SEARCHING PROBLEMS BETWEEN DOCTORS AND MEDICAL SALES REPRESENTATIVES WITH LOG-LINEAR ANALYSIS

Doktorlar İle Tıbbi Satış Mümessilleri Arasındaki Problemlerin Log-lineer Analiz İIle Araştırılması

Canmert BECANIM*
Hülya OLMUŞ**
Semra Oral ERBAŞ***

*İstatistikçi
**Araş. Gör. Dr., Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü
***Prof. Dr., Gazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü

ABSTRACT

As in many countries, medical advertisements are forbidden in our country. Therefore medical companies benefit from medical sales representatives to sell their product in today's environment of rivalry. Medical sales representatives establish a relation between doctors who propose medical products to their patients and medical manufacturer companies. Fast developments in the medical sector and conditions of increasing rivalry cause some problems between medical companies and doctors. For this purpose, a questionnaire was applied to doctors who work in Ankara, Turkey. In this study, variables that deal with the research subjects were first converted to a contingency table, and then relations between these variables were analyzed by using the log-linear model.

Key Words: *categorical data, contingency tables, log-linear model, odds proportion*

ÖZET

Birçok ülkede olduğu gibi, ülkemizde de ilaç reklamı yapmak yasaktır. Dolayısıyla ilaç firmaları, günümüzde yaşanan rekabet karşısında ürünlerini daha iyi pazarlmak için tıbbi satış mümessillerinden yararlanmaktadır. Tıbbi satış mümessilleri, ilaç üreten firmaların ürünlerinin hastalara ulaşmasında aracı olan doktorlarla ilişkiyi sağlayan kişilerdir. İlaç piyasasındaki hızlı gelişim ve artan rekabet koşulları, doktorlar ile tıbbi satış mümessilleri arasında bazı sorunlarında ortaya çıkmasına neden olmuştur. Bu amaçla Ankara ili ve çevresinde görev yapan doktorlar ile toplam 21 soruluk bir anket çalışması yapılmıştır. Yapılan çalışmada inceleme konusu olarak ele alınan değişkenler olumsallık çizelgeleri haline getirildikten sonra, bu değişkenler arasındaki ilişkiler log-lineer modellerin kullanımı ile incelenmeye çalışılmıştır.

Anahtar Sözcükler: *kategorik veri, olumsallık çizelgeleri, log-linear model, odds oranı*

INTRODUCTION

The number and variation of disease has risen up depending on the growing population of the world. As a consequence of this, the pharmaceutical companies are trying to develop alternative medication to cure these diseases (Haberman, 1974: 6).

The rivalry environment caused by the increase in the number of pharmaceutical companies enforces companies to create new marketing strategies and to develop new products. As in many countries, medical advertisements are forbidden in our country therefore the companies need to introduce their products directly to doctors who provide connection between companies and patients. The ban on medical advertisement brought about the need for people to market products of the pharmaceutical companies. This need for people who are liaisons between doctors and pharmaceutical companies gave rise to a new profession called medical sales representative.

Medical sales representatives besides introducing their companies' products to doctors they also try to increase company's profitability. Therefore, they have considerable responsibilities such as getting in direct contact with the doctors to influence them and market their products.

The occurrence of same kind of medication by different companies and the challenging situations have caused a highly competitive environment. Therefore medical sales representatives try to influence doctors that their products are better than their counterparts as well as increase their company's market share. As a result of this increase in competition among pharmaceutical companies, it is seen that some pharmaceutical companies are trying different methods to influence doctors and that the medical sales representatives are violating the rules of visiting. Unavoidably, as a consequence the problems between doctors and medical sales representatives have become unavoidable.

In this study, conspicuous problems arising between doctors and medical sales representatives are investigated. For this purpose, the opinions of doctors from various areas in Ankara and its surroundings are considered. The doctors asked for their opinions in the study were selected from small hospitals and state hospitals by stratified simple random sampling method. The effects in this study and their interactions can be examined using multivariate contingency tables. Until recently, chi square analysis has been used to determine the relationships between variables in the contingency tables however some problems arise when using these analyses. For example when there are more than two variables, it is very hard to interpret the results of the contingency tables by using chi square analyses. For this reason, in multivariate contingency tables chi square independency tests are insufficient for indicating and calculating the effects and their interactions (Haberman, 1988: 7). Log linear models which account for insufficient parts of chi square analyses and provide easier ways of interpreting complicated relationships without restricting the number of variables are preferred over the chi square analyses (Agresti, 1990: 1).

Log linear models are statistical techniques for determining the relationships between categorical variables. The advantage of log linear models is not only avoiding from making lots of cross tables but also finding unnoticeable interactions and determining real relationships between variables. Particularly after 1960s, the applications of log linear models for categorical variable analyses have increased considerably. In studies of Bishop, 1975: 4; Goodman, 1970: 8; Haberman, 1974:6 etc. the log linear models have found many application areas.

For the variables of interest in this study, Pearson's χ^2 and Wilks' likelihood ratio statistic L^2 are used for choosing the appropriate log linear model. In addition, after deciding on the appropriate log linear model, odds ratios will be used for interpretation.

In the light of these concepts, 414 doctors living in Ankara and its surroundings are surveyed and the contingency tables are constructed using the survey data and log linear analysis method is applied to these contingency tables. The odds ratios between the variables of the appropriate models are evaluated and the results are interpreted.

Log-linear models used in analyses of multivariate contingency tables:

In today's world the researches in social sciences are mostly observed as categorical variables. Multivariate categorical variables are statistical units which have more than one classes or orders.

Chi square independency tests are performed using multivariate contingency tables however as the number of categories of variables increase, it becomes very difficult to compare rows and columns where the categories of the variables are located and in some situations this comparison becomes impossible. In such situations, log linear models which do not put any restriction on the number of rows and columns of contingency tables and which do provide ways of testing different hypotheses than chi square tests should be preferred. Log linear models bring out the structural relationship between variables without classifying them as dependent and independent variables which is another reason for superiority of log linear model over chi square analysis (Agresti, 1984: 2).

Log-linear models for three way contingency tables:

In a three way contingency table, let Xijk denote the dependent variable and nijk denote the observation (i=1,2,…,I; j=1,2,…,J; k=1,2,…,K) Let also the log linear model for variables A, B and C is as shown below.

$$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC} \quad (1)$$

The parameters and dof in this equality are shown in Table 1 (Andersen, 1990: 1).

8 log linear models are obtained for three way contingency tables. The 8 log linear models, dofs and conditionally independent variable pairs are given in Table 2.

$A \perp C \setminus B$ shown in Table 2 means that given the levels of B the variables A and C are conditionally independent.

Odds and Odds Ratio:

Odds or odds ratios is one of the concepts used for interpreting variables in contingency tables and the log linear model related to this contingency table.

Table 1. Parameters and dof in the model and complete parameters in explicit form

Parameters	dof	complete parameters in explicit form
λ_o	1	$n_{...}$
λ_i^A	(I-1)	$n_{i..}-n_{...}$
λ_j^B	(J-1)	$n_{.j.}-n_{...}$
λ_k^C	(K-1)	$n_{..k}-n_{...}$
λ_{ij}^{AB}	(I-1) (J-1)	$n_{ij.}-n_{i..}-n_{.j.}+n_{...}$
λ_{ik}^{AC}	(I-1) (K-1)	$n_{i.k}-n_{i..}-n_{..k}+n_{...}$
λ_{jk}^{BC}	(J-1) (K-1)	$n_{.jk}-n_{.j.}-n_{..k}+n_{...}$
λ_{ijk}^{ABC}	(I-1) (J-1) (K-1)	$n_{ijk}-n_{ij.}-n_{i.k}-n_{.jk}+n_{i..}+n_{.j.}+n_{..k}+n_{...}$

Let p denote the probability of an event. The odds ratio for this event is calculated as below

$$\Omega = \frac{p}{1-p} \qquad (2)$$

If the probability that an event occurs (p) is greater than probability that an event does not occur (1-p) then the odds ratio will be greater than 1. In other words, if it is desired to obtain an odds ratio greater than 1 then the probability that an event occurs has to be greater than 0,5. As the p value gets closer to 0 then the odds ratio approximates to 0, else if the p value gets closer to 1, the odds ratio goes to infinity (Kroke&Burke, 1980: 9).

Choosing the appropriate model:

In the log linear application of a certain contingency table, the model obtained should be the most representative model for the population. For this reason, the goal of model selection methods is to find the most appropriate model for the population. Pearson's χ^2 statistic and Wilks' likelihood ratio statistic L^2 are two statistics used to search for the most appropriate model for the population. Which test should be preferred for choosing the most appropriate model? If the researcher has prior information about the population, then he would use Pearson's χ^2 statistic to find the model that fulfills his expectation however if the researcher has more than one model which he thinks are appropriate and he is trying to find the most representative model he should prefer Wilks' likelihood ratio statistic L^2. If there is more than one hierarchical model appropriate for the population then comparing all the models with the saturated model and using the L2(M) differences test statistics for each model would be the most practical way (Le, 1998: 10).

Table 2. Possible Log linear models for variables A, B, C

	Models	dof	Conditionally independent variable pairs
$M^{(0)}$	$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C$	IJK-I-J-K+2	None
$M^{(1)}$	$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{jk}^{BC}$	(I-1)(JK-1)	$A \perp C \setminus B$ $B \perp A \setminus C$
$M^{(2)}$	$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC}$	(J-1)(IK-1)	$B \perp C \setminus A$ $A \perp B \setminus C$
$M^{(3)}$	$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB}$	(K-1)(IJ-1)	$B \perp C \setminus A$ $A \perp C \setminus B$
$M^{(4)}$	$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$	K(I-1)(J-1)	$A \perp B \setminus C$
$M^{(5)}$	$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{jk}^{BC}$	J(I-1)(K-1)	$A \perp C \setminus B$
$M^{(6)}$	$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC}$	I(J-1)(K-1)	$B \perp C \setminus A$
$M^{(7)}$	$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$	(I-1)(J-1)	None
$M^{(8)}$	$\log_e m_{ijk} = \lambda_0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$	0	None

Criteria for choosing the appropriate model:

In situations where there are more than one appropriate models for a population, R2, λ ve AIC statistics are used to find the most representative model. R2 criteria is used for choosing between hierarchical log linear models and

$$R^2 = \frac{L^2(M^{(0)}) - L^2(M)}{L^2(M^{(0)})} \quad (3)$$

R2 takes values between 0 and 1. According to this criteria as R2 gets closer to 1 the next model M , and as R2 gets closer to 0 M(0) model is accepted to be the appropriate model. The lack of existence of the number of parameters in the model in R2 equation can confuse the researcher so another criteria λ alternative to R2 is suggested.

$$\lambda = \frac{L^2(M^{(0)})/sd\,(M^{(0)}) - L^2(M)/sd(M)}{L^2(M^{(0)})/sd\,(M^{(0)})} \quad (4)$$

λ is interpreted in a similar way as R2. The highest value that lambda can take is 1 and lambda is smaller than R2. For choosing the appropriate model λ (lamda) is used as the following. If there are two models that have the same R2 value then the model with the highest λ value is chosen to be the most appropriate model. Another criterion is the Akaike Information criterion (AIC) which was suggested by Akaike. One important feature of this criterion is that it approves models with few parameters as the most appropriate model does not take into account the models with many parameters. Akaike criterion can be calculated using the following equation.

$$AIC = L^2(M) - (q - 2sd(M)) \quad (5)$$

In the equation above, q denotes the number of observations in the table (Christensen, 1990: 5). The model with the smallest AIC values is considered to be the most appropriate model for the population. Please note that, none of these model selecting criteria guarantees that the model chosen is the right one. Different selecting criteria may suggest different appropriate models. The researcher should consider the population data used in the study and his knowledge and experience while deciding on the most appropriate model.

MATERIALS AND METHOD

In this study, 414 doctors living in Ankara and its surroundings are surveyed (Becanım, 2006: 3). The contingency tables are constructed using the survey data and log linear analysis method is applied to these contingency tables by the help of SPSS 13.0. The variables and their levels are given below.

A: What time interval do you think the medical sales representatives' should visit you? The variables are classified as (08:30-9:30), (09:30-10:30), (10:30-11:30), (11:30-12:30), (12:30-13:30), (13:30-14:30), (14:30-15:30) ve (15:30-16:30)

B: Where do you think the medical sales representatives' should visit you?

The variables are classified as policlinic, doctor's room, cafeteria, consulting room

C: Where is your workplace? State hospital, small hospital

The variables of interest and their distributions in the population are shown in Table 3, Table 4 and Table 5.

Table 3. The frequency table for doctors' workplaces

Unit	Number	Percentage (%)
small hospital	276	66.67
state hospital	138	33.33
Total	414	100

Table 4. The frequency table for doctors' opinions about the suitable visiting place

Visiting place	Number	Percentage(%)
policlinic	135	32.61
doctor's room	235	56.76
cafeteria	16	3.86
consulting room	28	6.77
Total	414	100

Table 5. The frequency table for doctors' opinions about the suitable the time intervals

Time intervals	Number	Percentage(%)
08:30-09:30	108	26.08
09:30-10:30	33	7.97
10:30-11:30	27	6.52
11:30-12:30	70	16.91
12:30-13:30	34	8.21
13:30-14:30	10	2.42
14:30-15:30	18	4.35
15:30-16:30	114	27.54
Total	414	100

RESULTS

The contingency tables are constructed for A, B, C and the most appropriate model is determined using log linear model analysis. R2, λ , AIC criteria are considered to decide on the most appropriate log linear model. The results for all possible log linear models for variables A, B, C are given in Table 6.

It can be seen from Table 6 that if R2 criterion is considered, 79% of the variability among variables is explained by (AB, AC, BC) model however although loosing 1% of the explained variability a simpler model (AB, BC) is preferred. (AB, AC, BC) model has the smallest AIC value. But when other model selection methods and criteria are considered (AB, BC) model seems to be the most appropriate model for the data.

Table 6. Results for all possible Log linear models for A, B, C

Model	dof	χ^2	p	L^2	p	R^2	€	AIC
A,B,C	52	130.148	0	121.634	0	-	-	161.634
A, CD	49	101.230	0	105.072	0	0.136	0.083	139.072
B, CD	45	113.651	0	110.579	0	0.091	-0.051	136.579
C, AB	31	51.144	0.013	53.175	0.008	0.563	0.267	51.175
AB, AC	24	39.145	0.026	42.120	0.013	0.654	0.249	26.12
AB, BC	28	34.273	0.192	36.612	0.128	0.699	0.441	28.612
AC, BC	42	87.330	0	94.017	0	0.227	0.043	114.017
AB, AC, BC	21	22.754	0.357	25.675	0.219	0.789	0.477	3.675
ABC	0	0	-	0	-	1	1	-64

Table 7. The distribution of doctors according to workplaces and policlinic and other levels

Workplace	Policlinic	Other (doctor's room, cafeteria, consulting room)
small hospital	89	187
state hospital	46	92

The odds ratios for the variables in the final model are calculated and the results are interpreted. As an example, the distribution of doctors according to workplaces and policlinic and other levels are shown in Table 7.

The Odds ratio from Table 7 is,

$$\theta = \frac{89/187}{46/92} = \frac{0.476}{0.50} = 0.95 \Rightarrow \theta_t = \frac{1}{0.95} = 1.05$$

The odds of preferring policlinics as a visit place for doctors who work at the hospital is 1.05 times as great as for doctors who work at small hospital.

The odds ratio interpretations for other variables are given below.

1. The odds of preferring doctor's room as a visit place for doctors who work at the small hospital is 1.65 times as great as for doctors who work at the hospital.
2. The odds of preferring cafeteria as a visit place for doctors who work at the hospital is 1.19 times as great as for doctors who work at small hospital.
3. The odds of preferring consulting room as a visit place for doctors who work at the hospital is 4.69 times as great as for doctors who work at small hospital.
4. The odds of preferring (8.30-9.30) as a visit time interval for doctors who work at the small hospital is 1.19 times as great as for doctors who work at the hospital.

5. The odds of preferring (9.30-10.30) as a visit time interval for doctors who work at the small hospital is 1.15 times as great as for doctors who work at the hospital.
6. The odds of preferring (10.30-11.30) as a visit time interval for doctors who work at the hospital is 3.12 times as great as for doctors who work at the small hospital.
7. The odds of preferring (11.30-12.30) as a visit time interval for doctors who work at the small hospital is 1.18 times as great as for doctors who work at the hospital.
8. The odds of preferring (12.30-13.30) as a visit time interval for doctors who work at the hospital is 1.01 times as great as for doctors who work at the small hospital.
9. The odds of preferring (13.30-14.30) as a visit time interval for doctors who work at the hospital is 1.22 times as great as for doctors who work at the small hospital.
10. The odds of preferring (14.30-15.30) as a visit time interval for doctors who work at the hospital is 1.44 times as great as for doctors who work at the small hospital.
11. The odds of preferring (15.30-16.30) as a visit time interval for doctors who work at the small hospital is 1.24 times as great as for doctors who work at the hospital.

DISCUSSION

As it is known, chi square analysis is the widely used method for analyzing contingency tables. However as the number of variables and levels of variables increase, the application of chi square independency tests becomes difficult and provides insufficient information and this situation leads to preference of log linear models.

Log linear model which brings out the structural relationship between variables without classifying them as dependent or independent has become a popular tool for analyzing contingency tables recently. In this study, conspicuous problems arising between doctors and medical sales representatives are investigated by the help of contingency tables, log linear analysis and odds ratios. The most appropriate log linear model is determined by using these contingency tables and the odds ratios for variables in the model are calculated. According to the calculated odds ratios the findings are summarized as follows.

- Most of the doctors who work at the small hospital prefer that the medical sales representatives should visit them in their rooms.
- Most of the doctors who work at the hospital prefer that the medical sales representatives should visit them in consulting rooms.
- The doctors who work at the hospital are the dominant group who prefer that the medical sales representatives should visit them in the cafeteria.
- Most of the doctors who work at the small hospital think that (11.30-12.30) is a convenient time for medical sales representatives' visits.
- However doctors who work at the hospital think that (10.30-11.30) is a convenient time for medical sales representatives' visits.

REFERENCES

Agresti, A. (1990) Categorical data analysis, John Wiley and Sons, New York.

Agresti, A. (1984) Analysis of ordinal categorical data, John Wiley and Sons, New York.

Becanım, C. (2006) "Log-Lineer Modeller ve Doktor-Tıbbi Satış Mümessilleri İlişkileri Üzerine Bir Uygulama", Gazi Üniversitesi Fen Bilimleri Enstitüsü, Yayınlanmamış Yüksek Lisans Tezi, Ankara.

Bishop,Y.M., Fienberg, S., ve Holland, P. (1975) Discrete multivariate analysis, M.I.T. Press, Cambridge, Mass.

Christensen, R. (1990) Log-linear models, Springer-Verlag, New York.

Haberman, S.J. (1974) The analysis of frequency data, Univ.Chicago Press, Chicago.

Haberman, J.S. (1988) "A warning on the use of chi-squared statistics with frequency tables with small expected cell counts", JASA, 63 (402) 555-561.

Goodman, L.A. (1970) "The multivariate analysis of qualiative data: interaction among multiple classification", J.Amer. Statist.Assoc., 65, 226- 256.

Kroke, D. and Burke, P.J. (1980) Log-linear models, Sage Publication Inc., London.

Le, C. T. (1998) Applied categorical data analysis, New York, Wiley.