

# Qualitative Data Analysis for Health Care Services

Taner Ersoz, Filiz Ersoz

**Abstract**—This study was designed enable application of multivariate technique in the interpretation of categorical data for measuring health care services satisfaction in Turkey. The data was collected from a total of 17726 respondents. The establishment of the sample group and collection of the data were carried out by a joint team from The Ministry of Health and Turkish Statistical Institute (Turk Stat) of Turkey. The multiple correspondence analysis (MCA) was used on the data of 2882 respondents who answered the questionnaire in full. The multiple correspondence analysis indicated that, in the evaluation of health services females, public employees, younger and more highly educated individuals were more concerned and complainant than males, private sector employees, older and less educated individuals. Overall 53 % of the respondents were pleased with the improvements in health care services in the past three years. This study demonstrates the public consciousness in health services and health care satisfaction in Turkey. It was found that most the respondents were pleased with the improvements in health care services over the past three years. Awareness of health service quality increases with education levels. Older individuals and males would appear to have lower expectancies in health services.

**Keywords**—Multiple correspondence analysis, optimal scaling, multivariate categorical data, health care services, health satisfaction survey, statistical visualizing, Turkey.

## I. INTRODUCTION

THE importance of the influence of public opinion on the health care provision has been acknowledge [1] regardless of a country's level of development [2], [3]. Public services and their responses have been taken as the most reliable indicator for the development of action plans and the improvement of the services [4], [5]. Health care systems have been improved in many countries for the past decade, and also needed to monitor the performance of the new systems widely recognized.

Service offering and service recipients' satisfaction in addition to the resources such as physical infrastructure, tools, equipment and manpower are determining factors in the evaluation of the systems. Faye et al. [6] discussed patient satisfaction and suggested that this should be considered a key determinant for strategic considerations in setting up health care service policies. Similar views and suggestions were reported by Fitzpatrick [7] and John [8]. It is understandable that patients are the direct objects of the efficiency of health care services and their opinions would reflect the value of the system.

A variety of statistical methods have been used to assess patient preference for the efficiency of health care systems [9].

Taner Ersoz is working Department of Actuarial and Risk Management at Karabuk University, Turkey (e-mail: tanererso@karabuk.edu.tr)

Filiz Ersoz is working Department of Industrial Engineering at Karabuk University, Turkey (e-mail: ferso@karabuk.edu.tr).

In this study, Multivariate Data Analysis techniques are used to summarize the original data and present relevant information in a graphic form to facilitate interpretation. The multiple correspondence analysis has been widely used by researchers because of its practicality in interpreting the survey data. This technique has also been used to establish optimal scale values for the levels of a categorical variable, which give quantitative meaning to the categories. Correspondence analysis shows the existence of relationship and how the variables are related. The purpose of multiple correspondence analysis, also known as homogeneity analysis, is to find quantifications that are optimal in the sense that the categories are distinct from each other as possible. This implies that objects in the same category are plotted close to each other and objects in different categories are plotted as far apart as possible.

The correspondence analysis starts with two-way cross classifications. The variables must be discrete, nominal, ordinal or continuous variables segmented into ranges. The technique defines a measure of distance between any two points, where the points are the values of the discrete variable.

The solution of multiple correspondence analysis is indicated by a set of eigenvalues. Each eigenvalue indicates the proportion of the total variance of the transformed variables that is explained by the consecutive dimensions of the solution. Therefore, an eigenvalue indicates the importance of a dimension for the solution, where importance denotes the ability of the dimension to discriminate between the respondents in terms of the quantified categories, or- from another perspective between the categories in terms of the quantified respondents [10].

Correspondence analysis is one of the multivariate graphical methods. Inertia in correspondence analysis is defined as the total mass Pearson Chi-square for the two-way divided by the total sum.

The map also has points to represent the columns. There are different ways of representing the columns jointly with the rows. In this map, the column profiles have been analyzed in the exactly way as described, as if the matrix were transposed and the whole process repeated in a symmetric fashion, leading to the principal coordinates of the columns. The rows and columns are then jointly plotted with respect to the same axes, both in principal coordinates. The merits and demerits of this joint display have been discussed in many texts [11], [12]. An alternative approach, called joint correspondence analysis, is shown to be a more natural generalization of the simple case.

The purpose of this paper is to present a tool which may better serve to evaluate the results obtained in the public opinion survey on the health care services in Turkey eventually. It may contribute to the policy making efforts for

the improvement of the health care services system in the country.

## II. MATERIAL AND METHODS

The data used in this study was obtained from the Health Services Satisfaction Survey conducted in Turkey in 2006 [13]. The survey was carried out with the cooperation of the Ministry of Health and TURKSTAT. Health survey was realized with the aim of providing indicators in health area not obtaining by administrative registration system and also constitute data source for the decision-makers and researchers. The satisfaction survey is designed to produce proper estimates for urban and rural in Turkey. The Sample size was found as 5253 for household survey. Total number of addresses was 3466 households for urban area, 1787 for rural area. It was used the strata and two-phase cluster sampling method. The differences of rural-urban was found settlements with a population. For having a population over 20 001 was regarded as urban, for the population of less than 20,000 and 20000 was regarded as rural.

A total of 17726 respondents all over 15 years of age were involved in the survey. Multiple Correspondence Analysis (MCA) was used on the data of 2882 respondents who answered all sections of the questionnaires. In the MCA analyses, 7 headings of variables were used as active variables. The variables were: gender, age, literacy, employment, work status and adequacy of services (health status and the improvements in health care in the last three years).

The data was statistically analyzed in a set of descriptive and multivariate statistics using IBM Statistical Package for the Social Sciences (SPSS) version 16.0 (2008). IBM SPSS categories display relationships among nominal variables in a perceptual map, the visual presentation shows the relationships among the categories of the variables, such as, health, social, economic, political or biological subjects etc.

MCA, also called homogeneity analysis, is a standard multivariate analysis technique widely used in partitioning into clusters with the same or similar categories.

The merits of correspondence analysis are ease of application and the portrayal of categorical data in a joint-space enabling comparison of both row and column relationships in terms of a squared distance model and allowing the examination of relationships between categories of nominal data in a contingency table (16-20). This analysis is an analog of principal components analysis for frequency data, designed to display the association among categorical variables in a small number of dimensions, designed to account for the largest proportion of the Pearson  $\chi^2$ , divided by the total sample size  $n$ . The distance used in the graphical display of the rows (and columns) of the table is the so-called chi-square distance between the profiles of rows (and columns). The method enables a researcher to scale nominal answers in multiple choice tests [14].

For a two-way table the scores for the row categories, namely  $X_{im}$ , and column categories,  $Y_{jm}$ , on dimension  $m = 1$ ,

...,  $M$  are derived from a singular value decomposition of residuals from independence, expressed as  $d_{ij}/\sqrt{n}$ , to account for the largest proportion of the  $\chi^2$  in a small number of dimensions. This decomposition may be expressed as;

$$\frac{d_{ij}}{\sqrt{n}} = \sum_{m=1}^M \lambda_m \chi_{im} y_{jm} \quad (1)$$

where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$  and  $M = \min(i_1, j_1)$ . In  $M$  dimensions, the decomposition [1] is exact. A rank  $d$  approximation in  $d$  dimensions is obtained from the first  $d$  terms on the right side of (1), and the proportion of  $\chi^2$  accounted for by this approximation is;

$$\sum_m^d N\% = \frac{\lambda_{m_2} \%}{\chi^2} \quad (2)$$

Thus, correspondence analysis is designed to show how you evaluate the improvements in the health care services in Turkey.

## III. RESULTS

In this research, multiple correspondence analysis was applied to the polling results of 17726 respondents in SPSS. However, after omitting incomplete data, the number of respondents was reduced to 12842. Details of respondents who completed is given Table I. The statistics in the last variable were the responses to the question "How do you evaluate the improvements in the health care services in the last three years?" The answers were also depicted as a Table II and a bar chart (Fig. 1). MCA was used on the data of 2882 respondents who answered all the questions in the questionnaires (Table III).

According to the evaluation of healthcare services, it was found that 52.99 % of the respondents the rated the evaluation positively, expressing satisfaction with the improvements in health care services in the past three years. The results are given in Fig. 1.

The correspondence analysis graphically portrays similarities and differences among the health services indicators. The correspondence map is based on the chi-squared statistics, which measure deviations from the expected values. Fundamental data in MCA are the total inertia of the data matrix and the percentages of the total inertia covered by each of the axes. The inertia is the chi-squared statistic divided by the grand total of all cell entries in the table. Total of 2882 respondents answered all the questions and their data were used in the MCA. Six headings were used as active variables in the MCA, gender, age, literacy, employment, work status, and adequacy of services. The dimensions are in descending order of eigenvalue. The mean Cronbach's Alpha is based on the mean Eigenvalue. Two dimensions with significant Chi-square value ( $\chi^2 = 48.402$ ;  $p < 0.01$ ) are given in Table III. The relative contribution of each dimension was shown as % variance with the total cumulative explained variance being

equal to 100.00%. The contribution of dimension 1 accounts for 37 % of the variance, while dimension 2 accounts for 29 %

of the variance. The two dimensions together account for 66 % of the total variance.

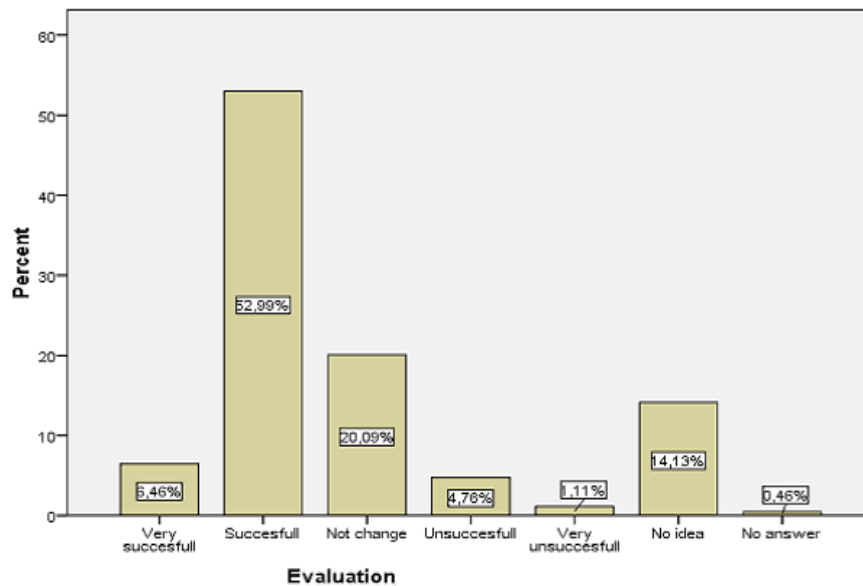


Fig. 1 Evaluation of healthcare services in the last 3 years

TABLE I  
 NUMBERS AND PERCENTAGES OF RESPONDENTS IN VARIABLES AND CATEGORIES

Variable	Categories	Count	%
Gender	Male	8485	47,9
	Female	9241	52,1
	<b>Total</b>	<b>17726</b>	<b>100,0</b>
Age	15-24	2401	19,1
	25-34	2798	22,3
	35-44	2590	20,6
	45-54	2069	16,5
	55-64	1346	10,7
	65-74	970	7,7
	75+	398	3,2
	<b>Total</b>	<b>12572</b>	<b>100,0</b>
Employment	Public	685	15,0
	Private	3878	85,0
	<b>Total</b>	<b>4563</b>	<b>100,0</b>
Literacy	Literate	14015	88,1
	Illiterate	1886	11,9
	<b>Total</b>	<b>15901</b>	<b>100,0</b>
Work Status	Payroll employees	2354	51,6
	Daily workers	423	9,3
	Employer	213	4,7
	Self employed	1070	23,4
	Unemployed	503	11,0
	<b>Total</b>	<b>4563</b>	<b>100,0</b>
Evaluation of Healthcare Services	Very successful	830	6,5
	Successful	6805	53,0
	No change	2580	20,1
	Unsuccessful	611	4,8
	Very unsuccessful	143	1,1
	No idea	1814	14,1
	No answer	59	0,5
<b>Total</b>	<b>12842</b>	<b>100,0</b>	

TABLE II  
 CASE PROCESSING SUMMARY OF MULTIPLE CORRESPONDENCE ANALYSIS (MCA)

Cases	Count
Valid Active Cases	2882
Active Cases with Missing Values	14844
<b>Cases Used in Analysis</b>	<b>17726</b>

TABLE III  
 MODEL SUMMARY

Dimension	Cronbach's Alpha	Variance Accounted for	
		Total (Eigenvalue)*	Inertia
1	0,757	2,959	0,370
2	0,651	2,324	0,291
<b>Total</b>		<b>5,283</b>	<b>0,660</b>

The joint plot figure (Fig. 2) was derived using the values of the coordinates in Table III. Each labeling variable produces a separate plot labeled with the values of that variable. The distance from an object to the origin reflects variation from the "average" response pattern. This average response pattern corresponds to the most frequent category for each variable. Objects with many characteristics corresponding to the most frequent categories lie near the origin.

Dimension 1 (the horizontal axis) discriminates the illiterate and those in the 65-74 age range by gender. The first dimension also separates those in the 74 + age range from others. On the other hand, expectations from health care services may vary according to age and gender of the respondents. The second dimension (the vertical axis) seems to separate female and hygiene from all other factors.

The analysis indicated that females are more complainant than males about the public health services of hospital bed, doctor, medical instruments, hygiene and nurse. Also, literates

are more complainant than illiterates; public employees more than private sector employees; young and middle-aged people more than older people.

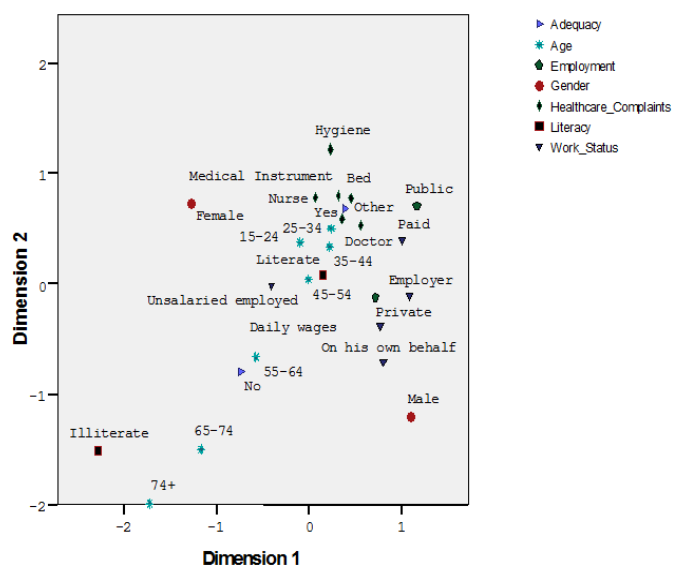


Fig. 2 Joint plot of category point

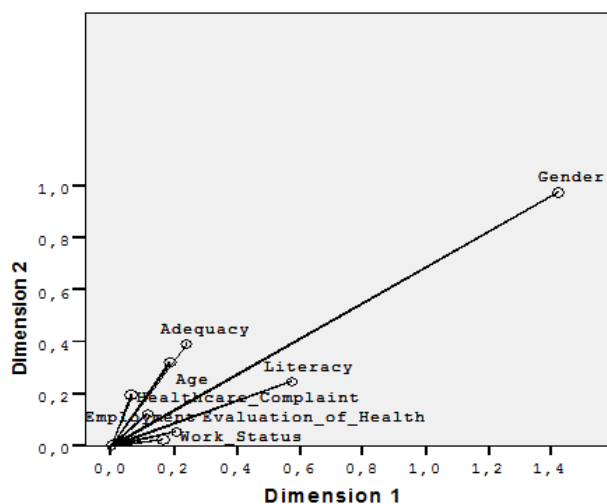


Fig. 3 Discrimination measures for each variable (Variable Principal Normalization)

For each variable, a discrimination measure was computed and presented on the two-dimensional graphical field (Fig. 3). The discrimination measure is also the variance of the quantified variable in that dimension. The average of the discrimination measures for any dimension equals the percentage of variance accounted for that dimension. Consequently, the dimensions are in descending order according to the average discrimination.

Fig. 3 shows that the first dimension is related to variables of work status, evaluation of healthcare and literacy. These variables have large and small discrimination measures on the first and second dimensions respectively. Thus, both these variables are widely distributed, but only on the first

dimension. Employment, healthcare complaint and adequacy are closer to the second dimension. Gender has the largest values on both dimensions, indicating that discrimination measures are the highest in both dimensions.

The optimal scale values for each variable are plotted in Fig. 4. Four headings of variables were used as active variables in the MCA: gender, age, literacy and the improvements in health care in the last three years. This map gives a comparable general view of the interrelationships among the variables.

The comparative positions of gender, age, literacy and the improvements in health care services in the last three years are reflected in Fig. 4. The values on the graph were derived from multiple correspondence analysis and plotted in two-dimensional graphical form. The information in the diagram reveals the assessment of the developments in health care services in the last three years by the respondents of the poll. It appears that females and younger individuals, who rated developments as very unsuccessful, are more critical about the developments in health care services than males and older respondents. The illiterate or less educated people expressed no idea, while older generation, aged 55-64 or more rated the health care services as successful.

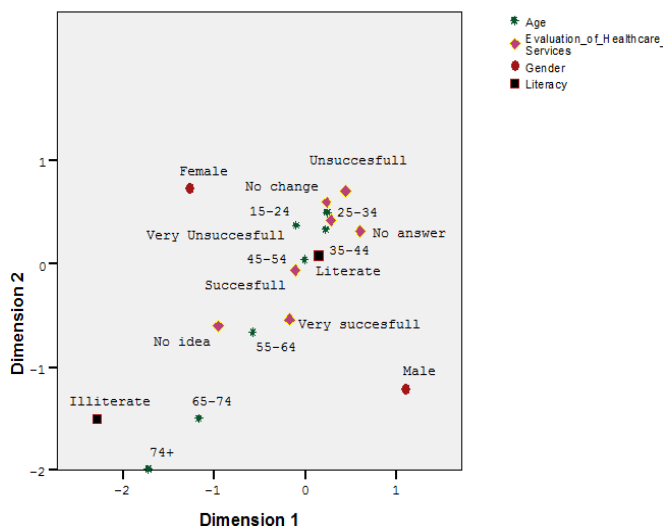


Fig. 4 Joint plot of category points for Health Data

#### IV. CONCLUSION

Multiple correspondence analysis is a particularly appropriate method for analyzing public opinion survey data. MCA adds statistical significance to the figures and ratios. It also facilitates comparable representation of the variables on the graphical field. Respondents' complaints about health services in the survey were mainly focused on doctor (52.8%) and equipment (22.4%). However, those complaints generally concerned shortages (rather than quality) of doctor and equipment.

MCA results indicated that females seem to be more complainant than males, literate more than illiterate and young and middle aged more than older people. The respondents using "public sector facilities" seem to be more complainant

than those in “private sector” in terms of doctor, hospital beds and other factors, indicating that private sector institutions provide better service than public. Awareness of health services increase with education level. Older and less well-educated people and males seem likely to have lower expectation about healthcare services. People, less educated persons and males might have lower expectations about health care services. On the other hand, females and younger people rated the health care as highly unsuccessful. Whereas illiterate respondents expressed no idea on the subject and the people aged (55-64) and older rated the system as successful. Their satisfaction might be related to the improved technology and services today compared to the past. Apparently younger and more educated people have wider vision and higher expectations in life.

This survey was carried out with the cooperation Turkish Ministry of Health and Turkish Statistical Institute in February in 2006. Although approximately half of the population (45.6%) satisfied about current health care system, the great and important changes in Turkish Health Care System took place in June 2006, just after this survey. Therefore, the Social Security Institution (SSI) was established in that time which brings Social Security Insurance for Craftsmen and Artisans (SSK), Social Security Insurance for other Self Employer and Retirement Fund for Government Official under a single roof to offer actuarial rights and obligations. The results of this study are compared to the results obtained from the new system, and should be seen advances in the health care system.

#### ACKNOWLEDGMENT

The authors would like to thank to Dr. Orhan Alpan for contribution and comments on this article. This work was supported by Research Fund of the Karabuk University. Project Number: KBÜBAP-17-YD-311.

#### REFERENCES

- [1] Delbanco TL: Quality of care through the patients' eyes. *BMJ* 1996, 313:832-833.
- [2] Bruster S, Jarman B, Bosanquet N, Weston D, Erens R, Delbanco TL: National survey of hospital patients. *BMJ* 1994, 309:1542-1546.
- [3] Delbanco TL: Enriching the doctor-patient relation: inviting the patient's perspective. *Ann Intern Med* 1992, 116:414-8.
- [4] Wensing M, Elwyn G: Methods for incorporating patients' views in health care. *BMJ* 2003, 326:877-879.
- [5] Organization for Economic Co-operation and Development. *OECD Health Data 2009: Statistics and Indicators for 30 Countries*.
- [6] Faye, G., Lumpkin, J., and Dant, R. (1992), 'Adaptation and Customer Expectations of Health Care Options', *Journal of Health Care Marketing*, 12 (3), Sept 92.
- [7] Fitzpatrick, R. (1991), 'Surveys of patient satisfaction: I—important general considerations', *British Medical Journal*, 302, 887-889.
- [8] John, J. (1992), 'Patient satisfaction: The impact of past experience', *Journal of Health Care Marketing*, 12 (3), 56-64.
- [9] Vavra, T. G. (1999). *Methods of Developing Measuring Patient Satisfaction*. Ankara: Kalder Publishing.
- [10] <http://website.leidenuniv.nl/~klootwavander/wakcv/homals.htm>, 28.01.2009.
- [11] Lebart L, Morineau A, Warwick K, *Multivariate Descriptive Statistical Analysis*, Chichester, UK: Wiley, 1984.
- [12] Blasius J, Greenacre MJ, *Visualization of Categorical Data*, San Diego: Academic Press, 1998.
- [13] "Health Services Satisfaction Survey", conducted by the Ministry of Health and TURKSTAT, 2006.

- [14] Hair, Jr., J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. (1998). "Multivariate Data Analysis," Fifth Edition, Upper Saddle River, New Jersey: Prentice-Hall.