***Turkish Studies*** *- International Periodical For The Languages, Literature and History of Turkish or Turkic*
*Volume 9/2 Winter 2014, p. 1517-1527, ANKARA-TURKEY*

# CLASSIFICATION PERFORMANCE COMPARISON OF DISCRIMINANT ANALYSIS AND LOGISTIC REGRESSION ANALYSIS*

*Necla TEKTAŞ***

## ABSTRACT

Classification satisfaction analysis as one of the multivariate statistic methods is used widely in various areas such as performance evaluation. Classification techniques are also divided into two groups depending on whether the classes are known a priori or not. If the classes are not known a priori, multidimensional scaling analysis and cluster analysis are used; otherwise discriminant analysis and logistic regression analysis are utilized. In this study, since groups are known a priori, discriminant analysis and logistic regression analysis are used to determine the rate of satisfaction. Application data are collected from students of Marmara University Technical Sciences Vocational School Electronics and Automation Department through surveys in order to identify their level of satisfaction with their department. The survey used for data collection consists of two sections. In the first part of the survey, there are questions related to the demographic characteristics of the students. In the second part, there are 52 Likert-scale statements targeted at determining their level of satisfaction. A Factor Analysis was implemented on the Likert-scale questions contained in the second part of the survey in order to determine the primary factors influencing the students' satisfaction levels. At the end of this analysis, four sub-factors were obtained. After that, a variable was obtained for every factor found by using SPSS 17 package software. These variables were used as independent variables in the analyses to be conducted. The dependent variable on the other hand was obtained as a dual variable on the basis of answers given to the statement: 'Are you happy to study at the Electronics and Automation Department? (Yes/No)'. To be able to determine the rate of satisfaction, data are examined and compared through the use of Discriminant analysis and logistic regression analysis. **Discriminant analysis is proven to be a better tool in determining the level of satisfaction.**

**Key Words:** Logistic Regression Analysis, Discriminant Analysis, Satisfaction, Factor Analysis, Frequency

---

*__Bu makale Crosscheck sistemi tarafından taranmış ve bu sistem sonuçlarına göre orijinal bir makale olduğu tespit edilmiştir.__

** Yrd. Doç. Dr. Marmara Üniversitesi Teknik Bilimler Meslek Yüksekokulu, El-mek: ntektas@marmara.edu.tr

# DISKRIMINANT ANALIZI VE LOJISTIK REGRESYON ANALIZININ SINIFLANDIRMA PERFORMANSLARININ KARŞILAŞTIRILMASI


## ÖZET

Çok değişkenli istatistik yöntemlerinden biri olan sınıflandırma; memnuniyet analizi, performans değerlendirilmesi gibi pekçok farklı alanda kullanılmaktadır. Sınıfların önceden bilinen gruplar olması veya önceden grupların bilinmemesi durumuna göre sınıflandırma teknikleri kendi içlerinde ikiye ayrılmaktadır. Sınıfların önceden bilinmemesi durumuna göre sınıflandırmada çok boyutlu ölçekleme analizi ve kümeleme analizi kullanılırken, sınıfların önceden bilinmesi durumunda ise diskriminant analizi ve lojistik regresyon analizi kullanılmaktadır. Bu çalışmada gruplar önceden bilindiği için memnuniyet oranını belirlemede diskriminant analizi ve lojistik regresyon analizi ele alınmıştır. Uygulama verisi olarak Marmara Üniversitesi Teknik Bilimler Meslek Yüksekokulu Elektronik ve Otomasyon Bölümünde öğrenim gören öğrencilerin bölümlerinden memnun olup/olmama düzeylerini belirlemek amacıyla bu bölümde öğrenim gören öğrencilerden anket aracılığı ile toplanmıştır. Veri toplama aracı olarak kullanılan anket iki bölümden oluşmaktadır. Anketin birinci bölümünde öğrencilere ait demografik özelliklerine ilişkin sorular bulunmaktadır. İkinci bölümde memnuniyeti belirlemek amacıyla 52 Likert ölçekli ifadeden oluşmaktadır.Anketin ikinci bölümünde yer alan Likert ölçekli sorulara memnuniyete etki eden belli baslı faktörlerin neler olduğunu tespit etmek için Faktör Analizi uygulanmıstır. Bu analizin sonucunda dört alt faktör elde edilmiştir. Bundan sonra bulunan her bir faktör için SPSS 17 paket programı aracılığıyla birer degisken elde edilmistir. Bu değişkenler yapılacak analizlerde bağımsız değişkenler olarak kullanılmıştır. Bağımlı değişken ise anketteki memnuniyeti ölçmeye yönelik 'Elektronik ve Otomasyon Bölümünde okumaktan memnun musunuz? (Evet/Hayır)' ifadesine verilen yanıtlardan yararlanılarak ikili bir değişken olarak elde edilmiştir. Memnuniyet oranını belirlemek için veriler Diskriminant analizi ve lojistik regresyon analizi ile incelenerek karşılaştırılmıştır. **Diskriminant analizinin memnuniyet oranını belirlemede daha iyi sonuç verdiği görülmüştür.**


**Anahtar Kelimeler:** Lojistik regresyon analizi, Diskriminant Analizi, Memnuniyet, Faktör Analizi, Frekans

## Introduction

In multivariate analysis, classification (grouping) is one of the most common problems the researcher faces. This study researches bundling of data from different groups and deciding the allocation of groups while measuring k number of properties of n number of data with p number of variables. In other words, the primary question in classification problem is: "which probability distribution do the observation values of p number of variables come from". From this perspective, classification problem is a statistical decision-making process. In this process, the researcher should decide on which group the individual comes from. In some cases, groups' probability distribution

and the parameters of these distributions are known. However in practice, usually each group is assumed to have a distribution correlated to the variable p and the parameters of this distribution are predicted through a chosen sample. Then, the researcher tries to solve the decision making problem. In this stage, there are two decision making topics for the researcher. One is to identify the variables that have a distinguishing effect by examining the discriminative properties, and the other is to classify the individuals into groups with the aid of these discriminative functions.

When classification methods are reviewed, it is possible to see two groups being formed. Classification techniques are also divided into two groups depending on whether the classes are known a priori or not. If the classes are not known a priori, multidimensional scaling analysis and cluster analysis are used; otherwise discriminant analysis and logistic regression analysis are utilized.

In this study, in order to determine the level of satisfaction through the survey, students are asked whether they are happy to attend Marmara University Technical Sciences Vocational School, Electronics and Automation Department, and classes are identified in advance based on their answers. Since classes are known a priori, the results are compared using the Discriminant Analysis and Logistic Regression Analysis methods.

**Discriminant Analysis**

The researcher, while trying to identify which group the individual belongs, assigns the individual to the appropriate group by using the variable p. In this stage, it is safe to say that each variable affects the decision on assignment. Another aim of the researcher might be to identify which variable or variables have more impact on classification of individuals. Hence, variables that cause different grouping of individuals can be identified.

While identifying the variables that affect both the classification and the group allocation, discriminant functions are defined that are in turn functions of p number of variables. Through these functions, it becomes possible to classify the individuals or identify the variables that affect the allocation [1].

Then, the purposes of the discriminant analysis can be summarized under two headings:

(1) To determine the discriminant functions and identify the most discriminative variables in group allocation.

(2) In cases of a unit of unknown origin, to decide which group it will be included in [2].

Discriminant function is created to maximize the discrimination between individuals. For this purpose, the equation

$$(\mathbf{W}^{-1}\mathbf{B} - \lambda\mathbf{I})\mathbf{V} = \mathbf{0} \qquad (1)$$

is examined. Here, W stands for within-group sum-of -squares matrix while B stands for between-group sum-of squares matrix. Solving the equation no. (1) means finding out eigenvalues and eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$. λ values obtained from here indicate eigenvalues and **V** the eigenvectors [3]. Using these eigenvectors, discriminant variables can be calculated as in equation no. (2):

$$Y_1 = v_{1,1}X_1 + v_{1,2}X_2 + \ldots\ldots + v_{1,p}X_p$$

$$Y_2 = v_{2,1}X_1 + v_{2,2}X_2 + \ldots\ldots + v_{2,p}X_p$$

. . (2)

. .

. .

$$Y_s = v_{s,1}X_1 + v_{s,2}X_2 + ..... + v_{s,p}X_p$$

In this equation $\mathbf{Y}_j$ denotes j_th (j=1,2,…,s and s=min(p,g–1)) discriminant variable; $v_{i,j}$ is the i_th variable's weight in j_th discriminant variable; and $\mathbf{X_i}$ refers to the observation eigenvector of i_th (i=1,2,…,p) variable [3]. Individual's group origin can be determined by using the discriminant variables in equation no (2). For this, Bayesian methods can be utilized.

The application steps of discriminant analysis are as follows:

(1) A priori group memberships are identified.

(2) If any, the differences between the groups for variables are determined by using Wilks' Λ statistic. After the MANOVA test is performed for this purpose, if there is a significant difference between the groups, analysis resumes. If there is no significant difference, the average of all groups is considered to be equal and hence the groups as identical. In this case, no discriminant analysis is performed.

(3) Variables are selected. In selection of variables, a priori knowledge or statistical methods can be applied.

(4) An examination is conducted to determine whether there are multiple correlations between the variables. To this end, combined within-group correlation matrix is reviewed. If the absolute value of the correlation in this matrix is greater than 75%, then some of the variables should be eliminated [4]. The variable set is determined at the end of this step.

(5) The eigenvalues of W−1B matrix and the eigenvectors of these eigenvalues are identified. These eigenvectors give the required weights for discriminant functions. Significance test of discriminant functions is also performed by using these eigenvalues. If any of the functions is significant, then allocation is considered successful.

(6) By using non-standardized discriminant function, discriminant function values are obtained for each individual. These values will later be used in the classification stage.

(7) A priori possibilities are identified for group memberships. Then using these probabilities and discriminant scores, a posteriori probabilities are attained. The highest a posteriori probability of the individual is identified. The group that gives this probability is considered the group where the individual belongs, thus the individual is classified.

(8) After each individual is classified, the success of the discriminant function can be determined by examining the percentage of accurate classification [5].

**Logistic Regression Analysis**

Let's assume that the number of individuals examined is n; explanatory variable vector is $Y = (Y_1, Y_2, ..., Y_n)'$; parameter vector is $\beta = (\beta_0, \beta_1, ..., \beta_P)'$, and error term vector is $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)'$. Taking p as the number of explanatory variables, explanatory variables matrix for a model that contains the invariable is indicated with an **X**. In this case, explanatory variable matrix is as shown in (3).

$$X = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} & \dots & X_{P,1} \\ 1 & X_{1,2} & X_{2,2} & \dots & X_{P,2} \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & X_{1,n} & X_{2,n} & \dots & X_{P,n} \end{bmatrix} \quad (3)$$

In the explanatory variable matrix, each column is composed of n number of observations of the explanatory variable. In this case, linear regression model can be shown as (4).

$$Y = X\beta + \varepsilon \quad (4)$$

In model (4), whether the explanatory variables are discrete or continuous does not affect the method used in prediction of the model or the parameter predictions obtained through this method. This is why explanatory variables fed to the model can have both discrete and continuous nature. On the other hand, whether the dependent variable in the model has a discrete nature has a considerable impact. Specifically, let us suppose that dependent variable is a Bernoulli variable that can assume two values as 0 and 1, and the probability of assuming these values is fixed. This type of dependent variable can be used frequently. Common examples include whether to consume a specific product, having a habit, satisfaction from a job, financial strength of a bank etc. This type of a dependent variable can be defined as shown in (5):

$$Y_i = \begin{cases} 1, & \text{if has a propertyrelated to i\_th individual} \\ 0, & \text{if does not have a propertyrelated to i\_th individual} \end{cases} \quad (5)$$

Let us show the probability of this variable assuming the value of 1 with $p_i$, and not assuming with $1-p_i$. In Logistic Regression Analysis, instead of predicting the parameters from the regression equation in model (4), $p_i$ probabilities are supposed to conform to the logistic distribution given in (6) [6].

$$p = G(X\beta) = \frac{e^{X\beta}}{1+e^{X\beta}} = \frac{1}{1+e^{-X\beta}} \quad (6)$$

It is difficult to attain the prediction of **β** in the form given in (6). This is because the expression (6) is not linear. When this equation is linearized, we obtain expression no. (7):

$$Logit_i = \ln\left(\frac{p_i}{1-p_i}\right) = \ln(e^{X_i\beta}) = X_i\beta \quad (7)$$

Since this equation is linear in terms of its parameters, prediction becomes easier. Parameter predictions can be obtained by using the likelihood function for this model and Newton-Raphson method [7]. Then, probability of success is predicted for each individual. If this probability is higher than 0.5, then $\hat{Y} = 1$ for that individual, and if lower, $\hat{Y} = 0$ **[6]**.

The application steps of logistic regression analysis are as follows:

(1) A priori group memberships are identified.

(2) The variables that will be fed into the model are identified. For this purpose, a priori information or statistical techniques may be used.

(3) The parameters of the model are predicted using the Newton-Raphson method. Then, the significance of the entire model is tested using the likelihood ratio. If the model is not

significant, the analysis ends. If the model is found to be significant, the analysis proceeds to the next stage.

(4) The individual significance of the predicted model parameters is then examined. For this purpose, the likelihood ratio or Wald statistical model may be used. After the significance of each coefficient is examined, their individuality ratios are checked and the effects of the explanatory variables on dependent variables may be interpreted.

(5) Using the predicted model parameters, it becomes possible to predict which observation comes from which group.

(6) To examine the goodness of fit of the model, the correct classification percentage and artificial $R^2$ criteria are used. If the goodness of fit of the model is at an acceptable level, the group predictions at stage 5 may be used. Otherwise, analysis proceeds to the 2nd stage where the variables that will be fed into the model are reviewed and the processes are repeated [5].

**Methods and Findings**

The data used in the research were collected using the simple random sampling method via face-to-face surveys with 163 freshman and sophomore students attending the Biomedical Device Technology, Electronic Technology and Electronic Communication Technology programs at Marmara University, Technical Sciences Vocational School, Electronics and Automation Department in the spring term of the 2011-2012 academic year.

The survey used for data collection consists of two sections. In the first part of the survey, there are questions related to the demographic characteristics of the students. In the second part, there are 52 Likert-scale statements targeted at determining their level of satisfaction.

Distribution of surveyed students as per the variable of gender: 135 (82.8%) were male and 28 (17.2%) were female. Since our students mostly composed of vocational high school graduates, the number of male students is higher. As for their distribution per the program variable, 61 (37.4%) were in the Electronic Technology, 42 (25.8%) in the Biomedical Device Technology and 60 (36.8%) in the Electronic Communication Technology programs. 80 (49.1%) of the students that participated the survey were freshmen while 83 (50.9%) of them sophomore.

**Identification of Factors Influencing the Level of Satisfaction**

A Factor Analysis was implemented on the Likert-scale questions contained in the second part of the survey in order to determine the primary factors influencing the students' satisfaction levels. As a result of the Kaiser Meyer Olkin sampling adequacy (0.85) and Barlett's globality tests (Chi-square =4560.246 and p=0.000), the Factor Analysis performed was considered suitable. According to the Reliability Analysis result, the Reliability (Cronbach Alpha) coefficient measuring the consistency of the scale was 0.949. During the analysis, the factors with an eigenvalue higher than 1 were selected and the Varimax Rotation was applied. The four factors obtained account for 62.87% of the total variability. At the end of this analysis, four sub-factors were obtained. These sub-factors were named: '$X_1$= Satisfaction with Academic Staff and Counseling', '$X_2$= Satisfaction with Education and Training', '$X_3$= Satisfaction with Physical Conditions' and '$X_4$= Satisfaction with Administration'. After that, a variable was obtained for every factor found by using SPSS 17 package software. These variables were used as independent variables in the analyses to be conducted. The dependent variable on the other hand was obtained as a dual variable on the basis of answers given to the statement: 'Are you happy to study at the Electronics and Automation Department? (Yes/No)'.

**Discriminant Analysis Application**

The Discriminant Analysis is a method that enables the variables in an X data set to be divided into two or more real groups, checks p number of properties of the units to derive functions that will enable these units to be assigned to their real groups and classes in the natural environment at an optimal level [8]. The Discriminant functions derived via the Discriminant analysis consists of linear components of their explanatory variables. In our study, the four variables obtained as a result of factor analysis were taken as explanatory variables and their effect on the satisfaction with the department was found with the Discriminant Analysis. As part of the Discriminant Analysis, firstly, the hypothesis on the equivalence of group covariance matrixes was tested. For this purpose, Box's M test was conducted and it was found that the covariance matrixes were homogeneous (Box's M = 5.358, p=0.888). In the Discriminant Analysis, a canonic discriminant function that enhances the ability to divide into groups and minimizes the Wilks' Lambda was obtained. The eigenvalue that shows the discrimination characteristic of discriminant functions is 0.107. The function accounts for 100% of the total variance. The Wilks' Lambda value showing the statistical significance of the Discriminant function found here is 0.904 (Chi-square= 16.29, df = 4, p<0.003). The lower this value, the more the discriminant power of the functions increases. The canonical discriminant function is specified below in the equation no. (8):

$$Y = 0{,}652 + 0{,}435X_1 - 1{,},546X_2 + 0{,}866X_3 + 0{,}733X_4 \qquad (8)$$

In this equation the denotations are as follows:

*Y*: 1 - satisfied, 0 - dissatisfied,

*X1*: Satisfaction with the Academic Staff and Counseling,

*X2*: Satisfaction with the Education and Training Services,

*X3*: Satisfaction with the Physical Conditions,

*X4*: Satisfaction with the Administrative Services.

As for the assessment of the magnitude of coefficients in order to compare the effects of variables on the satisfaction, the variable that had the highest effect on satisfaction was "Satisfaction with the Education and Training Services" followed by "Satisfaction with the Physical Conditions" in the second place and "Satisfaction with the Administrative Services" in the third place. It was found that the variable with the lowest effect on satisfaction was "Satisfaction with the Academic Staff and Counseling".

Table 1: Performances of the Discriminant Analysis Models in the Data Sets it was Obtained and its Validity Tested

| | | Satisfied | | Dissatisfied | | Total | Average Accuracy (%) |
|---|---|---|---|---|---|---|---|
| Data set | Observed | n | % | n | % | | |
| Original | Satisfied | 95 | 69.9 | 41 | 30.1 | 136 | 69.3 |
| | Dissatisfied | 9 | 33.3 | 18 | 66.7 | 27 | |
| Cross-validate | Satisfied | 93 | 68.4 | 43 | 31.6 | 136 | 67.5 |
| | Dissatisfied | 10 | 37 | 17 | 63 | 27 | |

When Table 1 is reviewed, 18 out of 27 dissatisfied students (66.7%) were assigned to the correct group and 9 (33.3%) were classified incorrectly in the data set (n=163) from which the function was obtained as a result of the Discriminant Analysis. 95 out of 136 satisfied students (69.9%) were assigned to the correct group whereas 41 (30.1%) were classified erroneously. In the data set in which the validity of the function was tested, 17 of the 30 dissatisfied students (63%) were assigned to the correct group whereas 10 of them (37%) were classified incorrectly. While 93 (68.4%) out of 136 satisfied students in the same group were assigned to the correct group, 43 (31.6%) out of them were assigned to the wrong one. In the data set used for developing the discriminant function model, the rate of correct grouping was 69.3% (n=163) on the average whereas the rate of correct grouping in the data set used for testing the validity of the model was 67.5% (n=163).

**Logistic Regression Analysis Application**

In order to find the effect of the satisfaction on the student's overall content, the effect of four factors obtained with the Factor Analysis (independent variables) on the satisfaction contentment (binary dependent variable) were found using the Logistic Regression Analysis. The dependent variable is the natural logarithm of the satisfaction probability of the student. The Logistic Regression model can be expressed as follows:

$$\ln\left[\frac{P_M}{1-P_M}\right]=\sum_{i=1}^{4}\beta_i \ln X_i \qquad (9)$$

In the equation no. (9), the denotations are as follows:

$P_M$: Probability of the Student's Satisfaction with the Service Offered,

$X_i$: Independent Variables,

$\beta_i$: Predicted Parameters,

Ln: Natural Algorithm.

In this case, the logistic regression model can be expressed as in (10):

$\ln Y= \beta_0+ \beta_1\ln X_1+ \beta_2\ln X_2+ \beta_3\ln X_3+ \beta_4\ln X_4+\varepsilon$ (10)

In the equation no.(10), the denotations are as follows:

$Y$: 1 - satisfied, 0 - dissatisfied,

$X_1$: Satisfaction with the Academic Staff and Counseling,

$X_2$: Satisfaction with the Education and Training Services,

$X_3$: Satisfaction with the Physical Conditions,

$X_4$: Satisfaction with the Administrative Services.

$\beta_0$: Fixed Term,

$\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$: Predicted Parameters,

$\varepsilon$: Error Term

ln: Natural Algorithm.

Table 2: Coefficients of the Logistic Regression Model

| | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|
| Satisfaction with the Academic Staff and Counseling | .831 | .416 | .006 | 1 | .040 | 1.032 |
| Satisfaction with the Education and Training | 1.896 | .548 | 11.968 | 1 | .001 | 1.150 |
| Satisfaction with the Physical Conditions | 1.050 | .477 | 4.842 | 1 | .028 | 2.859 |
| Satisfaction with the Administration | .875 | .445 | 3.864 | 1 | .049 | 2.399 |
| Constant | 0.614 | .314 | 3.958 | 1 | .047 | .073 |

$$\ln Y = 2{,}614 + 0{,}831 \ln X_1 + 1{,}896 \ln X_2 + 1{,}050 + \ln X_3 + 0{,}875 \ln X_4 \qquad (11)$$

The results of the Logistic Regression Analysis can be seen in Table 2. The Hosmer and Lemeshow test is used in order to test the compliance of the created Logistic Regression Model with the data [9]. The fact that the significance level of this test is higher than 0.05, indicates that the model was compliant with the data. In our study, this statistical value was found as p=0.653 (Chi-square=5.95, df=8); therefore, it was found that the model created was compliant with the data. The coefficients of four independent variables are also significant ($p<0.05$).

This model can be interpreted as follows: when the $X_2$, $X_3$ and $X_4$ are fixed, the increase in $X_1$ "Satisfaction with the Academic Staff and Counseling" by one unit increases the natural logarithm of the satisfaction probability of the student by 0.831 units. A similar comment can also be made for $X_2$, $X_3$ and $X_4$. For this reason, the coefficients found with the Logistic Regression analysis can be interpreted as the amount each independent variable would affect the student satisfaction. The fact that the coefficient is large shows that it has a higher effect on the student satisfaction. The results were found to be as follows: the variable that had the highest effect on the student satisfaction was "Satisfaction with Education and Training" followed by "Satisfaction with Physical Conditions" in the second place and "Satisfaction with the Administration" in the third place and the variable that had the lowest effect was found to be "Satisfaction with the Academic Staff and Counseling".

Table 3: Performances of the Discriminant Analysis Model in the Data Sets it was Obtained and its Validity Tested

| | | Satisfied | | Dissatisfied | | Total | Average Accuracy (%) |
|---|---|---|---|---|---|---|---|
| Data set | Observed | n | % | n | % | | |
| Original | Satisfied | 95 | 69.9 | 41 | 30,1 | 136 | 69.3 |
| | Dissatisfied | 9 | 33.3 | 18 | 66.7 | 27 | |
| Cross-validate | Satisfied | 135 | 85 | 1 | 25 | 136 | 84.7 |
| | Dissatisfied | 24 | 15 | 3 | 75 | 27 | |

When Table 3 is reviewed, 18 out of 27 dissatisfied students (66.7%) were assigned to the correct group and 9 (33.3%) were classified incorrectly in the data set (n=163) from which the function was obtained as a result of the Logistic Regression Analysis. 95 out of 136 satisfied students (69.9%) were assigned to the correct group whereas 41 (30.1%) were classified erroneously. In the data set in which the validity of the function was tested, 3 of the 27 dissatisfied students (75%) were assigned to the correct group whereas 24 of them (15%) were classified incorrectly. While 135 (85%) out of 136 satisfied students in the same group were assigned to the correct group, 1 (25%) out of them were assigned to the wrong one. In the data set used for developing the Logistic Regression Analysis model, the rate of correct grouping was 69.3% on the average whereas the rate of correct grouping in the data set used for testing the validity of the model was 84.7%.

**Conclusion**

In this study comparing the performances of the Logistic Regression Analysis and the Discriminant Analysis, the best average prediction ratios were obtained with the Discriminant Analysis in both the data sets where the model was derived and the data sets where the validity of the model was tested. While the Discriminant Analysis had a correct prediction rate of 67.5% in the data set in which the model validity was tested, this rate was 84.7% in the Logistic Regression Analysis. As a result of the application of the Discriminant Analysis on the data set, the relative effect of each variable on the student satisfaction was found. As can be seen in Table 4, the results were found to be as follows: the variable that had the highest effect on the student satisfaction was "Satisfaction with Education and Training" (100%) followed by "Satisfaction with Physical Conditions" in the second place (56%) and "Satisfaction with the Administration" in the third place (47%) and the variable that had the lowest effect was found to be "Satisfaction with the Academic Staff and Counseling" (23%). In order to compare these results with the results of the Discriminant Analysis and the Logistic Regression Analysis, the coefficients derived from these analyses were converted into relative magnitudes by assigning 100% to the magnitude of the most effective variable; the results can be seen in Table 4.

Table 4: Relative Comparison of the Effect of Independent Variables on the Satisfaction on Department based on the Techniques used in the Model

| | Discriminant Analysis | Logistic Regression Analysis |
|---|---|---|
| Satisfaction with the Academic Staff and Counseling | 28 | 43 |
| Satisfaction with the Education and Training | 100 | 100 |
| Satisfaction with the Physical Conditions | 56 | 55 |
| Satisfaction with the Administration | 47 | 46 |

Important conclusions can be drawn from Table 4. In both analysis techniques applied here, the variable with the highest effect on the student satisfaction was found to be "Satisfaction with the Education and Training". The ranking of the remaining three variables is same in both the Discriminant Analysis and the Logistic Regression Analysis; "Satisfaction with the Physical Conditions" is in the second place, "Satisfaction with the Administration" in the third place and finally "Satisfaction with the Academic Staff and Consulting" comes in the last place.

In conclusion, the Discriminant Analysis not only enabled better prediction rates, but also revealed the real relation between the variables better as compared to the other technique, the

Logistic Regression Analysis. For that reason, Discriminant Analysis can be recommended for both researchers and practitioners in addition to the conventional statistical methods for similar studies.

As in all research studies, this study also has its limitations. The results obtained are only valid for the data set used and no generalizations can be made. To be able to make generalizations, studies with numerous data sets should be conducted.

**REFERENCES**

[1] KSHIRSAGAR, A. M., *Multivariate Analysis*, Marcel Dekker Inc., 1999, New York, pp.194.

[2] ÜNSAL, A., *Diskriminant Analizi ve Uygulaması Üzerine Bir Örnek*, Gazi Üniversitesi, İ.İ.B.F. Dergisi, 2000, Vol. , pp. 19-36.

[3] TATSUOKA, M. M., *Multivariate Analysis: Techniques for Educational and Psychological Research*, 1971, New York: John Wiley & Sons, Inc., pp. 156-161.

[4] ÖZGÜR, E., *Çok Değişkenli İstatistiksel Analiz Teknikleri ve Bir Uygulama*, Yayınlanmamış Doktora Tezi, Danışman: Prof. Dr. Aydın ÜNSAL, Gazi Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri A.B.D.,Ankara, 2003, pp. 132-133.

[5] ÜNSAL,A. & Güler, H., *Türk Bankacılık Sektörünün Lojistik ve Diskriminant Analizi ile İncelenmesi*, VII. Ulusal Ekonometri ve İstatistik Sempozyumu, İstanbul, 2005,pp.2-5

[6] WOOLDRIDGE, J. M., *Introductory Econometrics: A Modern Approach, 2nd ed*., Thomson, 2003, Ohio: South-Western, pp.554-562

[7] MADDALA, G. S., *Limited-Dependent and Qualitative Variables in Econometrics*, 1999, New York: Cambridge University Press, pp. 24-25.

[8] ÖZDAMAR, K., *Paket Programlar ile İstatistiksel Veri Analizi*, 5.Baskı, Kaan Kitabevi, 2004, Eskisehir, pp. 355.

[9] HOSMER David W. & LEMESHOW Stanley, *Applied Logistic Regression*, John Wiley and Sons, 2000, New York.,pp. 156.