JOURNAL OF SOCIAL AND HUMANITIES SCIENCES RESEARCH

ISSN 2459-1149 SHSR journal

Economicsand Administration, TourismandTourism Management, History, Culture, Religion, Psychology, Sociology, FineArts, Engineering, Architecture, Language, Literature, EducationalSciences, Pedagogy&OtherDisciplines

2018 Vol:5 / Issue:26

pp.2547-2561

ArticleArrivalDate (Makale Geliş Tarihi) 19.08.2018
ThePublishedRel. Date (Makale Yayın Kabul Tarihi) 27.10.2018
ThePublishedDate (Yayınlanma Tarihi) 28.10.2018

INVESTIGATION OF LIFE SATISFACTION IN OECD COUNTRIES WITH MULTIVARIATE ANALYSIS METHODS

OECD ÜLKELERİNDE YAŞAM MEMNUNİYETİNİN ÇOK DEĞİŞKENLİ ANALİZ YÖNTEMLERİ İLE İNCELENMESİ

Research Assistant Ebru GUNDOGAN AŞIK

Department of Statistics and Computer Science, Karadeniz Technical University, Trabzon, TURKEY, ORCID: 0000-0002-9910-6555

Associate Professor Arzu ALTIN YAVUZ

Department of Statistics, Eskişehir Osmangazi University, Eskişehir, TURKEY, ORCID: 0000-0002-3277-740X

ABSTRACT

The quality of life index is an index used to measure the quality of life for countries. When this index value is calculated, countries are assessed in terms of the multivariate features they have. The significance of the variables studied in the index studies is not generally investigated statistically. In this study, it is aimed to create a classification mechanism with the help of other subindex values which are thought to affect life satisfaction and constitute quality of life. With the help of this classification mechanism, it was also determined whether the statistical effects of variables affecting life satisfaction are significant. There are many techniques for performing the classification process. However, the literature does not focus on the reliability of these techniques. As with all statistical techniques, the validity and reliability of the results obtained from these techniques depend on the provision of the necessary assumptions. For this reason, various classification techniques have been introduced and evaluated in this study. It has been mentioned which one should be preferred according to the situations that can be encountered in practice.

Keywords: Better Life, OECD, Robust, Logistic, Discriminant

ÖZET

Yaşam kalitesi endeksi, ülkelerin yaşam kalite düzeylerini ölçmek için kullanılan bir endekstir. Bu endeks değeri hesaplanırken ülkeler, içinde bulundukları çok değişkenli özellikler bakımından değerlendirilir. Endeks çalışmalarında ele alınan değişkenlerin önemi genellikle istatistiksel olarak araştırılmamaktadır. Bu çalışmada, yaşam memnuniyetini etkilediği düşünülen ve yaşam kalitesini oluşturan diğer alt endeks değerleri yardımıyla bir sınıflandırma mekanizması oluşturulması amaçlanmıştır. Bu sınıflandırma mekanizması yardımıyla yaşam memnuniyetini etkileyen değişkenlerin istatistiksel olarak etkilerinin anlamlı olup olmadığı da tespit edilmiştir. Sınıflandırma işleminin gerçekleştirilebilmesi için birçok teknik bulunmaktadır. Ancak çoğu zaman bu tekniklerin hangi durumlarda güvenilir sonuçlar verdiği konusu üzerinde durulmamaktadır. Bu çalışmada birden fazla sınıflandırma tekniği ele alınmış ve uygulamada karşılaşılabilecek durumlar doğrultusunda hangisinin tercih edilmesi gerektiği konusuna değinilmiştir.

Anahtar Kelimeler: Daha İyi Yaşam, OECD, Sağlam, Lojistik, Diskriminant

1. INTRODUCTION

The quality of life, which is being discussed more and more in the international literature, is in the common interest of many sciences. In particular, social scientists, philosophers, health scientists and politicians have started to deal with the concepts of quality of life and living standards back in the 1960s and 1970s. While the concept of quality of life used to be a problem of countries with unequal income distribution; with the development of the walfare state concept, it became important criteria all over the world.

The World Health Organization (WHO) defines the quality of life as the way people perceive their experiences in relation to their expectations, standards and goals in the context of one's own situation, culture and values system. Life quality can be defined as "a concept that affects the level of personal satisfaction that can be achieved in living conditions, and shows the personal reactions to diseases and physical, mental and social influences of daily life." The quality of life also is explained as the individual fulfilling their wishes, taking advantage of personal development opportunities, participating in various activities, having adequate resources in terms of quality, and thinking that these resources are sufficient (Shin and Johnson, 1978:479).

Abrams (1973:36) described the quality of life as a degree of satisfaction or dissatisfaction with people's various areas of life, and Andrews (1974:281) put it as the link between life and satisfaction. According to Patrick and Erickson (1993:421), life has two dimensions that can be divided into quality and quantity. Life quantity is identified with or difficult biomedical information such as mortality rates or life expectancy. On the other hand, life quality; can not be expressed only by the use of measurable indicators; and it generally defines a subjective assessment that expresses the complex aspects of life. The quality of life is complexly affected by one's physical health, mental state, social relations, level of independence, personal beliefs, and their relations with their surroundings. Quality life for the individual is to provide their needs without difficulty, to establish control over the environment, to make choices with free will, to find opportunities for self improvement and to live a meaningful life (Cilga, 1994:225). The main aim in life quality is to determine how satisfied the individual is from his psychological, social, physical and economic situation. Life standard is a tool for achieving the quality of life goals.

Many different quality of life approaches has emerged as well as different definitions of quality of life, since each researcher tries to explain the quality of life in accordance with their own research focus. The quality of life indicators used in the studies generally has been collected under two areas; objective indicators and subjective indicators. Physical well-being and financial conditions are among the objective indicators of quality of life. Furthermore, economic status of the person, conditions and place of residence and family status are also among the objective indicators. On other part, subjective indicators are associated with satisfaction from life. It is important that person evaluates their life and finds it positive. In this sense; quality of life is a subjective satisfaction or result evaluated by the person (Tüzün and Eker, 2003:4, Tekkanat 2008:69).

The quality of life index (QLI) is calculated and annually published to determine and measure the quality of life of countries. While this index value is calculated, they are assessed in terms of multivariate characteristics. The first studies on the QLI were made by Kasparian and Rolland. The better life concept first appeared in the 2000s. The OECD (2001) presented a report on the better life of the countries. Osberg and Sharpe (2002:294) developed an economically better life index for selected OECD countries. Only economic variables are considered in this study. Carlsson and Hamrin (2002:416) tested the theoretical model of the Life Satisfaction Questionnaire (LSQ), which was developed to assess the life satisfaction of Swedish women with breast cancer, using confirmatory factor analysis using Structural Equation Modeling (SEM). As a result, the result of confirmatory factor analysis shows that the factor structure of the original model is confirmed. Barbato et al. (2004:971) examined the quality of life with the Satisfaction with Life Domains Scale of 40 Italian patients with mental disorders living in the community. As a result of the questionnaire survey, the quality of life was confirmed to be able to give meaningful information about the subjective perception. Gough vd. (2006:21) stated that in better life studies only economic variables could not measure better quality of life for developed countries. Agulanna (2010:288) mentioned the ability to communicate as well as the social life of the people in Africa in his work on the quality of life for human life in African cultures. Kerenyi (2011:522) also worked on better quality of life index for countries. Kasparian and Rolland (2012:2225), in his work on the quality of life for OECD countries, has established an index of better quality of life on the basis of various data from OECD countries. Stevenson (2013:601) deals with better life and income, and has studied life satisfaction. Durand and Smith (2013:15) and Mizobuchi (2013:18) have came up with the better quality of life index available. Stevanovic (2014:1299) proposed sharing of additional data to show that the The Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) authors who study as two-dimensional tools are right. Durand (2015:8) assessed current the status of the better quality of life index. Jovanovic´ (2016:3176) assessed the validity of the Life Satisfaction Scale (SWLS) in adolescents and tested the measurement inversion by sex Thus, the use of SWLS among adolescents has been shown to be supported. It also examines the validity of SWLS and the comparison of adolescents' single-item life satisfaction.

The quality of life index, along with an index covering the countries' general happiness and emotional intensity, shows the measurable components of the countries with their surroundings and the way these components are perceived in terms of these countries. In recent years, a new index has been established to determine the quality of life of a country, including not only GDP but also variables such as health, education, work life, politics, social relations, environment and trust. According to Kasparian and Rolland (2012:2225), variables that should be included in the better life index are housing, income, business, society, education, environment, management, health, life satisfaction, safety and business life balance. For each variable, countries are evaluated with a scale ranging from 0 to 10 (Seker 2010:118). Having said that the variables covered by the better life index are detailed below (OECD 2013:23, Akar 2014:6).

- Housing: gives clues about housing costs in the country.

- Income and prosperity can be expressed as the source of funding for income, current and temporary expenditure. As the income increases in the country, the level of prosperity increases and social status develops.

- Jobs and outcomes is calculated with the proportion of individuals in the countries have a job that provides income. In general, societies with a high level of employment tend to be richer and politically more stable.

- Social communication measures the tendency of individuals in contact with each other. Especially when people spend their time together with their loved ones and families, it affects the quality of life positively.

- Education and skills is a measure of the value that is given to education. Education has a huge impact on the quality of life.

- Environmental conditions are the factors affecting an individual's quality of life. Especially air and water pollution increasingly affects the health of the individual and the country.

- Civil participation and management is related with the policy and voter participation in the country that has a decisive influence on the quality of life. Civil participation significantly affects the quality of life of countries in the field of public services.

- Health status is the most valuable situation affecting the lives of countries and individuals living in the country.

- Life satisfaction generally concerns situations that are good in the life of the societes. The more individuals are happy with their lives the more improve their quality of life is.

- Security refers to the general security situation of the countries. The high crime rate poses a danger to the safety of society. As a result, the quality of life of the countries is reduced.

- Work life balance is about work life responsibilities of the countries. For this reason, balancing in business life maximizes the quality of life of countries.

The quality of life concept and measurements are closely related to government policies in particular. Determination of happiness and pleasing values of the individual is very important for the decisions to be made in these areas and the applications to be achieved in order to reach more effective results.

The aim of this study is to classify OECD countries with multivariate statistical techniques in terms of better quality of life. The Quality of life index gives a ranking of countries. However, only the ranking does not provide the identification of countries that are similar. Detecting the similar countries is important for the policies that countries will make in order to pass to the category of countries with a higher level of life. In other words, when assessing the quality of life of the countries, it is important to determine where they are located in the system of countries as well as their policies. In addition, when the studies related to the index are examined, it is seen that the principal component analysis, which is one of the multivariate statistical methods, has been applied. It is aimed to make a dimension reduction in a multivariate data matrix with the principal component analysis. However, index data can be used not only for size reduction but also for classification. The significance of the variables studied in the index studies is not generally investigated statistically. In this study, it is aimed to create a classification mechanism with the help of other subindex values which are thought to affect life satisfaction and constitute quality of life. With the help of this classification mechanism, it was also determined whether the statistical effects of variables affecting life satisfaction are significant.

In the following sections, multivariate statistical analysis techniques for classification are introduced. The validity and reliability of the results obtained in the research are closely related to the use of accurate scientific methods. Various classification methods that can be applied depending on data structure are discussed in the study. These are logistic regression, robust logistic regression, logistic ridge regression, robust logistic ridge regression and discriminant and robust discriminant analysis. In the result section, relevant classification analyses were applied for OECD countries with better quality of life index values and their performances were compared. The results obtained based on data structure were evaluated and various suggestions were made as well.

2. MATERIAL AND METHODS

Multivariate statistical methods provide the necessary information for the appropriate solution to the structure of the problem being dealt with by determining the structure of the variable set that is formed by more than one variable and transforming it into a simpler form. It also creates statistical techniques that allow to measure and explain relations between variable groups.

This study deals with the classification of OECD countries according to their quality of life index values. There are many techniques for performing the classification. Nonetheless, the literature does not address the issue of the reliability of the results of these techniques. As with all statistical techniques, the validity and reliability of the results obtained from these techniques depend on the provision of the necessary assumptions. For this reason, several classification techniques have been dealt with in this work. Which one should be preferred in terms of situations that may be encountered in practice has been stated as well.

2.1. Discriminant Analysis

Discriminant analysis is one of the multivariate statistical analysis techniques that generally allows units to be assigned to one of a predetermined set of groups. It uses mathematical equations for this. These equations are called as "discriminant function" (Burmaoğlu et al. 2009:25; Klecka 1980:7; Lachenbruch 1975:20). The discriminant function aims to separate the units into different groups with the help of the discriminant function using the data matrix. Discriminant analysis was given by Fisher (1938:378) for two groups and by Rao (1948:172) for multiple groups. Fisher's classical discriminant function gives effective results in the case of equality of variance-covariance matrices under the condition that the independent variable vector comes from the normal distribution. If there are two groups, a discriminant function is obtained. In cases where the assumptions of the discriminant analysis are not valid, using the obtained discriminant functions for classification leads to unrealistic results.

For each group, Fisher's classical discriminant function is defined as in Equation 1.

$$D_{xi} = d_{0i} + d_{1i}X_1 + \cdots + d_{ki}X_k \tag{1}$$

Here, i represents each group, D_{xi} represents the discriminant function for each group, X_k independent variables and d_{ki} discriminant coefficients for each variable (Klecka,1980).

2.2. Robust Discriminant Analysis

In discriminant analysis, as in all statistical analyses, estimates of the mean vector and variance-covariance matrices are obtained from the sample. In data analysis, these observations, which are quite different from the rest of the data and are called outliers, disrupt assumption of the normality. This makes the discriminant functions obtained by Fisher's classical discriminant analysis disfunctional. In this case, the robust discriminant analysis is used to minimizes the effect of outliers.

Many studies have been conducted to obtain the robust case of classical discriminant function. Robust approaches to discriminant analysis are developed by Randles et al. (1978:566) using M estimators, by Chork and Rousseeuw (1992:195) and Hubert and Van Drissen (2004:311) using the minimum covariance determinant estimators, by He and Fung (2000:158) and Croux and Dehon (2001:480) using S estimators. All these studies are supported by simulation studies (Croux et al. 2005, Filzmoser et al. 2006).

In this study, we used the S estimators for the location-scale parameter (mean vector and variance-covariance matrix). On the other hand, the S estimators provide higher stability and convergence. In recent years, the robust discriminant function has begun to be estimated, often using S estimators. For a sample $\{x_1, \ldots, x_n\} \subset \mathbb{R}^{p+1}$, they are defined as couple $(\hat{\mu}, \hat{S})$ which minimizes $|S|$ under the restriction,

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\sqrt{(x_i-\mu)^t S^{-1}(x_i-\mu)}\right)=b \tag{2}$$

Over all possible pairs of $\mu \in \mathbb{R}^{p+1}$ and $S \in PTS(p+1)$ is PTS, then (p+1) show a symmetric positive definite matrix. Here b is a constant value chosen as appropriate for the distribution.

ρ , the Tukey's Biweight function

$$\rho(u)=\min\left(\frac{u^2}{2}-\frac{u^4}{2c^2}+\frac{u^6}{6c^4},\frac{c^2}{6}\right) \tag{3}$$

where c is a tuning constant to achieve the desired value of the breakdown point (Croux and Dehon, 2001:474).

2.3. Logistic Regression Analysis

In logistic regression, a categorial dependent variable Y having N (usually N=2) unique values is regressed on a set of p independent variables $X_1, X_2, \ldots, X_p$. For example, Y may be presence or absence of a disease, condition after surgery, or marital status. Since the names of these partitions are arbitrary, we often refer to them by consecutive numbers. That is, in the discussion below, Y will take on values 1, 2, …, N.

Let

$$X=(X_1, X_2, \ldots, X_p) \tag{4}$$

$$B_n=\begin{pmatrix}\beta_{n1}\\ \vdots\\ \beta_{np}\end{pmatrix} \tag{5}$$

The logistic regression model is given by the N equations

$$\ln\left(\frac{p_n}{p_1}\right) = \ln\left(\frac{P_n}{P_1}\right) + \beta_{n1}X_1 + \beta_{n2}X_2 + \ldots + \beta_{np}X_p \tag{6}$$

$$= \ln\left(\frac{P_n}{P_1}\right) + XB_n \tag{7}$$

Here, p_n is the probability that an individual with values $X_1, X_2, \ldots, X_p$ is in outcome n. That is,

$$p_n = \Pr\left(Y = n \middle| X\right) \tag{8}$$

Usually $X_1 \equiv 1$ (that is, an intercept is included), but this is not necessary.

The quantities $P_1, P_2, \ldots, P_N$ represent the prior probabilities of outcome membership. If these prior probabilities are assumed equal, then the term $\ln(P_n / P_1)$ becomes zero and drops out. If the priors are not assumed equal, they change the values of the intercepts in the logistic regression eqution.

2.4. Robust Logistic Regression Analysis

In the logistic regression model, if outliers are found then problem of wrong classification occurs. In the logistic regression model as the variable is modeled to be in between 0-1, if the likelihood of success is low, the observed situation can be considered successful, whereas if the likelihood of success is high, the situation can be determined as unsuccessful (Yavuzkanat 2013:57). For this reason, as in the discriminant analysis, the robust logistic regression model is used to give more accurate estimates by reducing the effect of outliers. In the robust logistic regression model, estimates are made using the Bianco and Yohai (1996:19) estimator, an M-class estimator.

Bianco-Yohai estimator is expressed as,

$$\beta_n = \arg\min \sum_{i=1}^{n} \phi(Y_i, \pi(x_i^{'}\beta)) \tag{9}$$

where $d_i(\beta)$ the variance of β in $\phi(Y_i, \pi(x_i^{'}\beta)) = \tilde{n}(d_i(\beta)) + \tilde{n}_0(\pi(x_i^{'}\beta))$ function and $\tilde{n}(t)$, restricted function.

2.5. Logistic Ridge Regression Analysis

In the logistic regression model, independent variables are assumed to be unrelated. However, independent variables may not always be unrelated. This is called multicollinearity (Montgomery et al., 2001:212). In the case of multicollinearity, problems arise such that estimators lose their minimum variance.

In the case of multicollinearity, the regression coefficients are ambiguous and the standard errors of these coefficients are infinite. In addition, the variance and covariance of the regression coefficients are increasing, and the R^2 value of the model is quite high. In addition, all or few of the independent variables are partially statistically significant. In the logistic regression model, as in the other models, the problem of multicollinearity arises. Alternative estimators have been proposed to deal with this problem (Arıcan, 2010:53).

One of the methods used in the case of multicollinearity is the ridge regressions. The ridge regression proposed by Hoerl and Kennard (1970:75) is widely used as biased regression estimators.

The logistic ridge regression method proposed by Dufy and Santer (1989:968) is derived from the constrained maximum likelihood method and is expressed as follows;

$$\hat{\beta}_{LRR} = \left[X'WX + k\mathrm{I}\right]^{-1}\left[X'WX\,\hat{\beta}_{ML}\right] \tag{10}$$

In this equation, k is the ridge parameter and is in the range k > 0.

2.6. Robust Logistic Ridge Regression Analysis

Robust logistic ridge regression analysis is similar to the logistic ridge regression analysis. The difference is replacing inital value estimated value of $\hat{\beta}_{ML}, k$ and W. W is ML estimator with those estimated by the WBY estimator. WBY is the weighted Bianco and Yohai estimator.

The robust logistic ridge regression method proposed by Ariffin and Midi (2017:180) is derived from the constrained maximum likelihood method and is expressed as follows;

$$\hat{\beta}_{RLRR} = \left[X'WX + k\mathrm{I}\right]^{-1}\left[X'WX\,\hat{\beta}_{WBY}\right] \tag{11}$$

3. RESULTS

The quality of life index is an index designed to measure the quality of life of countries. When the QLI is determined, some subindex values are calculated. A subindex that constitutes the better QLI is the life satisfaction index. In this study, a classification process was established using the life satisfaction index and the other subindexes.

Life satisfaction was determined as the dependent variable in the following expressions. This variable is coded as two levels.

$$y_i = \begin{cases} 0, & \text{Countries with a life satisfaction index value less than 5/}\textbf{have poor quality of life} \\ 1, & \text{Countries with a life satisfaction index value greater than 5/ }\textbf{have good quality of life} \end{cases}$$

Other variables that make the QLI are considered as independent variables. OECD countries' QLI data were obtained for the year 2016 from the OECD page (http://www.oecdbetterlifeindex.org). SPSS 18 and R programming language are used for analyses.

Descriptive statistics for the independent variables were calculated and shown in Table 1.

Table 1. Descriptive Statistics for Life Satisfaction

Variables	Minimum	Mean	Maximum
Housing	1.600	5.490	7.900
Income	0.400	3.587	10.000
Job	1.400	6.882	9.700
Communication	0.000	6.180	10.000
Education	0.700	6.440	9.300
Environment	2.300	6.737	10.000
Civil Participation	1.300	4.720	8.600
Health	2.900	7.197	9.600
Security	0.000	7.366	9.800
Work Life Balance	0.000	6.976	9.400

First, the discriminant analysis was applied for two group classification cases. It has been determined whether the assumptions of discriminant analysis have been met or not. It was determined that there was an outlier in the data set and robust discriminant analysis was applied.

The results of the discriminant analyses are given in Table 2. Since there are two groups, a discriminant function is derived. The obtained discriminant function was found to be significant. The calculated eigenvalue for the discriminant analysis was found as 2.601. This value explains for 99% of the variance. The canonical correlation coefficient is 0.850. The square of this coefficient (r^2) is 0.723. The independent variables explain the dependent variable by 72%. The correct classification ratio of the discriminant function was calculated as 92.1%.

One of the assumptions of the discriminant analysis is that the variables have a multivariate normal distribution. This assumption is ignored by most researchers. If the assumptions are not provided, it is clear that the results obtained from the applied technique will have no relation to the actual results. The multivariate normality test was applied to determine whether the variables had multivariate normal distribution or not and the results are given in Table 3.

According to Table 3, it is seen that the data matrix does not show multivariate normal distribution ($p < 0.05$). Observations that are far away from the dataset and have different characteristics from the rest of the dataset are also called outliers. Detection of outliers that have an effect on normality should be. The Q-Q plot is drawn for this purpose and is given in Figure 1. Countries identified as outliers in the data matrix in Figure 1 are Greece, Mexico and Turkey.

Table 2. Discriminant and Robust Discriminant Function for Life Satisfaction

Life Satisfaction	**CDF**	**RDF**
Constant	-36.317	28.016
Housing	-0.842	-0.544
Income	-2.797	-0.028
Job	4.787	1.840
Communication	0.412	0.627
Education	-2.007	-1.184
Environment	-0.510	-0.255
Civil Participation	1.594	0.173
Health	5.318	2.137
Security	0.850	0.043
Work Life Balance	4.033	1.213
Correct Classification Percentage	92.1%	94.7%

*CDF: Classic Discriminant Function, RDF: Robust Discriminant Function

Table 3. MV Normality Test

MV Normality		
Slope	**T**	**p**
0.668	10.216	0.000

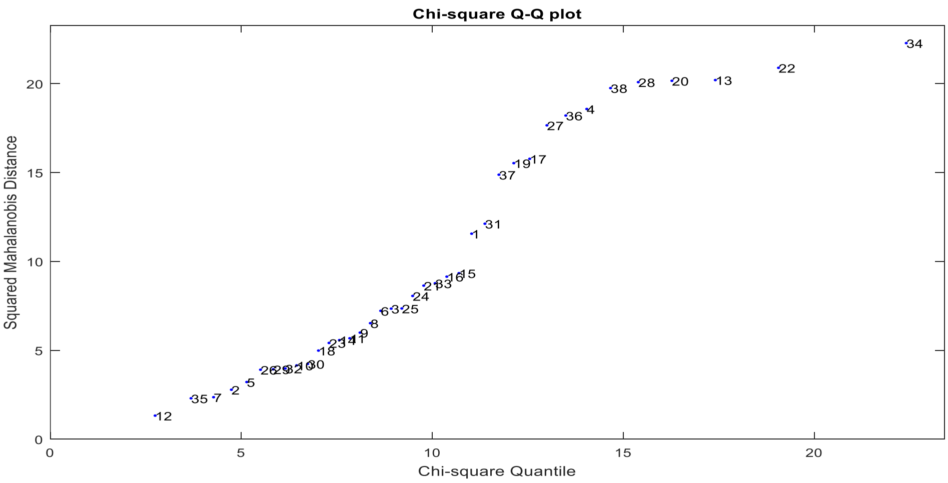

Figure 1. Graph of outlier

It is more useful to apply the robust discriminant analysis here because the data matrix (variables) does not have a multivariate normal distribution and there are outliers in the data matrix. Robust discriminant analysis results are given in the Table 2. The correct classification percentage of the obtained robust discriminant function was found to be 94.7%.

Logistic regression analysis is an alternative to discriminant analysis. Hence logistic regression analysis was applied to the same dataset. As in the discriminant analysis, when the residuals of the logistic regression analysis are examined, outliers are determined. Robust logistic regression was applied to reduce the effect of outliers. Logistic regression analysis results are given in Table 4 and Table 5. When Table 4 is examined, it is seen that the coefficient values obtained for logistic regression analysis is not significant (p> 0.05). Cox & Snell and Nagelkerke R^2 statistics are calculated in Table 5. These values represent the dependent variable explanatory ratio of the independent variables. This ratio is 74% according to the Cox & Snell R^2 statistic and is 100% according to the Nagelkerke R^2 statistic. The Hosmer-Lemeshow test statistic as applied to look at the overall significance of the model, and the p probability value obtained was found to be greater than the significance level of 0.05. Thus, it was decided that the model was generally significant.

Table 4. Logistic Regression, Robust Logistic Regression, Logistic Ridge Regression and Robust Logistic Ridge Regression Analysis Results for Life Satisfaction

Techniques	**LRA**				**RLRA**			
Variables	**Parameter**	**Std. Err.**	**p**	**OR**	**Parameter**	**Std. Err.**	**p**	**OR**
Constant	3433.150	720624.070	0.996	-	3440.438	63.545	**0.000**	-
Housing	-12.290	10785.650	0.999	4.6e-06	-12.319	0.720	**0.000**	4.5e-06
Income	132.660	28655.050	0.996	4.1e+57	132.941	2.558	**0.000**	5.4e+57
Job	-273.460	57788.210	0.996	1.7e-119	-274.042	5.156	**0.000**	9.6e-120
Communication	-132.150	28446.780	0.996	4.1e-58	-132.427	2.564	**0.000**	3.1e-58
Education	114.060	27953.040	0.997	3.4e+49	114.304	2.351	**0.000**	4.4e+49
Environment	-89.110	20601.070	0.997	1.9e-39	-89.303	1.686	**0.000**	1.6e-39
Civil Participation	41.830	10454.280	0.997	1.5e+18	41.922	0.807	**0.000**	1.6e+18
Health	-240.910	53384.370	0.996	2.3e-105	-241.421	4.441	**0.000**	1.4e-105
Security	136.380	32368.390	0.997	1.7e+59	136.667	2.711	**0.000**	2.3e+59
Work Life Balance	-132.440	28741.090	0.996	3.1e-58	-132.721	2.429	**0.000**	2.3e-58
Correct Percentage	-				100.00			
Tecniques	**LRRA**				**RLRRA**			
Variables	**Parameter**	**Std. Err.**	**p**	**OR**	**Parameter**	**Std. Err.**	**p**	**OR**
Constant	2.561	-	-	-	50.569	-	-	-
Housing	-0.073	0.226	**0.003**	0.929	-22.885	0.389	**0.000**	1.1e-10
Income	-0.054	0.228	**0.001**	0.947	4.361	0.224	**0.000**	78.298
Job	-0.066	0.249	**0.002**	0.936	-35.701	0.247	**0.000**	3.1e-16
Communication	-0.054	0.252	**0.003**	0.947	-28.324	0.208	**0.000**	5.0e-13
Education	-0.009	0.249	0.533	0.991	31.375	0.312	**0.000**	4.2e+13
Environment	-0.041	0.248	**0.040**	0.959	4.911	0.246	**0.000**	135.780
Civil Participation	-0.032	0.275	0.182	0.968	3.374	0.218	**0.000**	29.192
Health	-0.096	0.247	**0.000**	0.908	17.935	0.276	**0.000**	6.1e+07
Security	-0.024	0.240	0.126	0.976	5.258	0.278	**0.000**	192.145
Work Life Balance	-0.041	0.266	0.081	0.959	3.930	0.217	**0.000**	50.879
Correct Percentage	68.42				89.90			

*LRA: Logistic Regression Analysis, RLRA: Robust Logistic Regression Analysis, LRRA: Logistic Ridge Regression Analysis, RLRRA: Robust Logistic Ridge Regression Analysis, OR: Odds Ratio and p: Probability of Significance.

When Table 4 is examined, it is seen that the logistic regression analysis has 100% correct classification percentage. In addition, the logistic regression model needs to be revisited because of the reasons such as not significance of model coefficients and general significance of the model. In this case, the presence of outlier(s) in the logistic regression analysis needs to be determined. For this purpose, Box plot with residual values for logistic regression is drawn and given in Figure 2. Finland, Mexico, Italy, Russian Federation and Slovenia have been identified as outliers. Due to the existence of outliers, robust logistic regression analysis was applied instead of logistic regression analysis. Parameter estimation values for robust logistic regression analysis are shown in Table 4. When Table 4 is examined, it is determined that all coefficient values are significant at 0.05 significance level. It has been determined that the standard deviation values obtained for the coefficients are smaller than those found in the logistic regression analysis. Robust logistic regression analysis also seems to have 100% correct classification percentage. When the odds ratios are examined, it is seen that variables of income, education, civil participation and security increase the life satisfaction, whereas the variables of residence, job, housing, environment, work life balance, communication and health have an effect to decrease life satisfaction.

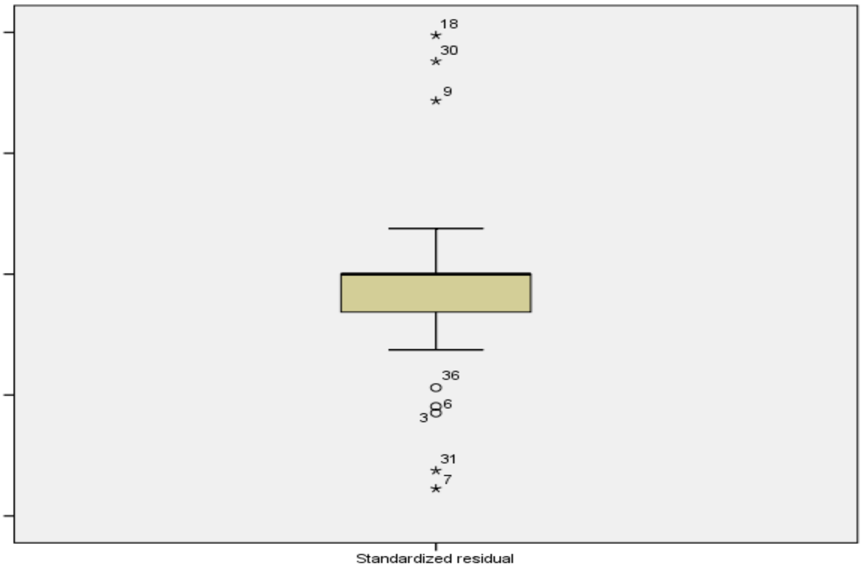

Figure 2. Countries With Outlier as a Result of Logistic Regression Analysis

In the logistic regression, the multicollinearity may provide the magnitude and sign of the estimates of the regression coefficients to be incorrect. In this case, it may lead to wrong results about the relations between dependent and independent variables. When Table 4 is examined, the fact that housing, work and health variables have an effect on reducing life satisfaction and 100% of the correct classification rate increases the doubt of the existence of multicollinearity.

In addition, the logistic regression model was generally found to be significant. When the individual coefficients of the coefficients are tested, it is seen that each coefficient is not significant. Cox & Snell and Nagelkerke R^2 values were found to be very high in Table 5. For these reasons, the existence of multicollinearity has been suspected. Robust logistic regression analysis is a method that reduces the influence of outliers. However, it seems that both the effect of outliers and the problem of multicollinearity needs to be solved for the data set.

Table 5. Logistic Regression Analysis Goodness of Fit Test Results for Life Satisfaction

Goodness of Fit Test	Value
Cox ve Snell R^2	0.738
NagelkerkeR^2	1.000
Hosmer and Lemeshow Test	0.000
Hosmer and Lemeshow p	1.000

Moreover, correlation coefficients between the independent variables are examined. VIF values were obtained. The results are shown in Table 6 and Table 7. When the simple correlation matrix was examined, the simple correlation coefficient between independent variables was found to be highly significant (r>0.75) thus it could cause a multicollinearity problem. It was observed from Table 7 that the VIF values were high.

Table 6. Correlations

	Housing	Income	Job	Commu.	Education	Environment	Civil Par.	Health	Security	Work Life Balance
Housing	1.000	**0.745**	0.645	0.449	0.503	0.561	0.325	0.678	0.635	0.393
Income		1.000	0.664	0.490	0.434	0.455	0.307	0.651	0.568	0.303
Job			1.000	0.383	0.516	0.495	0.280	0.458	0.433	0.181
Commu.				1.000	0.405	0.636	0.190	0.505	0.475	0.482
Education					1.000	0.473	0.114	0.275	**0.764**	0.517
Envinroment						1.000	0.349	0.486	0.503	0.355
Civil Par.							1.000	0.409	0.094	-0.126
Health								1.000	0.521	0.212
Security									1.000	0.444
Work Life Balance										1.000

Table 7. VIF Values

Variables	VIF values
Housing	27.812
Income	24.045
Job	11.360
Communication	79.330
Education	31.589
Environment	25.999
Civil Participation	34.722
Health	70.373
Security	36.304
Work Life Balance	51.327

Logistic ridge regression analysis was applied to the data in the presence of multicollinearity. The parameter estimates are shown in Table 4. When Table 4 is examined, it is seen that coefficient values for housing, income, business, communication, environment and health variables are significant because p probability values are smaller than 0.05 significance level. The correct classification percentage for the logistic ridge regression analysis model was found to be 68.42%. Logistic ridge regression analysis is more reliable than the logistic regression analysis, despite the fact that the correct classification percentage for the model is low. Housing, income, job, communication, environment and health are variables that decrease life satisfaction.

Parameter estimation values for robust logistic ridge regression analysis are shown in Table 4. When Table 4 is examined, it is determined that all coefficient values are significant at 0.05 significance level. It has been determined that the standard deviation values obtained for the coefficients are smaller than those found in the robust logistic regression analysis. Robust logistic ridge regression analysis also seems to have 89.90% correct classification percentage. When the odds ratios are examined, it is seen that variables of income, education, environment, civil participation, health, work life balance and security increase the life satisfaction, whereas the variables of residence, job, housing, communication and health have an effect to decrease life satisfaction.

The correct classification percentages of all methods used in this paper are given in Table 8.

Table 8. Correct Classification Percentage (CCP)

Techniques	**CCP (%)**
Logistic Regression Analysis	-
Robust Logistic Regression Analysis	100.00
Logistic Ridge Regression Analysis	68.42
Robust Logistic Ridge Regression Analysis	89.90
Discriminant Analysis	92.10
Robust Discriminant Analysis	94.70

4. CONCLUSION

Life satisfaction is an important variable considering the quality of life of countries. The study of how life satisfaction is influenced by other variables of quality of life is important for determining the policies that countries will follow to improve their welfare. When working with index data, it usually deals with the detection of variables that will explain more about the event. For this reason, only basic component analyses are used. In this study, life satisfaction, which is an important variable affecting the quality of life, is handled and modeled. Life satisfaction is measured with 10 variables in total. Life satisfaction provides a comparison between the prosperity levels of countries in many different areas. Especially today, the fact that the prosperity of the countries is defined only by the income does not show that the prosperity of the country is good. As a result of statistical analysis, this situation has been expressed more clearly.

When the correct classification percentages for discriminant analyzes are compared, it is seen that the robust discriminant analysis is 94.7% and the classical discriminant analysis has the correct classification percentage of 92.1% for the variables that are believed to affect quality of life. The quality-of-life data set does not show multivariate normal distribution. Therefore, it is more convenient to use robust discriminant analysis results.

Logistic regression analysis was then performed as an alternative to discriminant analysis. When the results of the logistic regression analysis are examined, it has been found that the error terms do not have normal distribution and have an outlier value. In addition, the existence of a multicollinearity was suspected because the model was generally significant, none of the coefficients were not significant when the coefficients were tested individually, and Cox & Snell and Nagelkerke R^2 values were too high.

Firstly, robust logistic regression analysis was applied because the assumption of normality and the effect of outliers were minimized. However, in this model, despite the fact that all coefficient values related to life satisfaction are statistically significant, the results of the opposite signs are observed in the theoretical substructure. Furthermore, the fact that the model has 100% correct classification has led to the suspicion of the existence of multicollinearity. When the correlation values between the explanatory variables and the VIF coefficients are examined, the existence of multicollinearity is determined. For this reason, robust logistic ridge regression analysis was applied. The correct classification percentage for this model was found to be 68.42%. The more reliable model among logistic regression models is logistic ridge regression analysis results. Logistic ridge regression analysis was found to be more reliable than the logistic regression analysis, although the correct classification percentage was low. Robust logistic ridge regression analysis was applied because the assumption of normality and multicollinearity. Robust logistic ridge regression analysis also seems to have 89.90% correct classification percentage. Robust logistic ridge regression was found to be most reliable than other logistic regression models.

Income has more influence on countries' life satisfaction than other variables. However, when the results of the robust logistic ridge regression analysis are examined, education and security have a more increasing effect on life satisfaction than income. Civil participation variable has an increasing effect on life satisfaction as much as income. Through life satisfaction, countries are able to produce policies for the index values they have been lagging behind other countries. Individuals are also could be able to put into practice exercises that maximize their life satisfaction. When the odds ratios of robust logistic ridge regression analysis are examined, it is seen that variables of income, education, environment, civil participation, health, work life balance and security increase the life satisfaction.

In this study, six different methods for modeling of life satisfaction are discussed and evaluated. The most important fact that the researchers had often overlooked is the assumptions of the method they are interested. Just like the water is boiled at 100 ^{0}C to based on the assumpiton of sea level conditions, the results from the statistical models depend on assumptions. In the modeling of life satisfaction, the robust discriminant analysis is preferred if only classification is concerned. However, when it is important to determine the affecting levels for life satisfaction, the use of robust logistic ridge regression analysis results will be more useful.

REFERENCES

ABRAMS, M. (1973), Subjective social indicators, *Social Trends (HMSE)*, 4, 35–50.

AGULANNA, C. (2010), Community and Human Well-Being In An African Culture, *Trames,* 14, 282-298.

AKAR, S. (2014), Türkiye'de Daha İyi Yaşam İndeksi: OECD Ülkeleri İle Karşılaştırma, *Journal of Economic Life*, 1-12.

ALBERT, A., and LESAFFRE, E. (1986), Multiple Group Logistic Discrimination, *Computational Mathematics With Applications,* 12(2) ,209-224.

ANDREWS, F.M. (1974), Social indicators of perceived life quality, *Social Indicators Research*,1, 279–299.

ATABEY, Ö. (2010), Lojistik regresyon modeli ve geriye doğru eliminasyon yöntemiyle değişken seçiminin hipertansiyon riski üzerine uygulamasında Bootstrap yöntemi, Gazi University, Ankara.

ARICAN, E. (2010), Nitel Yanıt Değişkene Sahip Regresyon Modellerinde Tahmin Yöntemleri, Çukurova University,Adana.

ARIFFIN, S. B., and MIDI, H. (2017), Robust Logistic Ridge Estimator in the Presense of High Leverage Multicollinear Observations, *Mathematical and Computational Methods in Science and Enginnering*, 179-184.

BARBATO, A., MONZANI E., and SCHIAI, T. (2004), Life satisfaction in a sample of outpatients with severe mental disorders: A survey in northern Italy, *Quality of Life Research*, 13,969–973.

BIANCO, A. M., and VICTOR, J. Y. (1996), Robust Estimation in the logistic regression model ,*Springer*, 17-34.

BİRCAN, H. (2004), Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama, *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* ,2,185-208.

BURMAOĞLU, S., OKTAY, E., and ÖZEN, Ü. (2009), Birleşmiş Milletler Kalkınma Programı Beşeri Kalkınma İndeksi Verilerini Kullanarak Diskriminant Analizi ve Lojistik Regresyon Analizinin Sınıflandırma Performanslarının Karşılaştırılması, *Kara Harp Okulu Yayınları*, Ankara.

CARLSSON, M., and HAMRIN, E. (2002), Evaluation of the life satisfaction questionnaire (LSQ) using structural equation modelling (SEM), *Quality of Life Research,* 11,415–425.

CHORK, C. Y., and ROUSSEEUW, P. J. (1992), Integrating a high-breakdown option into discriminant analysis in exploration geochemistry, *Journal of Geochemical Exploration,* 43, 191–203.

CILGA, İ. (1994), *Gençlik ve Yaşam Niteliği*, Ankara: T.C. Başbakanlık Gençlik ve Spor Müdürlüğü Basımevi.

CROUX, C., and DEHON, C. (2001), Robust linear discriminant analysis using S-estimators, The *Canadian Journal of Statistics*, 29, 473–492.

CROUX, C., FILZMOSER, P., and JOOSSENS, K. (2005), Robust linear discriminant analysis for multiple group: Influence and Classification Efficiencies, *Vienna University of Tehcnology*, 1–28.

DUFFY, D. E., and SANTNER, T. J. (1989), On the Small Sample Properties of Norm-Restricted Maximum Likelihood Estimators for Logistic Regression Models, *Communications in Statistics - Theory and Methods,* 18, 959-980.

DURAND, M., and SMITH, C. (2013), The OECD Better Life Initative: How's Life? And The Measuring Well-being, *OECD Publishing*, Hong-Kong, 1–20.

DURAND, M. (2015), The OECD Better Life Initiative: How's Life And The Measurement Of Well-Being, *Review of Income and Wealth,* 61,4-17.

FILZMOSER, P., JOOSSENS, K., and CROUX, C. (2006), Multiple Group Discriminant Analysis:Robustness and Error Rate, *Vienna University of Tehcnology,* 1–12.

FISHER, A. F. (1938), The Statistical Utilization of Multiple Measurements, *Annals of Eugenics*, 8, 376–386.

GOUGH, I., MCGREGOR, J. A., and CAMFIELD, L. (2006), Well-Being In Developing Countries, *WeD ESRC Research Group*, 1–58.

HAIR, J. F., ANDERSON, R. E., TATHAM, R. L., and BLACK, W. C. (1995), *Multivariate Data Analysis With Reading,* Boston :Prentice Hall International Editions.

HE, X., and FUNG, W.K. (2000), High breakdown estimation for multiple populations with applications to discriminant analysis, *Journal of Multivariate Analysis,* 72,151–162.

HOBZA, T., PARDO, L., and VAJDA, I. (2012), Robust median estimator for generalized linear models with binary responses, *Kybernetika,* 48, 768-794.

HOERL, A. E., and KENNARD, R.W. (1970), Ridge regression, Biased estimation for nonorthogonal problems, *Technomctrics*, 12, 69-82.

HUBERT, M. and DRIESSEN, K. V. (2004), Fast and robust discriminant analysis, *Computational Statistics and Data Analysis*, 45, 301–320.

JOVANOVIC, V. (2016), The validity of the Satisfaction with Life Scale in adolescents and a comparison with single-item life satisfaction measures: a preliminary study, *Quality of Life Research* 25, 3173–3180.

KASPARIAN, J., and ROLLAND, A. (2012), OECD's Better Life Index: Can Any Country Be Well Ranked?. Research Article, *Journal of Applied Statistics* 39, 2223– 2230.

KERENYI, A. (2011), The Better Life Index of the Organisation for Economic Co-operationand Development, *Public Finance Quarterly* 56, 518–538.

KLECKA, W.R. (1980), *Discriminant Analysis*, London:Sage Publications.

LACHENBRUCH, P.A. (1975), *Discriminant Analysis*, London: Hafner Press.

MIZOBUCHI, H. (2013), *Measuring World Better Life Frontier*, Discussion Paper Series, No. 13-01, 1-23.

MONTGOMERY, D.C., PECK, E.A., and GEOFFREY, V.G. (2001), *Introduction to the Linear Regression Analysis*, Canada:John Willey&Sons.

OECD, (2013), How's Life? 2013 Measuring Well-being, Paris: OECD Publishing.

OECD, (2017), *OECD Beter Life Index*, It was taken http://www.oecdbetterlifeindex.org site on June 14 2017.

OSBERG, L., and SHARPE, A. (2002), An Index Of Economic Well-Being For Selected OECD Countries, *Review of Income and Wealth* 48, 291-316.

PATRICK, D.L. and ERICKSON, P. (1993), *Health policy, quality of life: health care evaluation and resource allocation*, New York: Oxford University Press.

RANDLES, R.H., BROFFITT, J.D., RAMBERG, J.S., and HOGG, R.V. (1978), Linear and Quadratic Discriminant Functions Using Robust Estimates, *Journal of the American Statistical Association*, 73, 564–568.

RAO, R.C., (1948), The utilization of multiple measurements in problems of biological classification, *Journal of the Royal Statistical Society*, Series B, 10, 159–203.

SEKER, M. (2010), Mutluluk Ekonomisi, İktisat Fakültesi Sosyoloji Konferansları, 39. Seri, İstanbul, 115- 134.

SHIN, D.C., and JOHNSON, D.M. (1978), Avowed happiness as an overall assessment of the quality of life, *Social Indicators Research*, 5(4), 475-492.

STEVANOVIC, D. (2014), Is the Quality of Life Enjoyment and Satisfaction Questionnaire- Short Form (Q-LES-Q-SF) a unidimensional or bidimensional instrument?, *Quality of Life Research* 23, 1299-1300.

STEVENSON, B., and WOLFERS, J. (2013), Subjective Well-Being and Income: Is There Any Evidence of Satiation, *American Economic Association* 103, 598-604.

TÜZÜN, H.E., and EKER, L. (2003), Sağlık Değerlendirme Ölçütleri ve Yaşam Kalitesi, *Sağlık ve Toplum Dergisi* 13, 3-8.

TEKKANAT, C. (2008), Öğretmenlik bölümünde okuyan öğrencilerde yaşam kalitesi ve fiziksel aktivite düzeyleri, Pamukkale University.

World Health Organization, (2017), It was taken http://www.who.int/en site on August 15 2017.

YAVUZKANAT, P. (2013), Türkiye'de İllerin Yoksulluk Riskinin Ölçülmesi Üzerine Bir Yöntem Önerisi, Aile ve Sosyal Politikalar Uzmanlık Tezi, Ankara.