkiet (1998) proposed a class of measures of association between random variables with values in any measurable spaces, constructed using symmetric nondecreasing functions and the canonical coefficients derived from nonlinear canonical analysis.

In nonlinear caconical correlation analysis, Hsieh (2001) followed the same procedure as in CCA, except that the linear mappings are replaced by nonlinear mapping functions.

2.5. The Comparison among the Quantitative Methods

OVERALS is a technique for canonical correlation analysis with two or more sets of variables. Any three way table can be used as input for the OVERALS program. In OVERALS terminology the ways are called objects, variables and sets. Three measurement levels of the data can be handled as numerical, ordinal and nominal. They can be defined for each variable separately. OVERALS is considered as the most general model in the so-called Gifi-system of nonlinear multivariate analysis. The models that can be dealt with by OVERALS are PRINCALS (nonlinear principal component analysis), ANACOR (correspondence analysis), HOMALS (homogeneity analysis). In addition, the CORALS model (nonlinear canonical correlation analysis) is a special case of OVERALS. If linear teachniques are considered, linear PCA is a special case of PRINCALS and consequently of OVERALS.

When all the sets contain only one variable, we are dealing with a two-way table. Then we deal with PCA. A nonlinear version of PCA is called PRINCALS. Thus OVERALS with one variable per set is PRINCALS. If only numerical transformations are used it is linear PCA.

If we restrict not only the number of variables per set to one, but also the transformations to multiple nominal, we should use homogeneity analysis or HOMALS. If, in addition, the number of variables is restricted to two, OVERALS is similar to correspondence analysis or ANACOR.

CORALS is described as a model for canonical correlation analysis with optimal scaling. OVERALS with two sets of variables is similar to CORALS. CANALS, an alternative model for the CORALS, is a nonlinear generalization of ordinary canonical correlation analysis for two sets of variables (van der Burg et al., 1994: 18).

3. ANAYSIS OF SURVEY DATA

The data was obtained from a survey study whose variables and categories are given in Table 1. A labour union applied this study to 1264 workers in tourism sector in order to understand the feeling of the employees about their salary. We interpreted the results of this survey study by using quantitative methods.

Table 1. Variables from the survey study

SEX	(1) female, (2) male
AGE	(1) age <=25, (2) age 26 – 35, (3) age >=36
EDUCATION	(1) primary, (2) secondary, (3) lycee, (4) university
TYPE of LEARNING	(1) self learning, (2) course, (3) school, (4) others
SALARY	(1) unsatisfactory, (2) average, (3) satisfactory

Table 2 displays the frequencies and the percentages of 1264 employees according to their sexes, ages, education levels, types of learning the career and thoughts about salary. From Table 2, we understand that the variables of sex, type of learning the career and salary have nominal data, on the other hand, the variables of age and education have ordinal data.

We consider only two-dimensional solutions in order to interpret the results by the help of the plots.

Table 2. The frequencies and the percent of the variables

Variables	Frequency	Percent (%)
SEX		
Female	273	21.8
Male	981	78.2
AGE		
<= 25	501	42.0
26 – 35	519	43.5
>= 36	173	14.5
EDUCATION		
Primary	216	17.1
Secondary	234	18.6
Lycee	581	46.1
University	230	18.2
TYPE of LEARNING		
Self learning	656	53.0
Course	222	17.9
School	285	23.0
Others	75	6.1
SALARY		
Unsatisfactory	690	55.4
Average	386	31.0
Satisfactory	169	13.6

Table 3. The eigenvalues of quantitative methods

Eigenvalue	CA	HA	Nonlinear PCA	Nonlinear CCA
λ_1	0.47	0.41	0.34	0.35
λ_2	0.21	0.36	0.25	0.44
Total	0.68	0.77	0.59	0.79

When there are two variables, correspondence analysis can be used. Therefore, let take the variables sex and salary to examine by this analysis. When this analysis is applied to these variables, two-dimensional solutions account for 68% of the total inertia as shown in Table 3. From Figure 1, we can say that the female employees think that their salary is average, on the other hand, the male employees think that their salary is unsatisfactory. In addition, it is clearly seen that none of the employees think their salary is satisfactory.

When we take the variables sex, type of learning the career and salary, we use homogeneity analysis because these variables have nominal data. When all variables have nominal data, homogeneity analysis is used. When this analysis is applied to these variables, from Table 3 we see that two-dimensional solutions account for 77% of the total variance. From Figure 2, we can say that the males and the employees who learned the career by themselves think their salary is unsatisfactory and that the employees learned their career from course or school think their salary is average. Like CA, we again see that none of the employees is satisfied with their earnings.

Nonlinear principal component method can be used when the variables have both nominal and ordinal data. Therefore, we take all variables in Table 1. When this method is applied, two-dimensional solutions account for 59% of the total inertia in the data. From Figure 3, we observe that the female employees, self learning and the employees whose ages are lower than 25 think their salary is unsatisfactory. On the other hand, the employees graduated from lycee (high school), males, the employees who learned their career from course, and the employees whose ages are 26-35 think their salary is average. By applying this method, we again see that none of the employees is satisfied with their salary.

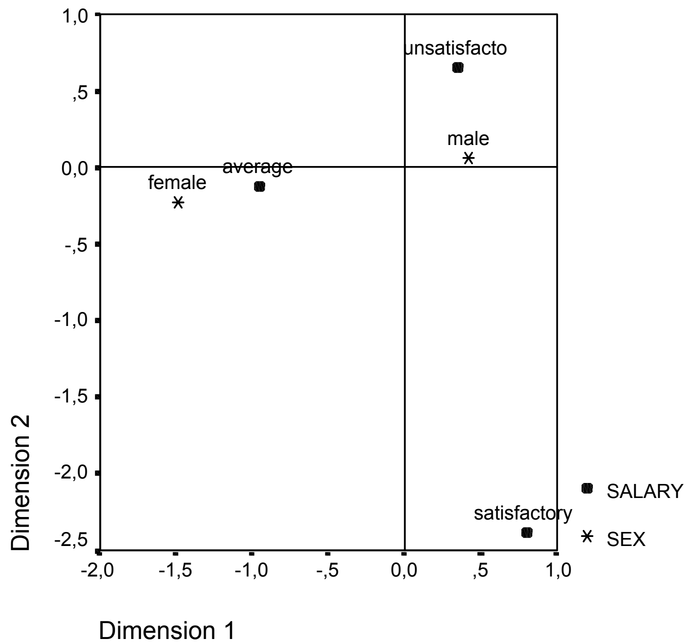

Figure 1. *The plot of correspondence analysis*

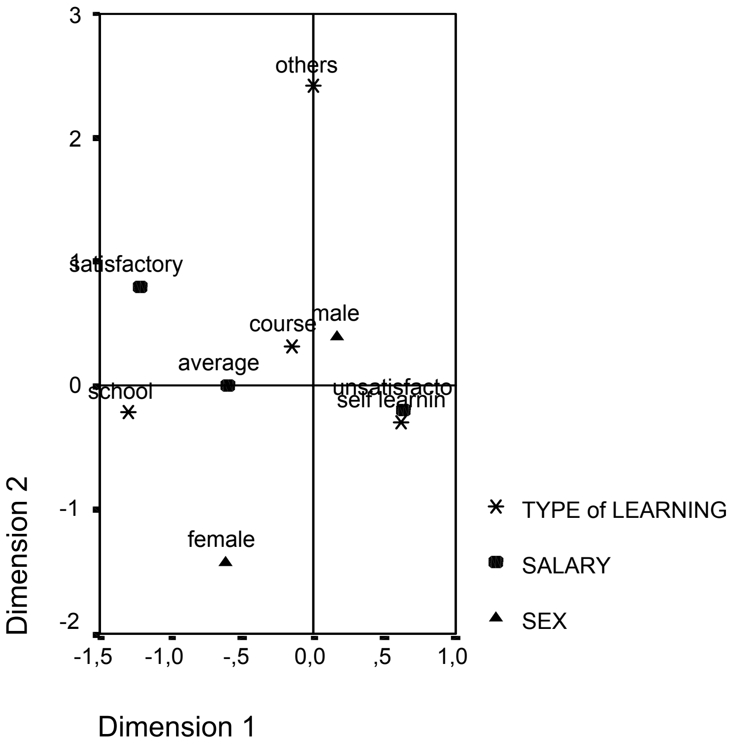

Figure 2. *The plot of homogeneity analysis*

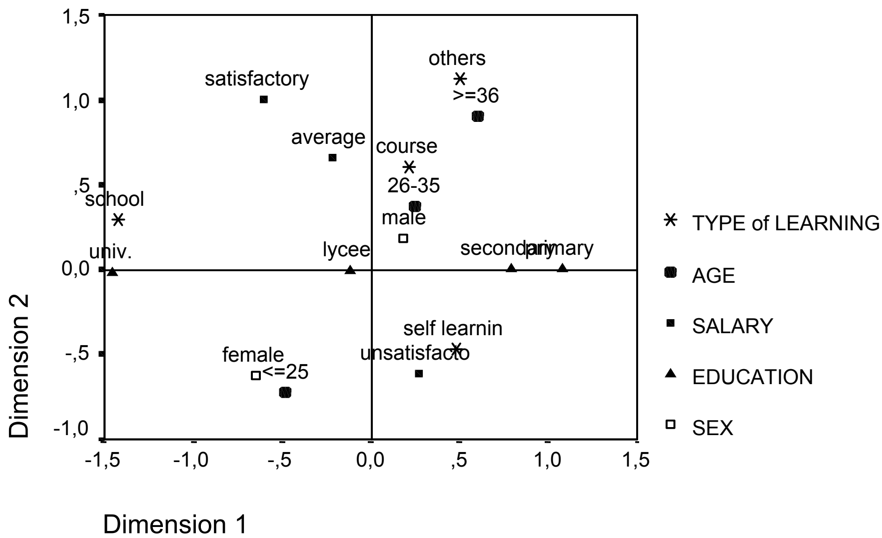

Figure 3. *The plot of nonlinear principal component analysis*

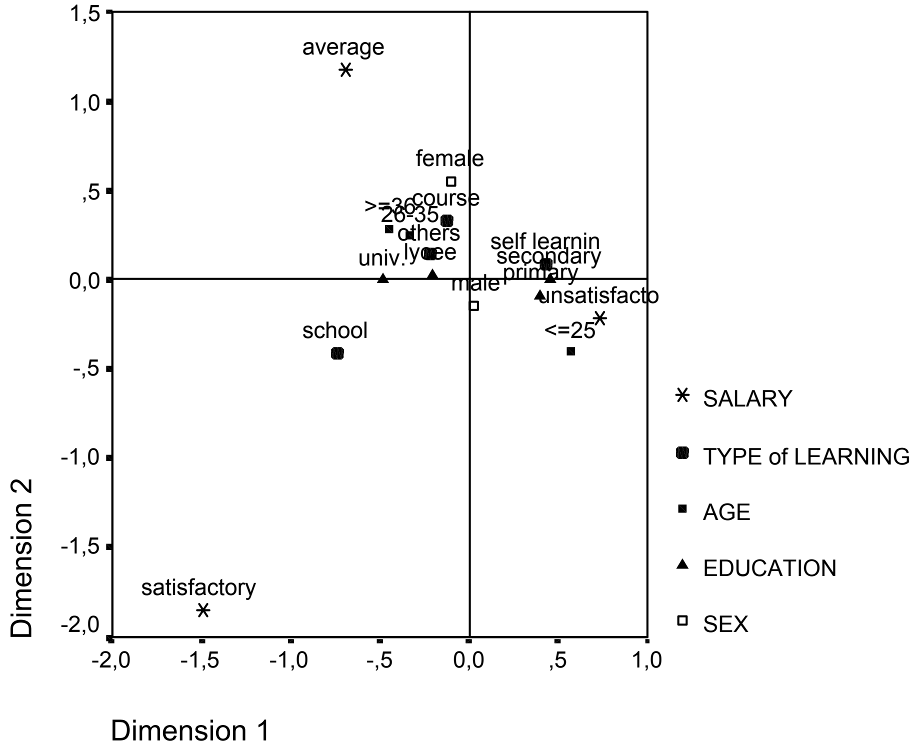

Figure 4. *The plot of nonlinear canonical correlation analysis*

Nonlinear Canonical Correlation Analysis can be used when the variables are grouped in more than one set. Therefore, let group age, education level, type of learning and sex in a set, which gives the personal identification, and salary, our interested variable, in another set. We obtain Figure 4 by this method. Figure 4 displays that the employees graduated from primary or secondary school, males, the employees who learned their career by themselves, and the employees whose ages are lower than 25 think their salary is unsatisfactory. In addition, we observe that the employees graduated from lycee or university, who are older than 25, learned their career from course or in other way and female employees think their salary is average. When we examine Figure 4 further, we again see that none of the employees is satisfied with their salary.

4. CONCLUSION

In this paper, by using CA, we find that the salary is thought as average by the female employees and as unsatisfactory by the male employees. By applying HA, we see that the salary is thought as average by the employees learning the career from course or school, on the other hand the salary is thought as unsatisfactory by the employees learned the career by themselves. By nonlinear PCA method, we understand that the salary is thought as unsatisfactory by the employees whose ages are lower than 25 and learned the career by themselves, however, the salary is thought as average by the employees whose ages are between 26-35 and graduated from high school. By nonlinear CCA method, we infer that the salary is thought as average by the employees whose ages are older than 25 and graduated from university, on the contrary, the salary is thought as unsatisfactory by the employees graduated from primary or secondary school. From all this quantitative methods we strongly conclude that employees are unsatisfied with their salary.

Turkey is now among a group of countries that are accepted as the top tourism earners in the world. In addition to its historical and cultural inheritance, favourable climatic conditions and abundance of natural resources, the Turkish tourism sector also offers a qualified and reasonably priced tourism product. With scores of tour guides and travel agents, tourism promotion is done professionally. The tourism sector tries hard and seeks new initatives to ensure its continuity. Despite of these positive improvements in tourism sector in Turkey, in this paper, by applying the quantitative methods to survey data, it is clearly shown that there is a tragic problem about the salary of the employees in tourism sector in Turkey. In order to succeed on this problem, public or private institutions should take a precaution as soon as possible.

REFERENCES

Bowden, R., Mitchell, T.A. and Sarhadi, M. (1997), "Cluster Based Nonlinear Principle Component Analysis", *Electronics Letters,* 33 (22), 1858-1859.

Dauxois, J. and Nkiet, G.M. (1998), "Nonlinear Canonical Analysis and Independence Tests", *The Annals of Statistics,* 26 (4), 1254-1278.

de Leeuw, J. and van der Heijden, P.G.M. (1988), "Correspondence Analysis of Incomplete Contingency Tables", *Psychometrika,* 53 (2), 223-233.

Erdemgil, R.D. (2002), An Analysis of Turkey's Tourism Sector and The Impact of Externalities, Working Paper, George Mason University.

Gifi, A. (1990), *Nonlinear Multivariate Analysis,* John Wiley and Sons.

Greenacre, M.J. (1988), "Correspondence Analysis of Multivariate Categorical Data by Weighted Least-Squares", *Biometrika,* 75 (3), 457-467.

Hsieh, W.W. (2001), "Nonlinear Canonical Correlation Analysis of the Tropical Pacific Climate Variability Using a Neural Network Approach", *Journal of Climate,* 14, 2528-2539.

Hoffman, D.L. and Franke, G.R. (1986), "Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research", *Journal of Marketing Research,* 23, 213-227.

Krzanowski, W.J. (1993), "Attribute Selection in Correspondence Analysis of Incidence Matrices", *Appl. Statist.,* 42 (3), 529-541.

Lebart, L., Morineau, A. and Warwick, K.M. (1984), *Multivariate Descriptive Statistical Analysis,* John Wiley and Sons.

Pack, P. and Jolliffe, I.T. (1992), "Influence in Correspondence Analysis", *Appl. Statist.,* 41 (2), 365-380.

Palmer, M.W. (1993), "Putting Things in Even Better Order: The Advantages of Canonical Correspondence Analysis", *Ecology,* 74 (8), 2215-2230.

Tenenhaus, M. and Young, F.W. (1985), "An Analysis and Synthesis of Multiple Correspondence Analysis, Optimal Scaling, Dual Scaling, Homogeneity Analysis and Other Methods For Quantifying Categorical Multivariate Data", *Psychometrika,* 50 (1), 91-119.

van der Heijden, P.G.M. and de Leeuw, J. (1985), "Correswpondence Analysis Used Complementary to Loglinear Analysis", *Psychometrika,* 50, 429-447.

van der Burg, E. and de Leeuw, J. (1988), "Homogeneity Analysis with k Sets of Variables: An Alternating Least Squares Method with Optimal Scaling Features", *Psychometrika,* 53 (2), 177-197.

van der Burg, E. and de Leeuw, J. and Dijksterhuis, G. (1994), "OVERALS: Nonlinear Canonical Correlation with k Sets of Variables", *Computational Statistics and Data Analysis,* 18 (1), 141-163.

van de Geer, J.P. (1993), *Multivariate Analysis of Categorical Data: Theory,* Sage Publications, Inc.