Classification of the Insurance Sector with Logistic Regression

BAHADTIN RUZGAR*, NURSEL SELVER RUZGAR**
*Banking and Insurance School, Actuaries Department, Marmara University
Goztepe Campus, Kadikoy, 34722, Istanbul, TURKEY
http://mimoza.marmara.edu.tr/~bruzgar/

**Vocational School of Social Sciences, Marmara University
Beyazit Campus, Eminonu, Istanbul, TURKEY

Abstract: - In cases when categorical results such as successful-unsuccessful, ill-not ill, good-fair-bad are obtained especially as a result of evaluation of data, the logistic regression is a rather suitable statistical method. In this study, the data of 53 companies that are active in the insurance sector in Turkey for 2004, 2005 and 2006 years were evaluated by using logistic regression method. However, as the data were not sufficient for all the insurance companies, 12 insurance companies were eliminated from the evaluation. 41 companies used for the analysis were divided into two groups depending on their activity area .17 companies were evaluated by using the data on individual accident, health and life branches and 24 companies by using data on fire, transportation, engineering, agriculture, all-risks, obligatory traffic, obligatory highway transportation, individual accident and other accident and health branches.

Key-Words: - Logistic regression, classification, discriminant analysis, insurance companies classification

1 Introduction

The researchers or model designers always endeavor to convert the data they obtain from real events or experiments to functional structures by means of various models. Though to establish mathematical models is rather difficult, they ensure very beneficial information. Classification of the data used in models constitutes the very important part of the statistical analysis but it is widely used by various science branches mainly in health. Some studies about logistic regression are; the comparison of the mobile nursing system that was established between the years of 1977-1985 in America to restrict the health expenses to the former system was examined with multi logistic regression analysis [1]. In America, the data obtained from extraordinary events such as wars, elections, political crisis and epidemic diseases were used to determine differences between the periods when such events occurred and other periods by means of logistic regression [2]. Binary logistic regression was used to calculate the retirement age of people depending on age, sex, economical and social statuses [3]. Between the years of 1980-1995, the data of bankrupted American companies were examined, 237 of the bankrupted companies were handled as samples in 1992 and the financial and non-financial values of their final bankruptcy resolutions were examined with logistic analysis and their classifications were tried to be estimated [4]. Again, in respect to health insurance, the health insurance classification of insured and uninsured low-income children in America between the years of 1995-1999 and the classification of uninsured ones according their sex, age and economical status were made by using the logistic regression model [5]. The national health researches of the Australian households were made by using 2001 data and their passing rate to private health insurance and the reasons of such passing such as economical, social and health factors were examined by means of multi logistic regression analysis [6]. In classification of car accidents in America, the logistic regression analysis was made by using the variables such as literacy rate, economical status and sex [7]. 57 big parent companies that were very important for the Japan economy between the years of 1998-2001 were classified as “financially under stress” and “peaceful” [8]. To determine whether the Treatment Center established for purpose of treating visual disabled or blind people in order they could find new jobs was beneficial for such people, logistic modeling was used [9]. Between the years of 1980-2004, the disability risk and disability risk insurances were examined in America and a classification was made by using the logistic regression according to workability limits, non-workability situations and the need to get health aid of people who retired for reason of a physical disability [10]. In another study, the logistic regression was used in determination and classification of car insurance tariffs of insurance companies [11]. When theoretical studies related to the logistic regression are viewed, it can be determined that the widely used logistic regression models and development of coefficient estimation methods have caused logistic regression models to be examined in a more detailed manner. In the logistic

coefficient estimation procedure, the popular discriminate function approach was used [12]. χ^2 Likelihood rate (G^2), pseudo likelihood estimations, consistency benefit and hypothesis tests were examined in the logistic regression [13]. The distribution of fault terms and approach of parameter values to real values were examined in the logistic regression [14]. Mostly, a researcher or experimenter desires to find whether there is a relation between two or more variables and to express such relation with an equation [15]. For instance, an engineer may want to know the relation between the pressure and temperature, an economist between the income level and consumption expenses, an insurer between the number of sold policy and its profitability and an educator between the absent days of students and their success ranks. An equation showing the relation between two (or more) variables not only demonstrates the functional form of the relation between variables but also estimates any variable if the value of another is known [16]. Determination the relation between two or more variables is generally necessary for two types of information. These are, firstly, the reliability of estimations on values of any variable by means of observation results on another variable and, secondly, the rate of some determinative factors related to observed differences in variable values. In other words, in case two rational variables are connected to each other, the information on one of these variables can be used to estimate the values of other variable. For this reason, the functional type, direction and rank of the relation between variables must be known. When a dependent variable is a classified variable which depends on two situations while independent variables can be continuous, discrete or classified, logistic regression has a quite functional relation and it is suitable for category classification by using the structure of regression analysis.

2 Methodology

The purpose of the study is to control the explained success performances of 53 insurance companies and to check whether the companies that are divided into two groups according to geometrical mean and median are in correct classification in relation to rates having deviated end values. In the study, the logistic regression analysis was made by considering the policy numbers and total premium productions of 53 insurance companies, but only 41 companies were evaluated [17], for 2004 and 2005 years in basis of 12-month branches. The statistics of companies have been regularly broadcasted in internet and the companies are being classified according to the changes in their statuses when compared to their statuses of previous year. The success ranking of companies is made as companies in the first 10 and companies between 11 and 20. Whether such classification of 41 companies (17 companies are group I and 24 companies are group II) collides with the classification of "successful" and "unsuccessful" companies according to geometrical mean and median was determined with a comparison. The first six-month data of 2006 year were used for control and the classification obtained from models was compared to real classification of companies.

3 Results

17 companies (Group I) of 41 companies were examined by considering the data on individual accident (IA), health (H) and life (L) branches and 24 companies (Group II) by considering the data on fire (F), transportation (T), engineering (E), agriculture (A), all-risks (AR), obligatory traffic (OT), obligatory highway transportation (OHT), individual accident (IA) and other accident (OA) and health (H) branches. A classification was made with 2 different regression models by using the data of 17 companies in Group I and 24 companies in Group II for 2004 and 2005 years. Firstly, when they were classified by considering their success percentages when compared to previous year for 2004 and 2005 years, the companies in the first 20 were classified as "successful" and others as "unsuccessful". Besides, for data of 2004 and 2005 years, separate models were found. Secondly, the policy numbers of companies for 2004 and 2005 years were considered and as they were formed from deviated data, they were classified as" successful-unsuccessful" according to geometrical mean and median. For each 2004 and 2005 years, the companies of which policy numbers were within and above the geometrical mean were classified as successful and below the geometrical means as unsuccessful. The companies within and above the median depending on their policy numbers were classified as successful and below the median as unsuccessful. In this way, the logistic regression equations were found. In Table 1 and Table 2, the logistic regression equations belong to 17 companies in Group I and in Table 3 and Table 4, the logistic regression equations belong to 24 companies in Group II are given. When we examine the statistics related to the logistic regression, we find that Cox-Snell R^2 (CS-R^2) and Nagelkerke R^2 (Nag R^2) values that show the degree of relation between the dependent and independent variables in the logistic regression models are higher and -2LogL (-2log likelihood=-2LL) statistic is lower. When the model exactly represents the data, the likelihood is 1 and -2LL statistics is zero. For this reason, the lower -2LL statistic always shows a better model [18, 19]. When the statistics related to testing of meaningfulness of model are examined, Chi-square (χ^2) statistic, -2LL statistic and Blok Chi-square (B χ^2) statistic must be considered. The χ^2 statistic tests the

logistic regression model in general. The χ^2 statistic firstly shows the fault only when there is a fixed term in the model and then it determines whether all the logistic coefficients except the fixed term are equal to zero. The χ^2 statistic conforms to χ^2 distribution with degree of freedom equals to difference between the parameter number of examined model and parameters of model with fixed term [20]. In logistic regression, -2LL statistic shows the fault of model when an independent variable is added to model. For this reason, it is the measure of unexplained variance in a dependent variable and the non-meaningful statistic is a desirable situation. The B χ^2 statistic shows the changing in the χ^2 statistic when a block variable is added to model [21]. When the relation measure in a logistic regression analysis is examined, we see that a widely accepted statistic logistic regression that is similar to R^2 statistic does not exist. In the regression analysis, R^2 shows the percentage of explained variance of dependent variable but the variance of a dependent variable in the logistic regression analysis depends on the probability distribution of that variable. For this reason, R^2 in regression analysis must not confuse with R^2 in logistic regression. The mostly used R^2 statistics for the logistic regression are CS-R^2 statistic and Nag R^2 statistic [18, 22]. CS-R^2 statistic may have a value higher than zero. Its value below 1 strengthens the interpretation of statistic. Nag R^2 statistic was developed to ensure CS-R^2 statistic to have values between 0 and 1. The statistic closer to 1 means the relation is high.

Table 1 The logistic regression values of 17 companies in the first 20 of Group I (successful) and of other companies (unsuccessful) for 2004 and 2005 years.

First20$_{2004}$=-7,351484+4,579242E-04*IA -6,741806E-6*L					
R^2	D.F.	χ^2	Model Prob	Correctly Class. %	Misclass. Rows Sec. Actual/Predicted
0.54	2			94.12	
Step	-2 LL	CS-R^2	Nag R^2	Overall %	The cut value
1	4.18	0.62	0.88		
First20$_{2005}$=-1,276982+3,363836E-05*IA-1,554465E-03*H					
R^2	D.F.	χ^2	Model Prob	Correctly Class. %	Misclass. Rows Sec. Actual/Predicted
0.37	2	8.16	0.0167	82.35	4(1/0),6(1/0), 11(0/1)
Step	-2 LL	CS-R^2	Nag R^2	Overall %	The cut value
1	11.61	0.34			0.5

When 17 companies in Group I are classified according to their success percentages rate when compared to previous year in classification of successful-unsuccessful by considering whether they are in the first 20 or not of all insurance companies- it is seen that they are in correct classification with 94.12% for 2004 year and 82.35 % for 2005 year. In the 2004 year logistic regression model, number 4 company, Anodolu Hayat ve Emeklilik, is estimated as unsuccessful in reality it is successful. In the 2005 year logistic regression model, number 4, Anadolu Hayat ve Emeklilik, and number 6, Aviva Hayat ve Emeklilik, companies are estimated as unsuccessful in reality they are successful. Number 11 company, Garanti Emeklilik, is estimated as successful in reality it is unsuccessful. In the 2005 year logistic regression model, Nag R^2 is 0.505 and -2LL is 11.610 and these demonstrate that the model is not in a good classification.

Table 2 The logistic regression values of 17 companies in Group I that are classified as successful and unsuccessful according to geometrical mean and median of their policy numbers for 2004 and 2005 years.

Policy_num$_{2004}$= -367.218+ 2.41603E-03*L (GM) Policy_num$_{2004}$=-248.6836+9.521043E-04*L+8.481388E-04*IA (Me)					
R^2	D.F.	χ^2	Model Prob	Correctly Class.%	Misclass. Rows Sec. Actual/Predicted
0.60 /0.63	1/2	22.1/23.5	0/0	100/100	-/-
Step	-2LL	CS- R^2	Nag R^2	Overall %	The cut value
1/1	0/0	0.73/0.75	1/1	100/100	0.5/0.5
Policy_num$_{2005}$=-84.39449+1.015181E-03*L+8.428923E-04*IA(GM) Policy_num$_{2005}$=-41.95804+3.689567E-04*L+8.586143E-04*IA (Me)					
R^2	D.F.	χ^2	Model Prob	Correctly Class. %	Misclass. Rows Sec. Actual/Predicted
0.63/0.63	2/2	23.5/23.5	0/0	100/100	-/-
	-2LL	CS- R^2	Nag R^2	Overall %	The cut value
	0/20.9	0.74/0.14	1/0.19	100/58.8	0.5 /0.5

When 17 companies in Group I are classified as successful and unsuccessful according to geometrical mean and median in respect to their policy numbers of 2004 and 2005 years, it is seen that this classification is made in 100% correctness. Nag R^2 is 1 and -2LL is 0.000 in the logistic regression model made according to geometrical mean and median for 2004 year. However, though the classification is not faulty, in the logistic regression model of 2005 according to median, Nag R^2 is 0,185 and -2LL is 20,973. This model is not a suitable logistic regression model.

Table 3 The logistic regression values of 24 companies in the first 20 of Group II (successful) and of other companies (unsuccessful) for 2004 and 2005 years.

First20$_{2004}$=-63.82845+9.545645E-04*OHT+7.222392E-04*T					
R^2	D.F.	χ^2	Model Prob	Correctly Class. %	Misclass. Rows Sec. Actual/Predicted
0.6	2	31.76	0	100	-
Step	-2 LL	CS- R^2	Nag R^2	Overall %	The cut value
1	0	0.7	1	100	0.5
First20$_{2005}$= -109,7679+ 2,055966E-03*OHT					
R^2	D.F.	χ^2	Model Prob	Correctly Class. %	Misclass. Rows Sec. Actual/Predicted
0.59	1	31.76	0	100	-
Step	-2 LL	CS- R^2	Nag R^2	Overall %	The cut value
1	0	0.734	1	100	0.5

When 24 companies in Group II are classified in respect to their percentage success rates when compared to

previous year – their classification of successful-unsuccessful by considering whether they are in the first 20 or not among all other companies- it is seen that classification for 2004 and 2005 years are made correctly with 100%. Nag R^2 is 1 and -2LL is 0.000 for 2004 and 2005 years and this shows that the logistic regression models are correct. Besides, none of the companies is classified faulty in reality and estimation.

Table 4 The logistic regression values of 24 companies in Group II that are classified as successful and unsuccessful according to geometrical mean and median of their policy numbers for 2004 and 2005 years.

Policy_num$_{2004}$= -644.5532+0.0141135*F+ 0.0136799*T (GM) Policy_num$_{2004}$= -644.5532+ .0141135*F+ .0136799*T (Me)					
R^2	D.F.	χ^2	Model Prob	Correctly Class. %	Misclass. Rows Sec. Actual/Predicted
0.6/0.6	2/2	32.6/32.6	0/0	100/100	-/-
Step	-2 LL	CS- R^2	Nag R^2	Overall %	The cut value
1/1	0/0	0.75/0,74	1/1	100/100	0.5/0.5
Policy_num$_{2005}$= -109.7679+ 2.055966E-03*OHT (GM) Policy_num$_{2005}$= -271.143+ 1.729338E-03*OHT+ 3.164847E-03*T-1.054694E-02*OT (Me)					
R^2	D.F.	χ^2	Model Prob	Correctly Class. %	Misclass. Rows Sec. Actual/Predicted
0.6/0.6	1/3	31.8/33.3	0/0	100/100	-/-
Step	-2 LL	CS- R^2	Nag R^2	Overall %	The cut value
1/1	0/0	0.7/0.8	1/1	100/100	0.5/0.5

When 24 companies in Group II are classified as successful and unsuccessful according to geometrical mean and median in respect to their policy numbers of 2004 and 2005 years, it is seen that this classification is made in 100% correctness. When the policy numbers in years of 2004 and 2005 are examined separately, it is determined that Nag R^2 is 1 and -2LL is 0.000. This shows that the logistic regression models are in correct classification.

When a classification is made according to the first six-month of 2006 by considering the data of 2004 year (companies in the first 20 are deemed as successful and others as unsuccessful), it is estimated that 16 of 17 companies (94%) are in correct classification and 1 (6%) of them in faulty classification. When a classification is made by considering the data of 2005 year, it is estimated that 14 of 17 companies (82%) are in correct classification and 3 (18%) of them are in faulty classification. When the companies above according to the first six-month of 2006 the geometrical mean are grouped as successful and ones below the geometrical mean as unsuccessful by considering the geometrical mean of policy numbers of 17 companies in Group I; it is estimated that 15 of 17 companies (83%) are in correct classification and 2 of them (12%) are in faulty classification. When a classification is made by considering the data of 2005 year, it is estimated that 16 of 17 companies (94%) are in correct classification and 1 (6%) of them is in faulty classification. When the companies above the median are classified as successful and ones below the median as unsuccessful, it is estimated that 13 of 17 companies (76.5%) are in correct classification and 4 of them (23.5%) are in faulty classification. When a classification is made by considering the data of 2005 year, it is estimated that 16 of 17 companies (94%) are in correct classification and 1 (6%) of them is in faulty classification. 12 of 17 companies in Group I (59%) are classified correct with all methods.

Table 5 The comparison of the first six-month real data of 17 companies in Group I for 2006 year to the estimated classification values found with application of the logistic regression equations estimated for 2004 and 2005 years to the values of 2006 year.

Company Name (Grup I)	**First 10-20**	**2004**	**2005**	**GM. Pol. Num**	**2004**	**2005**	**Me. Pol. Num**	**2004**	**2005**
Acıbadem Saglık Ve Hayat	0	0	0	0	0	0	0	0	0
Ak Emeklilik	0	0	0	1	0	0	1	0	0
American Life	0	0	0	0	0	0	0	0	0
Anadolu Hayat Ve Emeklilik	1	0	0	1	1	1	1	1	1
Ankara Emeklilik	0	0	0	0	0	0	0	0	0
Aviva Hayat Ve Emeklilik	1	1	0	1	0	1	1	0	1
Axa Oyak Hayat	1	1	1	1	1	1	1	1	1
Basak Emeklilik	1	1	1	1	1	1	1	1	1
Birlik Hayat	0	0	0	1	1	1	1	0	1
Demir Hayat	0	0	0	0	0	0	0	0	0
Garanti Emeklilik	0	0	1	1	1	1	1	0	1
Genel Yasam	0	0	0	0	0	0	0	0	0
Guven Hayat	0	0	0	0	0	0	0	0	0
Isvıcre Hayat	0	0	0	0	0	0	0	0	0
Koc Allianz Hayat- Emeklilik	0	0	0	1	1	1	1	1	1
Vakıf Emeklilik	0	0	0	0	0	0	0	0	0
Yapı Kredi Emeklilik	1	1	1	1	1	1	1	1	1
Faulty Cell Number		**1**	**3**		**2**	**1**		**4**	**1**

When a classification is made according to the first six-month of 2006 by considering the data of 2004 year (companies in the first 20 are deemed as successful and others as unsuccessful), it is estimated that 18 of 24 companies (75%) are in correct classification and 6 (25 %) of them in faulty classification. When a classification is made by considering the data of 2005 year, it is estimated that 20 of 24 companies (83%) are in correct classification and 4 (17%) of them are in faulty classification. When the companies above according to the first six-month of 2006 the geometrical mean are grouped as successful and ones below the geometrical mean as unsuccessful by considering the geometrical mean of policy numbers of 24 companies in Group II; it is estimated that 23 of 24 companies (96%) are in correct classification and 1 of them (4%) are in faulty

classification (in classification made by considering the data of 2004 year).

Table 6 The comparison of the first six-month real data of 24 companies in Group II for 2006 year years to the estimated classification values found with application of the logistic regression equations estimated for 2004 and 2005 years to the values of 2006 year.

Company Name (Grup II)	**First 10-20**	**2004**	**2005**	**GM. Pol Num**	**2004**	**2005**	**Me. Pol Num**	**2004**	**2005**
Aig	0	0	0	0	0	0	0	0	0
Ak Sigorta	1	1	1	1	1	1	1	1	1
Anadolu	1	1	1	1	1	1	1	1	1
Ankara	1	0	1	0	0	0	0	0	0
Aviva	1	0	0	1	1	0	0	1	0
Axa Oyak	1	1	1	1	1	1	1	1	1
Basak	1	1	1	1	1	1	1	1	0
Birlik	0	0	0	0	0	0	0	0	0
Finans	1	0	0	1	1	1	0	0	0
Garanti	1	0	0	1	1	1	1	1	0
Generali	0	0	0	0	0	0	0	0	0
Gunes	1	1	1	1	1	1	1	1	1
Guven	1	0	1	1	0	1	1	0	0
Hur	0	0	0	0	0	0	0	0	0
Ihlas	0	0	0	0	0	0	0	0	0
Isık	0	0	0	0	0	0	0	0	0
Isvıcre	1	1	1	1	1	1	1	1	1
Koc Allianz	1	1	1	1	1	1	1	1	1
Ray	1	1	1	1	1	1	1	1	1
Seker	0	0	0	0	0	0	0	0	0
T.Genel	1	0	0	1	1	0	1	1	0
Teb	0	0	0	0	0	0	0	0	0
Toprak	0	0	0	0	0	0	0	0	0
Yapı Kredi	1	1	1	1	1	1	1	1	1
Faulty Cell Number		**6**	**4**		**1**	**2**		**2**	**4**

When a classification is made by considering the data of 2005 year, it is estimated that 22 of 24 companies (92%) are in correct classification and 2 (8%) of them is in faulty classification. When the companies above the median are classified as successful and ones below the median as unsuccessful, it is estimated that 22 of 24 companies (92%) are in correct classification and 2 of them (8%) are in faulty classification (in classification made by considering the data of 2004 year). When a classification is made by considering the data of 2005 year, it is estimated that 20 of 24 companies (83%) are in correct classification and 4 (17%) of them is in faulty classification. 17 of 24 companies (71%) in Group II are classified correct with all five methods.

4 Conclusion

In this study, the applicability of the logistic regression in classification of general sizes of insurance companies in Turkey, establishment of future oriented strategies and correct recognition of strategies of companies is tried to be demonstrated. 41 companies that have sufficient data from 53 insurance companies in Turkey were examined by means of the logistic regression to categorize whether they are being successful and unsuccessful. In examination, firstly, the companies were listed in order according to their success percentages when compared to their previous year successes (for 2004 and 2005 years), the first 20 companies were categorized as successful and others as unsuccessful and the logistic regression was applied. The logistic regression equations had sufficiency in classification of Group I companies and high sufficiency in classification of Group II companies. Besides, from the logistic regression equations estimated for 2004 and 2005 years, it was examined whether a correct classification was made for the first 6-month real data of 2004 year and it was determined that the models are valid in great extent. On the other hand, it was determined that the policy numbers of companies could be used as a dependent variable. For policy numbers, the geometric mean and median was used in categorical separation and the logistic regression equations were found for 2004 and 2005 years. It was seen that the logistic regression equations formed a suitable separation for Group I and Group II companies. Besides, for the first 6-month real data of 2004 year, estimations were made from the logistic regression models of 2004 and 2005 years and it was compared that whether the real classifications and estimated classifications of the first 6-month real data of 2006 year were the same. As a result of that comparison, it was determined that the estimations confirmed the same separation in high probability. In conclusion, the logistic regression is a good method in categorizing of insurance companies as successful and unsuccessful.

References

[1]Weissert, W. G. & Elston, J. M. & Koch, G. G., *Risk of Institutionalization***,** 1977-1985, http://aspe.hhs.gov /daltcp/reports/instrkes.pdf, 4, 2006.

[2]King, Gary. & Zeng, Langche.; *Logistic Regression in Rare Events Data***,** http://gking.harvard.edu/files/0s. pdf, 137, 2006

[3]Williamson, J. B. & McNamara T. K., *Why Some Workers Remain in The Labor Force Beyond the Typical Retirement Age*, http://www.bc.edu/centers/crr/papers/ wp_ 2001-09.pdf, 6, 2006.

[4]Branıv, R.& Agarwal, A. & Leach, R., *Predicting Bankruptcy Resolution*, http:// bear.cba.ufl.edu/ agarwal/ vita/PredictingBRResolutionJBFA.pdf, 6, 2006.

[5]Fisher, R. & Campbell, J., *Health Insurance Estimates for States*, http:// www.Census.gov/hhes/www /saipe/ asapaper/fishercampbellasa02.pdf, 1-2, 2006.

[6]Temple, J., *Explaining the private health insurance coverage for older Australians*, http://www.elecpress. monash.edu.au/pnp/cart/download/free.php ?paper=178, 17, 2006.
[7] Sullivan, K., *Transportation & Work: Exploring Car Usage and Employment Outcome in The LSAL Data***,** http://www.ncsall.net/fileadmin/ resources/ research /op_ sullivan.pdf**,** 4-5, 2006.
[8]Lee, T. & Yeh,Y. & Liu, R., *Can Corporate Governance Variables Enhance the Prediction Power of Accounting-Based Financial Distress Prediction Models?*, http://cei.ier.hit-u.ac.jp/ working/2003/2003 WorkingPapers /wp2003-14.pdf, 4, 2006.
[9]Capella, M. E. & McDonnall, *Predictors of Competitive Employment for Blind and Visually Impaired Consumers of Vocational Rehabilitation Services* http://eric.ed.gov/ERICDocs/ data ericdocs2/ content_storage_01/ 0000000b/ 80/ 2c/ 7d/ 1a.pdf, 6, 2006.
[10] Chandra, A. & Samwick, A. A., *Disability Risk and The Value of Disability Insurance*, http://www.nber.org /papers/w11605, 3-4, 2006.
[11]Galiano, M. M. & Christmann, A., *Insurance: An R-Program to Model Insurance Data*, http://www.sfb475 .uni-dortmund.de / berichte / tr49-04.pdf, 5-8, 2006.
[12]Cornfield, J., *Joint Dependence of the Risk of Coronary Heart Disease on Serum Cholesterol and Sistolic Blood Pressure: A Diskriminant Function Analysis*, Federation Proceeding, 21:58-61, 1962.
[13]Robert, G.& Rao, N. K. & Kumar, S., *Logistic Regression Analysis of Sample Data*, Biometrika, 35, 58: 79th Survey, 1987.
[14]Duffy, D. E., *On Continuity-corrected Residuals in Logistic Regression*, Biometrica 77: 287-293, 1990
[15] Tatlıdil, H., *Uygulamalı Cok Degiskenli Istatistiksel Analiz*, Ankara, 225-232, 1992.
[16]Menard, S., *Applied Logistic Regression Analysis*, Sage Pub. USA, 37-42, 1995.
[17]http://www.tsrsb.org.tr/tsrsb/Istatistikler/Genel+ Sektör +verileri/Türk+sigorta+sektörü+verileri/
[18]Agresti, A., *Categorical Data Analysis*, Wiley Pub., Florida, 165-257, 2002.
[19] Ozdamar, K., *Paket Programlar ile Istatistiksel Veri Analizi*, Kaan Kitabevi, Eskisehir, 601-608, 2004.
[20]Weisberg, S., *Applied Linear Regression*, Wiley Pub., Canada, 255-265, 2005.
[21]Albayrak, A. S., *Uygulamalı Cok Degiskenli Istatistik Teknikleri*, Asil Yayin, Ankara, 439-462, 2006.
[22]Hair, S. F.& Anderson, R. E. & Tahtam, R. L. & Black, W. C., *Mutivariate Data Analysis*, Prentice-Hall, NJ, 276-281, 314-321, 1998.