A STATISTICAL CLASSIFICATION STUDY OF COUNTRIES' HUMAN DEVELOPMENT LEVEL BY DISCRIMINANT ANALYSIS

Serhat BURMAOĞLU [*]
Erkan OKTAY [**]

Alınış Tarihi: 14 Şubat 2012 **Kabul Tarihi: 20 Ağustos 2015**

Abstract: The classification done by Human Development Index comes into prominence through countries, which take into consideration the level of development instead of economic growth. The United Nations Development Programme has been classifying countries by using this index since 1990. The aims of this study are to determine the importance of the variables used for preparing Human Development Index and by developing the discriminant function, providing classification with fewer variables for the future. In this analysis, the classification of the Human Development Index by United Nations Development Programme (UNDP) is examined and necessary transformations for ensuring a discriminant analysis of the examined data are made. The obtained variables are then used in a discriminant analysis. One discriminant function is constructed since only very developed and mid-developed country-groups are analyzed. As a result, a function with a high classification success of 92.5% is obtained. Interpretation of the coefficients of variables involved in the function and the effect of variables on classification have been analyzed.

Keywords: Human Development, Discriminant Analysis, Classification.

DİSKRİMİNANT ANALİZİ YARDIMIYLA İNSANİ GELİŞMİŞLİK SEVİYELERİNE GÖRE ÜLKELERİN İSTATİSTİKSEL SINIFLANDIRILMASI

Öz: İnsani Gelişmişlik İndeksi'ne göre ülkelerin sınıflandırılmasında iktisadi büyüme yerine gelişmişlik düzeyinin dikkate alınması ön plana çıkmaktadır. Birleşmiş Milletler Kalkınma Programı, bu indeksi kullanarak ülkeleri 1990 yılından beri sınıflandırmaktadır. Bu çalışmanın amacı, İnsani Gelişmişlik İndeksi'nde kullanılan değişkenlerin önemini tespit edebilmek ve bir diskriminant fonksiyonu geliştirerek ileriki sınıflandırmalar için daha az değişken kullanılarak bir sınıflandırılmanın yapılmasını sağlamaktır. Bu analizde, Birleşmiş Milletler Kalkınma Programı'nın (UNDP) sınıflandırılması dikkate alınmış ve diskriminant analizi için verilerde gerekli dönüşümler yapılmıştır. Elde edilen veriler diskriminant analizi için kullanılmıştır. Analizde yalnızca çok gelişmiş ve orta düzeyde gelişmiş ülke grupları kullanıldığı için tek bir diskriminant fonksiyonu oluşturulmuştur. Sonuç olarak, % 92,5 gibi yüksek bir sınıflama başarısına sahip bir diskriminant fonksiyonu elde edilmiştir. Söz konusu fonksiyonda kullanılan değişkenler yorumlanarak bu değişkenlerin yapılan sınıflamaya etkileri analiz edilmiştir.

Anahtar Kelimeler: İnsani Gelişmişlik, Diskriminant Analizi, Sınıflama.

I. Introduction

Countries have researched for years, attempting to ensure predetermined economic goals in order to improve living standards and sustain the welfare and a peaceful environment. In order to ensure these goals, research

[*] Doç. Dr. İzmir Kâtip Çelebi Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Sağlık Kurumları Yönetimi Bölümü
[**] Prof. Dr. Atatürk Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü

has shown that merely focusing on economic factors aimed at improving the welfare environment are not enough since such factors are limited to the field of finance.

To reach only economic goals are called as growth in field of finance. Economic growth is linked to development in production factors which provides more real wage year by year. Development, on the other hand, not only improves the production and the per-capita real revenue, but also changes and restores the socio-cultural structure in less-developed country.

As can be understood from these definitions, growth is not only a concept for less-developed countries; but is also relevant for developed countries.

According to Gaspar (1995, 208), growth is the realization of potential for improving monetary welfare, and diminishing human suffering. Another definition suggests that growth should include nutrition, health, shelter, employment, physical environment, socio-cultural environment, involvement in decision-making, human reputation, sense of belonging, and other variables (Bhanoji, 1991: 1452).

Human development is defined as the development process of people's choices by the 1990 United Nations Development Program, Human Development Report. Economic, social, political, and cultural activities providing human development are accepted as dimensions of human development. However, the development in three major fields, which allows people to expand their choices, is accepted as the essential indicators of human development. These essential indicators are income, education, and health and nutrition (UNDP, 1990: 9).

The primary objective of human development is to improve the humanitarian choices and make growth more democratic and sharing-based. Humanitarian choices include indicators like improving income, employment opportunities, education, health, and a clean and secure physical environment. Moreover, with these indicators, sharing in the decision-making process and economic and political freedom are emphasized (UNDP, 1991: 7).

In 1994, the Human Development Report, growth was defined once more in broader terms. The essential objective of human development is the creation of an appropriate atmosphere and opportunities for improving the current and prospective potential of people. The Human Development Process is not only relevant to the best improvement of people's potential, but also relevant to the most appropriate use of provided potential in economic, social, political, and cultural fields (UNDP, 1994: 13).

Four different development indexes have been used by the United Nations since 1990. These indexes are the Human Development Index, the Development Index Based on Gender, the Index of Gender Reinforcement, and the Human Poverty Index. Different calculation methods and variables are used in these four indexes.

Only the Human Development Index data are used in this study. The Human Development Index was first put forward for consideration by United Nations Development Programme in 1990 and is constituted of three major indicators. Revenue is described as per-capita GDP according to purchasing power parity; life expectation is described as the expectation at birth, and education is described as the rate of mature literates and school attendance.

According to the value of the Human Development Index in the Human Development Report, countries are classified based on human development. These countries are also grouped according to the index values (UNDP, 1996: 136).

In subsequent years, the Human Development Index was evaluated by its handling of different topics. Some of these topics represented are climate changes, cultural identity, water problems, and democratization (UNDP, 2004; UNDP, 2005; UNDP, 2006; UNDP, 2007/2008).

The aim of this study is to evolve a linear model for classification based on the indexes made by United Nations by reclassifying through the use of a discriminant analysis. Additionally, it will consider whether the variables contribute to the model or not and through this, the ways of classification with minimum variables and minimum costs are to be investigated. As a result, a model with a high classification success is suggested for the human development classification. It also submits a proposal for medium-developed countries in fields which would get better results as they develop. The study consists of five sections. After the introduction, the application section discusses the discriminant analysis as an analysis method. In the third section, classifications existing in literature and forecast methods are mentioned. In the fourth section, the arrangement and analysis of data is provided through a quadratic discriminant analysis, and a stepwise discriminant analysis is made to identify the high rates of discrimination of variables. In the last section, the results are interpreted.

II. Discriminant Analysis

Discriminant analysis is a statistical technique, which allows a researcher to study the differences between two or more sample groups with various variables at the same time. Generally, mathematical equations are utilized in grouping units and these equations, known as a discriminant function are used in order to determine the groups' mutual features so as to allow a determination of the most similar groups. The characteristics used for distinguishing the groups are called discriminant variables. In short, discriminant analysis is a transaction that presents the differences between two or more groups by using discriminant variables. It is a broad term that consists of some mutually relevant statistical approach (Klecka, 1980).

Discriminant functions, obtained through discriminant analysis, are formed from linear components of forecast variables. Discriminant functions reveal which forecast variables affect the differences between inter-groups. These variables, which affect the differences between inter-groups, are called discriminant variables. Another function of discriminant analysis is to determine the group of any unit if the group from which it comes is unknown.

Discriminant analysis also determines in which variables the differences intensify most and states factors, which are active in differentiating groups. The comparison between the classification, with respect to analysis results, and original group memberships gives an opportunity whether the function is sufficient or not.

Discriminant analysis, as in the MANOVA method, is a method which aims to develop a discriminating criterion that is different from the common mean compared to the groups' means. Therefore, in order to apply discriminant analysis to the data sets, these data sets should carry the following assumptions.

- X data matrix should fit multivariable normal distribution.
- The variance and the covariance of the variables should be homogenous. The variables that exists in X matrix should be the samples drawn from the multivariable main body having common covariance matrix.
- There should not be any correlation between the variables' means and variances.
- There should not be multicollinearity between the variables.

X matrix should not include unnecessary variables that do not act a part in discriminating the groups from each other, and should include the right and necessary variables, which provide discrimination of groups from each other.

Some researchers describe discriminant analysis methods with other additional words, aiming to indicate methods like Fisher's Linear Discriminant Analysis, Kernel Based Discriminant Analysis, Maximum Likelihood Discriminant Analysis, Bayesian Discriminant Analysis, and Laplacian Linear Discriminant Analysis (Tang et al, 2005; Liang and Shi, 2004; Lu et al, 2005; Zheng, 2005; Srivastava and et al, 2007). In this study, since quadratic discriminant analysis is used, only the mathematical foundations of this subject will be mentioned without reference to other methods.

Discriminant function is known as a powerful tool in preventing abnormality; however, it cannot be used in oblique distribution. Alternative functions are used in situations where these assumptions are deteriorated. Quadratic discriminant function is a function in which the data is distributed normally and it is used when the groups' variance-covariance matrices are different. The assumption of equality of covariance matrices is a seldom-observed situation (Lachenbruch, 1975: 20).

In quadratic discriminant analysis, in calculating the coefficients, instead of common covariance matrix (S), the differences of groups' covariance matrices are used.

$$Q(x)=\frac{1}{2}\log\frac{|S_j|}{|S_i|}-\frac{1}{2}(x^{-(i)}S_i^{-1}x^{-(i)}-x^{-(j)}S_j^{-1}x^{-(j)}+x(S_i^{-1}x^{-(i)}-S_j^{-1}x^{-(j)}))-\frac{1}{2}x(S_i^{-1}-S_j^{-1})x \quad (1)$$

Initially this function is developed for two groups but is also used by twos for multi-group situations. In the function, S_i and S_j are the covariance matrices regarding i.th and j.th groups, respectively. If $S_i=S_j=S$, the quadratic function will be the same with linear function.

The misclassification possibility in this method, where if the value of function $Q(x)\geq 0$ then the individual is classified in the R_i region, and if not then classified in the R_j region, can be calculated as the following:

$$R_{Q(x)}=\left[1+\exp(Q(x)-\log(\hat{q}_j/\hat{q}_i))\right]^{-1} \quad (2)$$

In the case that covariance matrices are not equal, in addition to the previous transactions, if ($\Sigma_1\neq\Sigma_2$), classification regions R_1 and R_2 are calculated so that:

$$k=\frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right)+\frac{1}{2}(\mu_1{}'\Sigma_1^{-1}\mu_1-\mu_2{}'\Sigma_2^{-1}\mu_2) \quad \text{in for}$$

$$R_1=-\frac{1}{2}x'(\Sigma_1^{-1}-\Sigma_2^{-1})x+(\mu_1{}'\Sigma_1^{-1}-\mu_2{}'\Sigma_2^{-1})x-k\geq\ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad (3)$$

$$R_2=-\frac{1}{2}x'(\Sigma_1^{-1}-\Sigma_2^{-1})x+(\mu_1{}'\Sigma_1^{-1}-\mu_2{}'\Sigma_2^{-1})x-k<\ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad (4)$$

Classification regions are defined as X's quadratic function. In the case that covariance matrices are equal, $\Sigma_1=\Sigma_2$, so $-\frac{1}{2}x'(\Sigma_1^{-1}-\Sigma_2^{-1})x$ quadratic term will disappear and classification regions can be calculated as when the covariance matrices are equal .

If π_1 and π_2 groups have multivariable normal density function and average and covariance matrices are accepted as $\mu_1;\Sigma_1$ and $\mu_2;\Sigma_2$, allocating x_0 to π_1 group is possible if following is provided: Otherwise, x_0 will be allocated to π_2 group.

$$-\frac{1}{2}x_0'(\Sigma_1^{-1}-\Sigma_2^{-1})x_0+(\mu_1{}'\Sigma_1^{-1}-\mu_2{}'\Sigma_2^{-1})x_0-k\geq\ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad (5)$$

III. The Scientific Studies Used in Classification and Estimation Methods

It has been explained in the second section that discriminant analysis is used for classification and estimation. In this section, the applications of discriminant analysis will be briefly expressed.

In their study Balcaen and Ooghe (2005) studied statistical techniques, which have been used over the last 35 years, for the classifications of failures in business life and the problems related to these techniques. They compared multiple discriminant analysis, logit models, conditional probability models and one-variable analysis methods in the study.

Sueyoshi (2004) examined classification performances of discriminant analysis with standard integer programming models and with two-phased integer programming models. He applied these models on the data obtained from Japanese banks.

Berg (2007) attempted to calculate the bankruptcy estimations of companies by using linear discriminant analysis, generalized linear models and artificial neural networks.

Çilan et al (2009) examined the digital distinction between those who are members of European Union and those who are not, with discriminant analysis. The classification performance was considered successful with 74.1% in the analysis they made. Prior to the analysis normality assumption was tested.

Bosse (2008) modeled the discrimination of credibility of small enterprises, with multiple discriminant analysis and achieved a classification performance of 86.6%.

Wu et al (2008) made an analysis of financial problems of Chinese public enterprises by using stochastic artificial neural networks and discriminant analysis. With multiple discriminant analysis, he obtained a classification performance of 81.25% for short-term estimations, and 56.25% for long-term estimations. On the other hand, the analysis made by artificial neural networks showed a classification performance of 87.5% for short-term estimations, and 81.25% for long-term estimations.

Chen et al (2008) examined the criteria defined for the selection communication tool with discriminant analysis and studied the relationship between variables by developing a model for estimation.

Pompe and Bilderberd (2005) utilized the discriminant model developed with the multiple discriminant analysis to estimate the bankruptcy of small and medium sized enterprises.

Malhotra and Malhotra (2003) used discriminant analysis and artificial neural networks for evaluation of the liabilities of clients. They observed the five samples used in the application and noted that the artificial neural networks were more successful in classification.

Cheng and Titterington (1994) compared various artificial neural networks and statistical methods. They showed that there are strong relations between feed forward neural networks and discriminant analysis, and logistical regression.

Odom and Sharda (1990) compared the estimation ability of artificial neural networks and multivariable discriminant analysis. It was determined that artificial neural networks showed better performance in the analysis method made for a diminishing sample size.

Leshno and Spector (1996) compared the estimation ability of artificial neural networks with linear discriminant analysis and quadratic discriminant analysis. It was observed that the chosen artificial neural networks models yielded more accurate results than the classic discriminant analysis models.

Lee et al (2005) used artificial neural networks to estimate the number of bankruptcies of Korean companies. Discriminant analysis and logistical regression were compared in terms of estimation accuracy. Provided that network, used with back propagation algorithm, provides the best solution if its target vector is definite and even if the sample size decreases.

IV. Research

Except for UNDP classification works, making a classification by using the developmental data with statistical methods is not shown in literature in which is the most important contribution of this study. Another important contribution of this study is to show the mid-developed countries which variable is comparatively more influential for them and to help them to draw a roadmap to improve in Human development.

A. The Scope of the Research

In this research, the discriminant analysis is made by using the values and results obtained from 155 countries. When the variable values are considered, 35 countries of 155 were not included in the process and the classification process was performed as shown in Table 1.

Table 1: *Case Processing Summary*

Unweighted Cases		N	Percent
Valid		120	77.4
Excluded	Missing or out-of-range group codes	0	.0
	At least one missing discriminating variable	35	22.6
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	.0
	Total	35	22.6
Total		155	100.0

In this analysis, very developed (1) and mid-developed (2) country classification is used as dependent variables. Initially, the analysis started with 28 variables; however, by using the omitted 12 variables, the number of countries falls to 74, so an attempt was made to analyze 120 countries by omitting these variables. Since revenue, education, and life expectation are used as sine qua non of the Human Development Report, it is decided that no more elimination can be made to 16 variables so all 16 variables were decided to be used in this analysis. The independent variables are shown in Table 2.

Table 2: *Selected HDI Variables*

	Label	N
Total Population (2005)	IV1	154
Rural Population (2005)	IV2	155
Seat in Parliament Held by Women (Percent of Total)	IV3	150
Public Health Expenditure (% of GDP)(2007)	IV4	154
Private Health Expenditure (% of GDP) (2007)	IV5	154
Health Expenditure PPP (PPP US$)(2007)	IV6	152
Life Expectancy at Birth (2002-2005)	IV7	151
Net Enrollment Rate to Primary School	IV8	141
Telephone mainlines (per 1000 people) (2005)	IV9	152
Cellular subscribers (per 1000 people) (2005)	IV10	154
Internet users (per 1000 people) (2005)	IV11	153
GDP (Billion $) (2005)	IV12	152
Exports (% of GDP) (2005)	IV13	146
Imports (% of GDP) (2005)	IV14	147
Electricitiy Consumption (Kw-H)(2004)	IV15	151
Prison Population (2007)	IV16	155
Valid N		120

The possibilities of countries according to the predetermined groups are shown in Table 3. There are 120 countries in the sample, of which 56 countries are very developed and 64 countries are mid-developed.

Table 3: *Prior Probabilities of Groups*

Countries' Classification of Development	Prior	Cases Used in Analysis	
		Unweighted	Weighted
High Development	.467	56	56
Medium Development	.533	64	64
Total	1.000	120	120

The possibilities of the groups are established as 0.467 for very developed countries and 0.533 for mid-developed countries. These possibilities will be used in evaluating the classification rate. Before starting the results of the analysis, the assumption of normality, the assumption of equality of covariance matrices, and the assumption of multicollinearity are examined, and then the classification analysis is covered.

B. Testing Assumptions

As mentioned in the literature concerning discriminant analysis, three very important assumptions were identified prior to the analysis, and the analysis is not made due to obtained values, or is continued by using different methods. Normality, equality of covariance matrices, and multicollinearity assumptions lead these assumptions. The literature expresses that if the assumptions are not satisfied, this will pose a problem in terms of classification results, and desired high rates of classification can not be achieved.

In the case that covariance matrices are equal, linear discriminant analysis is made; however in the case that covariance matrices are not equal, quadratic discriminant analysis is made to obtain the classification results.

(1) The Assumption of Normality

Sharma (1996, 380-382) studied the multivariate normality and mentioned that if main mass is normal and sample size is (n > 25), then the Mahalanobis distance is distributed with Chi-square distribution. It is studied that whether all of the variables were normally distributed or not, and identified that some of the investigated variables were not normally distributed, in order to normalize these variables, the logarithm of variables were taken and stated that they were normalized. The variables, which are not initially distributed normally according to Kolmogorov-Smirnov test and normalized by logarithm, are:

- Total Population (LogIV1)
- Health Expenditures PPP (PPP US$)(2007) (LogIV6)
- GDP (Billion Dollars) (LogIV12)
- Exports (% of GDP) (LogIV13)
- Imports (% of GDP) (LogIV14)
- Prison Population (LogIV16)
- Electricity consumption (kW-H) (LogIV15)
- Telephone mainlines (per 1000 people) (2005) (LogIV9)
- Internet users (per 1000 people) (2005) (LogIV11)

The analysis results, based on the specified expression can be seen in Table 4. The correlation values between the Inverse Cumulative Chi-Square value with the Mahalanobis distance is 0.979 and is significant at 0.01 significance level. The same situation also can be seen in the scatter diagram in Figure 1.

Table 4: *Correlation Analysis of Mahalanobis Distances and Chi-Square Values*

		Mahalanobis Distances	Chi-Square Value
Mahalanobis Distances	Pearson Correlation	1.000	.979**
	Sig. (2-tailed)		.000
	N	120.000	120
Chi-Square Value	Pearson Correlation	.979**	1.000
	Sig. (2-tailed)	.000	
	N	120	120.000
**. Correlation is significant at the 0.01 level (2-tailed).			

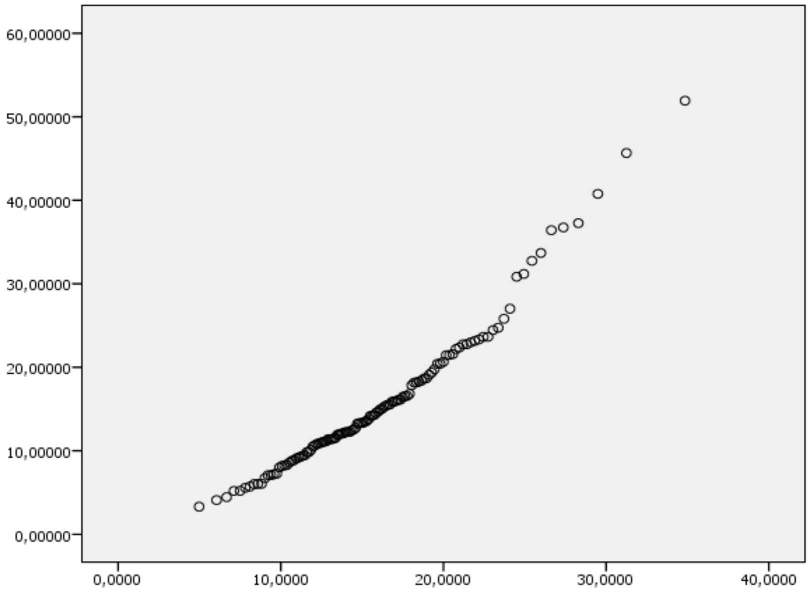


Figure 1: *Scatter Diagram of Mahalanobis Distances and Chi-Square Values*

It is desired that the correlation value between the Inverse Chi-square values and the cumulative Mahalanobis distances is equal to 1. The resulting correlation value is close to 1, which reminds that the Multivariate Normal Distribution assumptions are satisfied. To test the normality of the correlation coefficient, the corresponding value in the probability table with n = 100 and 0.01 probability is 0.981. The resulting correlation coefficient is very close to this value.

Multivariate normality test can be conducted by analyzing the multivariate outlier unit values of variables as expressed by Kalaycı (2008, 212-214). In examining outlier units, the Mahalanobis distances are used. Calculated distances are divided by the number of variables used in the analysis so that the deviation values are calculated. As noted, these values fit to the t distribution. To describe a unit as outlier, the value of the unit should be significant in 1% level of significance. So, MD^2/sd value must be greater than (t) 5.014. None of the 120 variables used in the analysis are considered to be outlier.

(2) Assumption of Equality of Covariance Matrices

The Box's M test, which tests equality of covariance matrices, was used. When the intra-group covariance matrices (Within Groups) option was used, the equality of covariance matrices could not be achieved. ($P < 0.05$) Therefore, for quadratic discriminant analysis, covariance matrices for separate groups were the option used and a significant value of 0.703 was found. The values are found in Table 5.

Table 5: *Box's M Test Result*

Box's M		.147
F	Approx.	.146
	df1	1.000
	df2	41262.347
	Sig.	.703

Equality of covariance matrices for separate groups is examined with Box's M test. Since the significance value is greater than ($p > 0.05$), the null hypothesis is accepted, as the covariance matrices are equal for separate groups. Equality of covariance matrices between groups could not be achieved so that quadratic values were taken into consideration.

(3) Assumption of Multicollinearity

With the purpose of identifying multi-connection, collinearity was investigated with a linear regression analysis. Table 6 contains the results.

Table 6: *Multicollinearity Test*

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	2.627	.392		6.705	.000		
	IV1	3.104E-5	.000	.007	.114	.909	.750	1.334
	IV2	.001	.002	.029	.360	.719	.411	2.434
	IV3	.000	.003	-.009	-.132	.895	.645	1.550
	IV4	-.028	.026	-.119	-1.090	.278	.228	4.383
	IV5	-.006	.028	-.018	-.226	.822	.447	2.236
	IV6	.000	.000	.285	1.787	.077	.108	9.276
	IV7	-.011	.007	-.169	-1.677	.097	.270	3.710
	IV8	.003	.003	.064	.872	.385	.510	1.962
	IV9	.000	.000	-.355	-2.543	.012	.141	7.104
	IV10	.000	.000	-.393	-3.557	.001	.225	4.449
	IV11	.000	.000	-.089	-.734	.465	.184	5.421
	IV12	7.107E-6	.000	.018	.140	.889	.169	5.921
	IV13	.005	.002	.227	2.020	.046	.218	4.597
	IV14	-.005	.003	-.230	-1.913	.058	.190	5.256
	IV15	-6.240E-6	.000	-.060	-.618	.538	.288	3.468
	IV16	-2,480E-7	,000	-,108	-,937	,351	,207	4,838

In analyze of VIF and tolerance values of multiple linearity tests it is observed that the VIF values were smaller than 10 and tolerance values were greater than 0.30. This situation can be interpreted as multiple linear relationship is not present. In addition, very small values of t are concluded by some authors as one of the reasons for the multi-linearity problem. When Figure 2 is examined, t values are not very close to 0.

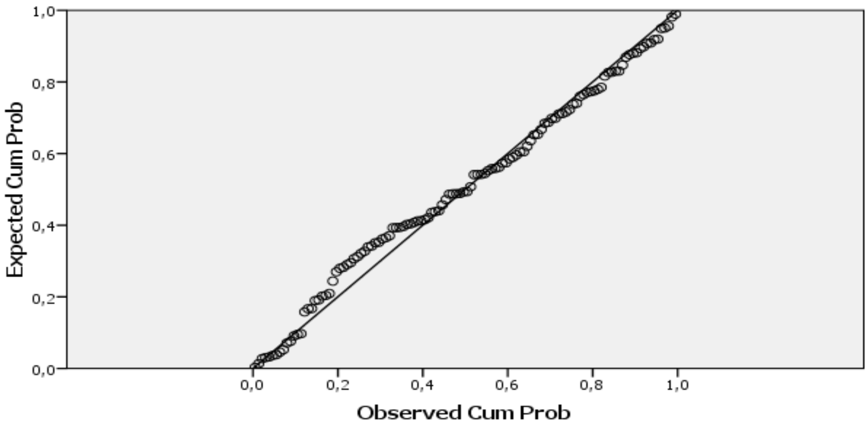


Figure 2: Development Classifications of Countries

Figure 2 demonstrates the visual at which level points at the bottom and top of the linear line deviate from a normal distribution. As shown in Figure 2, regression residual values of development classification variables does not deviate substantially from normality.

C. Research Results

(1) Results of Discriminant Analysis

Initially, since two groups (very developed and mid-developed) have been identified, one discriminant function has been derived. The large eigenvalue, offers that large portion of the variance in the dependent variables are explained by the obtained function. Although it is not an absolute value, the values of more than 0.40 are accepted as good. As shown in Table 7, in the model, the eigenvalue is found at 2.385, and explains the 100% of variance. Moreover, a canonical correlation coefficient of 0.839 is found. The square of the coefficient is 0.704. It can be said that independent variables explain the dependent variables at the rate of 70.4%.

Table 7: *Eigenvalue*

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.385[a]	100.0	100.0	.839

Table 8: *Wilk's Lambda Value*

Test of Function(s)	Wilk's Lambda	Chi-square	Df	Sig.
1	.295	134.142	16	.000

The statistics of Wilk's Lambda shows the parts of total variance of discriminant scores which are unexplained by the differences among the groups. In the model, 0.295, in other words, 29.5% of total variance cannot be explained by the differences among the groups.

Table 9 shows the obtained function and the coefficients of the variables in the function.

Table 9: *Unstandardized Discriminant Function Coefficients*

	Function
	1
LogIV1	.922
LogIV6	2.676
LogIV9	-.528
LogIV11	.160
LogIV12	-.495
LogIV13	-.565
LogIV14	.385
LogIV15	.440
LogIV16	-.424
IV2	-.002
IV3	.001
IV4	-.127
IV5	-.189
IV7	.045
IV8	-.019
IV10	.002
(Constant)	-7.119

The discriminant function, which is used to determine the development level of countries, is constituted to determine Z development level as the following:

$$Z=-7.119+0.922*(LogIV1)+2.676*(LogIV6)-0.002*(IV2)+0.01*(IV3)-0.127*(IV4)-0.189*(IV5)+0.045*(IV7)-0.019*(IV8)-0.528*(IV9)+0.002*(IV10)+0.160*(LogIV11)-0.495*(LogIV12)-0.565*(LogIV13)+0.385*(LogIV14)+0.440*(LogIV15)–0.424*(LogIV16)$$

As can be shown in the model, according to LogIV6-Purchasing Power Parity per Capita, health expenditure creates the biggest impact on the dependent variable by a single unit increase, which creates a positive impact with a value of 2.676. High imports create negative impact as shown in LogIV13, and high exports create a positive impact as shown in LogIV14. Also, electricity consumption (IV15) and communication (LogIV9, LogIV11) create positive impacts. The point which attention should be given is that high-developed countries are shown as 1 and mid-developed countries are shown as 2. In other words, since the variables, which create positive impacts increase the points of the dependent variable, these variables classify the country among the mid-developed countries.

With this point of view, when the coefficients of the variables are handled, in order to be a high-developed country 7-negative-coefficient-variables should be increased.

Moreover, the average discriminant function of each group is shown in Table 10. According to the Table 10, the mean value of first group and second group's distance to the function are 1,637 and –1,433 respectively.

Table 10: Distances of Group Means to Discriminant Function

Development Classification of Countries	Function
	1
High developed	1.637
Medium developed	-1.433

In the discriminant analysis, the effects on function are analyzed by a structural matrix.

Table 11: *Structure Matrix*

	Function
	1
IV10	.838
LogIV6	.817
LogIV15	.689
LogIV9	.637
LogIV11	.637
IV7	.598
IV2	.427
IV4	.421
LogIV12	.414
IV8	.276
IV3	.248
LogIV14	.147
IV5	-.123
LogIV16	.057
LogIV13	-.045
LogIV1	.011

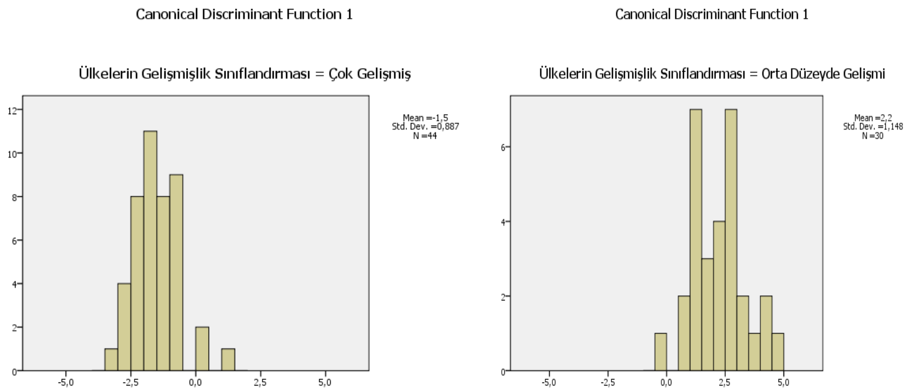
The structural matrix, which is shown in Table 11, is used in determining the importance of independent variables and shows the correlation of each variable with the discriminant function. When the following variables of the model are analyzed; number of subscriptions to mobile phone per 1000 people (IV10), per capita health expenditures according to purchasing power parity (LogIV6), electricity consumption (LogIV15), life expectation at birth (IV7), number of telephone lines per 1000 people (LogIV9), and internet line for 1000 people (LogIV11), it is found that these variables have a strong correlation with function in the model. Categorically, health, energy, and communication variables are more effective in classification and discrimination. Economic variables lag behind these variables, which can be evaluated as an important outcome.

Table 12: *Classification Results*

		Countries' Development Classification	Predicted Group Membership		
			High Developed	Medium Developed	Total
Original	Count	High Developed	51	5	56
		Medium Developed	4	60	64
	%	High Developed	91.1	8.9	100.0
		Medium Developed	6.2	93.8	100.0

As shown in Table 12, the model is successful in classifying with a 92.5% of classification rate. However, in order to test the accuracy of this classification, relative chance criterion and maximum chance criterion should be calculated and then compared. The sample size is 120. Therefore, the high-developed group constitutes 59% of the sample, and the mid-developed constitutes 41%. Chance value is the selection probability of the high-developed group, that is 0.59, and the selection probability of the mid-developed group, that is, 0.41. Here, the maximum chance criterion is 0.59. The relative chance criterion is $(0.59)^2+(0.41)^2=0.5162$. The classification rate, which is obtained by discriminant analysis, is much greater than that value.

In the classification, there are 9 countries, which are classified wrongly. Uruguay, Mexico, Panama, Belarus, and Albania are found to be mid-developed by the analysis while they are high developed. Turkey, Colombia, Tunisia, and Jamaica are found as high developed while they are mid-developed. The rankings of the variables are the development rankings identified by United Nations Development Program. Misclassified countries are between the very developed and mid-developed countries, so this shows the similarity of the classification with the United Nations Development Program's classification. In the analysis, Turkey is identified as a very developed country according to the values of the independent variables; however, it is identified as a mid-developed country by the United Nations Development Program.



(a) (b)

Figure 3: *High developed (a) and medium developed (b) countries' distribution*

(4) Results of the Stepwise Discriminant Analysis

A stepwise discriminant analysis was done to identify which variables are more effective in discrimination. Therefore, instead of showing the topic in detail as when a quadratic discriminant analysis is carried out, the variables, which stayed in the model and the classification, success of these variables is emphasized.

In the Stepwise discriminant analysis, Wilk's method, Mahalanobis method, and the method of the smallest F rate were used. Table 13 shows the variables included in the model and the coefficients of the model.

Table 13: *Comparison of a Stepwise Discriminant Analysis*

	Mahalanobis Method	The Smallest F Rate Method	Wilks Method
	Function	Function	Function
IV4	-.158	-.158	-.158
IV5	-.238	-.238	-.238
LogIV6	2.744	2.744	2.744
IV10	.001	.001	.001
(Constant)	-6.641	-6.641	-6.641

As shown in Table 13, the groups are classified with 4 variables in the model. Generally, communication and health variables constitute an important part. Although public and private health expenditures (IV4, IV5) create a negative impact, individual health expenditure is the most important deterministic variable.

According to this model, all methods were observed to be successful at the rate of 90.7%, which is also remarkable. When the Casewise values were analyzed, in all three methods, Turkey was observed to be classified among the high developed countries.

V. Results

Classification is done using a variety of methods in research in social sciences, and statistical inferences are put forward with respect to the results of such classifications. The data set used is that used by United Nations for ranking of high and mid developed countries in terms of Human Development Index. In this study, raw data, made up of index values, were used instead of index values, and the countries were re-classified.

The assumption of normality, the assumption of the equality of the covariance matrix and the assumption of multicollinearity were tested in discriminant analysis. In testing of the assumption of normality, normality transformations were made due to the fact that some variables did not prove a one-variable normal distribution, and this revealed that these variables were normally distributed with a multi-variable in the tests made afterwards. A quadratic discriminant analysis was used instead of linear discriminant analysis because the assumption of equality of the covariance matrix could not be satisfied. As a result of the analysis the nonexistence of the multicolllinearity was shown.

As a result of the discriminant analysis, it was observed that independent variables were able to explain dependent variables with a rate of 70.39 %. In an evaluation of the importance of the independent variables used, it was determined that communication data, life expectancy at birth, exported goods as a percentage of GDP and public expenditure on health services had positive impact on the ability of a country to become developed. This is considered realistic in determining the development level of countries. After all, it is expected to observe that advanced and productive countries have high export levels, well-established communication systems, and large investments in health care and the eventually high level of life expectancy at birth. Obviously, rural population, imported goods, and the rate of registration to elementary education variables create negative impacts on the model. In the classification of high-developed countries, according to the discriminant analysis results, exports, communication, and health investments are observed as important for investment. In other words, by working together, healthy people can move a country to a very developed level by work aimed at production.

In the classification made by discriminant analysis, a 92.5% classification rate is achieved. Incorrectly classified countries are in between high developed and mid-developed countries, so this shows similarity with the United Nations Development Program's classification. Turkey is one of the incorrectly classified countries. Turkey is identified as a very developed country according to the values of independent variables, but is identified as a mid-developed country by the United Nations Development Program, which is very meaningful.

Discriminant analysis was also made stepwise and the distinguishing variables were examined in details. As a result of this analysis, a model was developed with four variables and it was observed that public and private health expenditures create negative impacts on the model; however, individual health expenditure and mobile phones per 1000 people create positive impacts. The most important aspect, which was pointed out, with its high impact, is individual health expenditure according to the purchasing power per capita.

The classification rates of reclassification of high-developed and medium-developed countries are observed to be very high in this study. It also shows that in both micro and macro level, these methods could be used with the variables emphasized. In following researches, the relationship between the data and variables used in this study could be analyzed with canonical correlation analysis and the relationships between the variables could be more detailed to offer a strategic roadmap to medium-developed countries for improving their human development levels.

References

Balcaen, S., H. Ooghe. 35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems. *The British Accounting Review*. 2006. Vol.38. p. 63-93.

Berg D. Bankruptcy Prediction by Generalized Additive Models. *Applied Stochastic Models in Business and Industry*. 2004. Vol.23. p. 129-143.

Bhanoji, Rao. Human Development Report: 1990 Review Assessment. *World Development*. 1991. Vol.19. p.10.

Bosse, D. A. Bundling Governance Mechanisms to Efficiently Organize Small Firm Loans. *Journal of Business Venturing*. 2008. Vol.24. p. 183-195.

Chen, K., D.C. Yen, S. Hung, A.H. Huang. An Exploratory Study of the Selection of Communication Media: The Relationship between Flow and Communication Outcomes. *Decision Support Systems*. 2008. Vol.45. p. 822-832.

Cheng, B., D.M. Titterington. Neural Networks: A Review from a Statistical Perspective. *Statistical Science*, 1994. Vol. 9(1). p. 2-30.

Çilan, Ç.A., B.A. Bolat, E. Coşkun. Analyzing Digital Divide Within and Between Member and Candidate Countries of European Union, *Government Information Quarterly*, 2009. Vol.26. p.98-105.

Gaspar, Des. *Kalkınma Ahlakı Yeni Bir Alan mı? Piyasa Güçleri ve Küresel Kalkınma*. İngilizceden Çeviren: İdil Eser. Yapı Kredi Publications. 1995. No.2. İstanbul.

Klecka, W. *Discriminant Analysis*. 1980. Sage Publications. London.

Lachenbruch, P.A. *Discriminant Analysis*. 1975. Hafner Press. London.

Lee, K., D. Booth, P. Alam. Comparison of Supervised and Unsupervised Neural Networks in Predicting Bankruptcy of Korean Firms. *Expert Systems with Applications*. 2005. Vol.29 (1). p. 1-16.

Leshno, M., Y. Spector. Neural Network Prediction Analysis: The Bankruptcy Case. *Neurocomputing*. 1996. Vol.10. p. 125-147.

Liang Z., P. Shi. Kernel Discriminant Analysis and Its Theoretical Foundation. *The Journal of the Pattern Recognition Society*. 2004. Vol.38, p.445-447.

Lu, J., K.N. Plataniotis, A.N. Venetsanapoulos, J. Wang. An Efficient Kernel Discriminant Analysis Method. *The Journal of the Pattern Recognition Society*. 2005. Vol.38. p. 1788-1790.

Malhotra, R., D.K. Malhotra. Evaluating Consumer Loans Using Neural Networks. *The International Journal of Management Sciences*. 2003. Vol.31. p. 83-96.

Odom, M., R. Sharda. A Neural Network Model for Bankruptcy Prediction. *Proceedings of the International Joint Conference on Neural Networks*. 1990. Vol.2. p. 163-168.

Pompe, P.P.M., J. Bilderbeek. The Prediction of Bankruptcy of Small-and-Medium Sized Industrial Firms. *Journal of Business Venturing*. 2005. Vol.20. p. 847-868.

Sharma S. . *Applied Multivariate Techniques*. 1996. John Wiley and Sons Inc. Canada.

Srivastava S., M. Gupta, B. Frigyik. Bayesian Quadratic Discriminant Analysis. *Journal of Machine Learning Research*. 2007. Vol.8. p. 1277-1305.

Sueyoshi, T. A Methodological Comparison between Standard and Two Stage Mixed Integer Approaches for Discriminant Analysis. *Asia-Pacific Journal of Operations Research*. 2004. Vol.4. p. 513-528.

Tang H., T. Fang, P. Shi. Laplacian Discriminant Analysis. *The Journal of the Pattern Recognition Society*. 2005. Vol.39. p. 136-139.

UNDP (1990). *Human Development Report 1990*. Oxford University Press. New York.

UNDP (1991). *Human Development Report 1991*. Oxford University Press. New York.

UNDP (1994). *Human Development Report 1994*. Oxford University Press. New York.

UNDP (1996). *Human Development Report 1996*. Oxford University Press. New York.

UNDP (2004). *Human Development Report 2004*. Oxford University Press. New York.

UNDP (2005). *Human Development Report 2005*. Oxford University Press. New York.

UNDP (2006). *Human Development Report 2006*. Oxford University Press. New York.

UNDP (2008). *Human Development Report 2007/2008*. Oxford University Press. New York.

Wu D.D., L. Liang, Y. Zijiang. Analyzing Financial Distress of Chinese Public Companies Using Probabilistic Neural Networks and Multivariate Discriminate Analysis. *Socio Economic Planning Sciences*. 2008. Vol.42. p. 206-220.

Zheng, W. A Note on Kernel Uncorrelated Discriminant Analysis. *The Journal of the Pattern Recognition Society*. 2005. Vol.38. p. 2185-2187.