

# Investigation of Factors Affecting General Mortality in Some Countries by Quantile Regression Method Alternative to Least Squares Method (LSM)

Yüksel Akay Ünvan

Ankara Yıldırım Beyazıt University, Faculty of Management, Banking and Finance Department, Turkey

## ABSTRACT

This study was conducted to find out the factor or factors affecting the overall mortality rates in a total of 31 countries, including 28 European Union countries.

The data set consisting of 2014 year data was analyzed using the Eviews 9 program. After the descriptive statistics and covariance matrix were determined, the regression model was established by the LSM. It has been observed that this model does not provide the assumption that it does not contain outliers, which is one of the regression assumptions. Therefore, 3 Quantile Regression models were established by using the values of 0.25, 0.50 and 0.75. Interpretations were made according to these regression equations. Factors affecting the General Mortality (OLM) are as follows. In the quantile model of 0.25; the Ratio of People With Asthma (RA) has a negative effect and the Ratio of People With Blood Pressure (TAN) has a positive effect. In the 0.50 quantiles model; only the TAN variable has a positive effect. In the last model with a value of 0.75 quantiles, again the TAN variable has a positive effect.

The general result according to the models established for 3 quantile values is that the AST variable has a negative (decreasing) effect on General Mortality (OLM) while TAN variable has a positive (increasing) effect.

**Key Words:** General Mortality, Asthma, Blood Pressure, Least Squares Method, Quantile Regression

## Introduction

Death occurs as a concrete fact in all times when living beings exist with the formation of the Earth. These deaths can be in natural way as well as a result of a number of factors. In this study, in order to determine some factors that have an impact on the General Mortality Rate, a data set that includes some of the death and health statistics for some countries including the European Union (EU) countries, Turkey, Iceland and Norway was formed. The EU member states included in the study are; Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, the United Kingdom, and Croatia [1]. The data set is consisting of a total of 31 countries and was taken from the Eurostat [2] website. The General Mortality for the countries included in the study are shown in Figure 1. The least General Mortality was observed in Turkey, Ireland and Iceland.

**Literature Review About The Studies Related With Mortality/Death Numbers:** Keys, Menotti & Aravanis (1984) obtained the following results in their study: 15 cohorts (four regions) in seven countries differed by causes of death, mainly reflecting large differences in coronary deaths. Only mean blood pressure helped explain cohort differences in all-cause mortality. Three-quarters of the change in coronary mortality was calculated by differences in mean serum cholesterol and blood pressure in cohorts. Age, serum cholesterol, blood pressure and smoking for coronary death are very important in all regions except Japan, where coronary deaths are scarce for evaluation. Physical activity is important only in Southern Europe, where differences are related to socioeconomic status. Relative body weight tends to be a negative risk factor everywhere. Between the United States and Northern Europe there are similar cross-estimates of death cholesterol, blood pressure, smoking habits, physical activity factors. Analysis of time trends, age, blood pressure and smoking were important in the relationship between mortality and entry characteristics, and cholesterol tended to decrease over the last 5 years.

Gaziano, Bitton & Anand, et al. (2010) reported that coronary heart disease (CHD) is the largest cause of death in developed countries and is one of the leading causes of disease burden in developing countries. In 2001, there were 7.3 million deaths worldwide due to CHD. Three-quarters of global deaths due to CHD occurred in low- and middle-income countries. The rapid increase in CHD burden in most low- and middle-income countries results from socio-economic changes, an increase in life duration, and the acquisition of lifestyle risk factors. However, mortality from CHD varies considerably in developing countries. The varying incidence, prevalence and mortality rates reflect different levels of risk factors, other competing causes of death, the availability of resources to combat cardiovascular disease, and the epidemiological transition phase in which each country or region finds itself.

Mozaffarian, Fahimi & Singh, et al. (2014) reported that high sodium intake increases blood pressure, a risk factor for cardiovascular disease, but the effects of sodium intake on global cardiovascular mortality are uncertain. Data were collected from questionnaires of 66 countries (74.1% of adults worldwide) on urinary excretion and dietary sodium intake. The cause-related mortality was taken from the Global Burden of Disease Study (2010). Using comparative risk assessment, the current sodium intake; cardiovascular effects were calculated in comparison to the reference to sodium intake of 2.0 g per day by age, country and sex. In 2010, the estimated average level of global sodium consumption ranged from 3.95 g per day to regional mean levels from 2.18 to 5.51 g per day. Globally, 1.65 million deaths per year due to cardiovascular causes are attributed to sodium intake above the reference level. 61.9% of these deaths occurred in men and 38.1% in women. These deaths account for about 1 in 10 deaths from cardiovascular causes. Cardiovascular mortality associated with sodium intake above the reference level is highest in the Georgian country and lowest in Kenya. As a result, in this modeling study, 1.65 million deaths due to cardiovascular causes in 2010 were attributed to sodium consumption above 2.0 g reference level per day.

Keys, Menott & Karvonen, et al. (2017) reported that 2.288 of the 11.579 male subjects they took from 15 cohorts aged 40-59 years died within 15 years. The mortality including differences in mean age, blood pressure, serum cholesterol and smoking habits was 46%, 80% for coronary heart disease, 35% for cancer and, 45% for stroke. Mortality differences are not related to differences between mean relative body weight, obesity, and physical activity. Since cohorts differ in average diets; mortality rates were positively

correlated with mean percentage of dietary energy from saturated fatty acids and percentage of dietary energy from monounsaturated fatty acids. However, the ratio of monounsaturated fats to saturated fatty acids was negatively related. In addition, oleic acid makes up almost all differences in monounsaturated products between cohorts.

Afshin, Forouzanfar & Reitsma, et al. (2017) analyzed data from 68.5 million people to evaluate trends in overweight and obesity prevalence among children and adults between 1980 and 2015. According to the data, 107.7 million children and 603.7 million adults are obese in 2015. As a result of the analysis; since 1980, the prevalence of obesity has doubled in more than 70 countries and has been found to increase continuously in many other countries. Although the prevalence of obesity among children is lower than in adults, the rate of increase in obesity in childhood has been greater than the increase in obesity in adults in many countries. The high body mass index accounts for 4 million deaths worldwide. Furthermore, more than two thirds of deaths associated with high body mass index (BMI) occur due to cardiovascular disease. The burden of disease associated with BMI has increased since 1990. The high prevalence of BMI and the rapid increase in disease burden emphasize the need to focus on the surveillance of BMI and on the identification, implementation and evaluation of evidence-based interventions for this problem.

Roth, Abate & Abete (2018) 's study of "The Global Burden of Disease, Injury and Risk Factors Study (GBD) 2017" provides a comprehensive assessment of specific deaths for 282 causes in 195 countries and regions from 1980 to 2017. According to the results of the study; non-communicable diseases (NCDs) account for the largest rate of deaths, with 73.4% of the largest causes of death. The rate of infectious, maternal, neonatal and nutritive (CMNN) causes of total deaths in 2017 is 18.6% and the rate of injuries is 8.0%. The total number of NCD-related deaths increased by 7.61 million, or 22.7%, from 2007 to 2017. The mortality from NCDs was 7.9% globally; the mortality for CMNN causes was 22.2% and the general mortality decreased by 31.8%. Deaths from substance use disorders increased from 284,000 worldwide in 2007 to 352,000 in 2017. Between 2007 and 2017, total deaths caused by conflict and terror were increased by 118%. Globally, the number of deaths for men (except those older than 85) is higher in 2017 than in women. There has also been a large increase in mortality from neoplasms and cardiovascular diseases. The leading causes of deaths in 1990; neonatal diseases, lower respiratory tract infections and diarrhea diseases were ranked second, fourth and fifth in 2017. Developments in global

health are unevenly distributed among societies. As a result, injuries, substance abuse disorders, armed conflict and terror, deaths from neoplasms and cardiovascular disease increase the threats to global health.

**Literature Review About The Studies Using Quantile Regression:** Özel & Sezgin (2012), stated the purpose of their study as determining the degree and direction of influence of Turkey's trade openness on economic growth level. In the analysis, the Bootstrap Quantile Regression method was applied to the quarterly data of 1998.Q1-2011.Q4, and it was determined that trade deficit had a statistically significant and increasing effect on growth. Çelik & Selim (2014), have made their work in order to analyze income differences that occurred between the urban and rural working women and men in public and private sectors in Turkey. In the study, by using the micro Household Labor Force Survey data sets prepared by the Turkey Statistical Institute in 2011, the least squares method (LSM) and quantile regression model results were compared. When the LSM results of the male and female employees in the public sector in the urban sector were examined, the female employees in public sector had higher income than men in all age groups. According to the results of quantile regression, income-age profile of all age groups of private sector female employees in the city continued in the reverse U-shape. The same has not been the case for private sector male workers in the city. It has been observed that the income gap between the 36-45 age group and over 46 age group has decreased gradually.

Yavuz & Aşık (2017) introduced Quantile Regression in their studies and compared it with the LSM estimators on an engineering application. According to the results obtained for the concrete breaking test, it was determined that the results of the model obtained by the LSM cannot be used for interpretation purposes. Then, quantile regression was applied and BE3 variable was found to be significant for all three models.

In his study, Büberkökü (2018) examined the sensitivity of short-term and long-term interest rate risk of 6 major deposit banks traded in BIST with the two-factor Arbitrage Pricing Model. 3-month interbank money market interest rate was used as the short-term interest rate and 10-year government bond interest rates were used as the long-term interest rate. Quantile regression was used in model estimations. Thus, more efficient and consistent estimators were obtained according to LSM method. Findings show that banks are sensitive to both short and long term interest rate risk and are negatively affected by the increase in interest rates.

Gürler, Birecikli & Eryavuz (2018), using the Household Budget Survey raw data between 2003 and 2014, identified the differences between the groups by dividing consumption expenditures and food expenditures as sub-items by 20%. For this purpose, 3 different models with 5 different quantile values and LSM were tested and interpreted. In their study, Topbaş & Unat (2018) investigated the relationship between income and consumption with the quantile regression model, based on the traditional consumption function. For this aim, the structure of household consumption trends has been investigated in terms of expenditure groups using the data obtained from households budget survey made by Turkey Statistics Institution (TSI) from 2005 to 2016 year in Turkey. According to the results of the study, the households in the low spending group have a low consumption tendency and it is observed that consumption tendency increases as the lowest to the highest level for all other tranches of the Quantile Regression. In the study, it was found that the marginal consumption trend for each quantile showed a stable structure over time. However, significant differences have been observed between the groups (in cross-section).

## Materials and Methods

In this study, in addition to 28 European Union member countries; Iceland, Norway and Turkey 'data (the rates) is also used in the dataset. In the study where a total of 31 countries were used as observations, 8 different variables were given for these countries. The following methods and analyzes were applied on this data set consisting of 31 observations and 8 variables using Eviews 9 program.

### Regression with Least Squares Method (LSM):

One of the most commonly used regression methods is the LSM method. In this method, the estimated values of  $\beta_0$  and  $\beta_1$  parameters are determined. The univariate regression equation for LSM is given as follows:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad i = 1, 2, \dots, n \quad (1)$$

The error term shown in formula (2) consists of the difference between the real and the estimated value.

$$\hat{\varepsilon} = Y_i - \hat{Y}_i \quad (2)$$

The error terms given in Formula (2) have both positive and negative values together with zero, and the sum of these differences should be zero as given in Formula (3) (Dirican, 2012).

$$\sum_{i=1}^n \hat{\varepsilon}_i = \sum_{i=1}^n (Y_i - \hat{Y}_i) = 0 \tag{3}$$

As given in Formula (4), the main purpose of the LSM is to make the sum of the squares of the difference between the actual value and the estimated value of the dependent variable to a minimum.

$$\sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \tag{4}$$

If all observations are collected by taking the squares of the generated error term, the sum of the error squares is obtained. In order to minimize the sum of squares in the LSM method, derivatives of the above expression according to  $\beta_0$  and  $\beta_1$  predictions are taken and equalized to zero, resulting in the equation systems  $\beta_0$  and  $\beta_1$  (Alma and Vupa, 2008).

**Residual Analysis**

**DFFITs:** The DFFITs<sub>i</sub> statistic which is proposed by Belsley et al. (1980) and based on the deletion of the observation, is a diagnostic method that examines the effect of ith observation on the estimation values by deletion of that observation (Acarlar, 2011; 107). Let the estimation value  $\hat{y}_i$  for the ith observation calculated by the regression coefficients obtained from the whole data, and the predictive value  $\hat{y}_{(i)}$  for the ith observation calculated by the regression coefficients obtained by deleting the first observation from the data. Accordingly, DFFITs<sub>i</sub> is defined as in Formula (5).

$$DFFITs_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\hat{\sigma}_{(i)} \sqrt{h_{ii}}} \quad i = 1, 2, \dots, n \tag{5}$$

Belsley et al. (1980) proposed a critical value of  $2\sqrt{p/n}$  for this statistic. According to this critical value; the observations that meet the condition  $|DFFITs_i| > 2\sqrt{p/n}$  are contradictory observations.

**COVRATIO:** Abuzaid et al. (2011; 323-324) provide the following information and formulas about COVRATIO in their study. Hussin et al. (2004), proposed the model given in the formula (6) for circular observations with a linear relationship between X and Y circular variables X and Y circular variables  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

$$y_i = \alpha + \beta x_i + \varepsilon_i \pmod{2\pi} \tag{6}$$

Here,  $\varepsilon_i$  is a random error with a concentration parameter  $\kappa$  ve and a circular average of 0. The estimator of the concentration parameter is as in Formula (7).

$$\hat{\kappa} = A^{-1} \left( \frac{1}{n} \sum \cos(y_i - \hat{\alpha} - \hat{\beta}x_i) \right) \tag{7}$$

A function is the ratio of the modified Bessel function of the first kind of order zero. The inverse of the function A is given by Dobson (1978) as follows.

$$A^{-1}(w) \approx \frac{(9-8w+3w^2)}{(8-8w)} \tag{8}$$

The function of A is also multiplied by the estimator of the concentration parameter  $\kappa$  as in Formula (9).

$$A(\hat{\kappa}) = \frac{1}{n} \sum \cos(y_i - \hat{\alpha} - \hat{\beta}x_i) \tag{9}$$

The covariance matrix used in the COVRATIO formula and the formulas for the determinant of the covariance matrix are given in Formula (10) and Formula (11).

$$COV = \frac{1}{\hat{\kappa}A(\hat{\kappa}) \{n \sum x_i^2 - (\sum x_i)^2\}} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \tag{10}$$

$$|COV| = \frac{1}{\hat{\kappa}A(\hat{\kappa})} \tag{11}$$

From this, the formula COVRATIO for the i<sup>th</sup> observation is obtained as in Formula (12).

$$COVRATIO_{(i)} = \frac{|COV|}{|COV_{(-i)}|} = \frac{\hat{\kappa}_{(-i)} A(\hat{\kappa}_{(-i)})}{\hat{\kappa} A(\hat{\kappa})} \tag{12}$$

For any i<sup>th</sup> observation, where n is the sample size, if the  $|COVRATIO_i - 1|$  value containing the formula COVRATIO is equal to or greater than  $(6/n)$ , that observation is called an outlier.

**Quantile Regression:** Quantile regression has emerged as an alternative as the models established by LSM are sensitive to extreme values. Excessive values will increase the error of the installed model. This situation disappears with Quantile Regression (Çınar, 2019). Quantile Regression was developed for selected quantiles of conditional distribution of dependent variable. Unlike the classical regression model, there is no requirement to provide any assumptions about the homogeneity of error variance and the distribution of errors. Therefore, it can be considered as a more flexible approach compared to LSM. The regression with LSM is based on the minimization of the conditional mean of the dependent variable and the sum of the residual squares, while the Quantile Regression functions are based on the minimization of the weighted sum of the absolute residues (Yavuz & Aşık, 2017).

The quantile regression method is a settlement model. The simple position model is shown as in Formula (13).

$$Y_t = \beta + e_t \tag{13}$$

In the model,  $Y_t$  is a random and independent variable with a symmetric F distribution function and has an identical distribution and  $\beta$  median. In this model, the sample  $\theta^{\text{th}}$  quantile is;

$$\min_{\beta} \frac{1}{n} \{ \sum_{i: y_i \geq \beta} \theta |y_i - \beta| + \sum_{i: y_i < \beta} (1 - \theta) |y_i - \beta| \} \tag{14}$$

and is obtained by minimization of the expression given in Formula (14). This quantile, with the linear regression model below;

$$y_i = x_i' \beta + e_i \tag{15}$$

based on the signs of the  $\theta^{\text{th}}$  quantile regression observation values,

$$\beta \min_{\beta} \frac{1}{n} \sum_{i=1}^n (\theta - \frac{1}{2} + \frac{1}{2} \text{sgn}(y_i - x_i' \beta)) (y_i - x_i' \beta) \tag{16}$$

and is estimated as given in Formula (16).

Here,  $\text{sgn}(a)$  is the sign of  $a$  and takes 1 if it is positive and -1 if it is negative or zero. Estimates are based on signs of observation values instead of magnitude of observation values, which makes quantile regression a robust method (Topbaş & Unat, 2018).

### Findings

**Variables:** The definitions of the 8 different variables used in the study are given in Table 1. The rates of these variables are expressed as a percentage of the individuals in the whole population who have the characteristics of the variable.

**Descriptive Statistics:** Some descriptive statistics of the 8 variables used in the study are given in Table 2. These statistics provide information about the distribution of data in each variable. The average value for the variable “General Mortality” (OLM) is 9,641; while the median value is 9,300. In other words, the average mortality of 31 countries in the data set is 9,641% and the median value is 9,300%. In addition, the highest mortality in the data set for the OLM variable was found to be 15,100% and the lowest mortality was 5.1%. The average value for another variable, “Rate of People Who Never Eat Fruit and Vegetables” (MEYV) is 35,796; while the median value is 34,700. In other words, the average of those who do not eat any fruits and vegetables in 31 countries in the data set is 35,796% and the median value is 34,700%. In addition, the highest rate in the data set for the MEYV variable was 65,100% and the lowest rate was 16.1%. Similar interpretations can be made for the 6 variables given in Table 2.

**Correlation Matrix:** A correlation matrix describes the relationship between the  $m$  variables where the cross elements are equal to the 1. The square obtained from the variance-covariance matrix is a symmetric  $m \times m$  matrix. Both of these matrices contain similar information, but the correlation matrix makes it easier to associate variables with each other (Horn and Johnson, 1985). Since the diagonal values in the correlation matrix are between the same two variables, they are always 1 (100%). When Table 3 examined, it is seen that the highest positive relationship is between OLM and TAN variables with 72.2%. In addition, the biggest negative relationship is between and OLM and AST variables with - 67.5%. Based on Table 3, similar interpretations can be made according to other variables.

### Regression Model with Least Squares Method (LSM ):

First, a regression model was established with Least Squares method (LSM), which is one of the most classical and most common regression methods. In this model where the dependent variable is determined as OLM variable in Table 4, only variables whose  $p$ -values are less than 0.05 significance value can be included in the regression equation. The model has an  $R^2$  (correlation coefficient) of 72.15%. This coefficient expresses the relationship between the variables in the model. In addition, since the  $p$ -value of the model that is 0,00 is less than 0,05, the model is accepted. In other words,  $H_0$  absence hypothesis is rejected. So the model is suitable for use. The established model equation is given as in Formula (17). However, it is necessary to look at the appropriateness of some regression assumptions before interpreting the model equation.

$H_0$ : Variable/Model is not significant. Not suitable for use.

$H_1$ : Variable/Model is significant. Suitable for use.

$$\text{OLM} = -0,472023 * \text{AST} + 0,256787 * \text{TAN} + 6,311919 \tag{17}$$

Leverage graphs given in Figure 2 is another method that makes some inferences about the variables in the regression model established by LSM. The red line in the graphs shows the dependent LSM variable. The variables that this red line slopes downward affect the dependent variable negatively and the variables that have upward slope affect the dependent variable positively. Thus, even if there are no coefficients given in the table above, inference can be made about the model. According to the graphs in Figure 2, AST variable affects OLM negatively, while TAN variable and C constant positively affect. Variables in graphs where the red line representing the OLM variable are straight are not significant, so they cannot be taken into the model.

**Table 1.** The Definition of the Variables

Variables	Definition of the Variables
OLM	General Mortality
AST	Rate of People with Asthma
DEP	Rate of People with Severe Depression
DIY	Rate of People with Diabetes
FIZ	Rate of People who has No Aerobic, No Physical Movement
MEYV	Rate of People who Never Eat Fruit and Vegetables
SOL	Rate of Chronic Lower Respiratory Disease (Except Asthma)
TAN	Rate of High Blood Pressure People

**Table 2.** Descriptive Statistics

	OLM	AST	DEP	DIY	FIZ	MEYV	SOL	TAN
Mean	9.641	5.461	0.687	6.416	48.561	35.796	3.790	21.223
Median	9.300	4.900	0.700	6.400	47.700	34.700	3.900	20.900
Maximum	15.100	9.400	1.500	10.00	88.200	65.100	7.700	31.900
Minimum	5.100	2.000	0.200	4.200	18.700	16.100	1.100	12.700

**Regression Assumptions**

**Normal distribution:** In the regression models made with LSM, the model should be distributed in accordance with the normal distribution. According to the results obtained from Figure 3, since the p-value of the distribution is  $0.735 > 0.05$ ,  $H_0$  hypothesis cannot be rejected. The model shows a distribution that is suitable for normal distribution. However, since compliance with this assumption is not sufficient by itself, other assumptions are needed to be examined.

$H_0$ : A distribution that is suitable for normal distribution is available.

$H_1$ : There is no suitable distribution for normal distribution.

**Heteroskedasticity (Changing Variance):** The lack of changing variance is another assumption. The Breusch-Pagan-Godfrey Test is a common test for testing changing variance. According to Breusch-Pagan-Godfrey Test results, since the p-value in Table 5 is  $0.4225 > 0.05$   $H_0$  absence hypothesis cannot be rejected. Also there is no problem of changing variance. Providing these two assumptions is not sufficient for the use of LSM. In addition, an outlier analysis is required.

$H_0$ : Changing variance is not available.

$H_1$ : Changing variance is available.

**Outlier Value Review:** Figures 4 and 5 show graphs for the DFFITS and COVRATIO values of the observations, respectively. According to both graphs, outlier values exceeding the limits are seen. Since these outliers may produce misleading information

about the regression results, it is more appropriate to use the Quantile Regression method, which is an alternative to the regression models made with the LSM and in which the assumptions are not taken into consideration.

**Quantile Regression:** In the Quantile Regression, the data is analyzed part by part. This is due to the fact that the tail portions are also important in the distribution of data. In this study, the data were examined in 3 parts with 3 different quantile values. First, a quantile value of 0.25 was used. The 0.25 quantile forms a model by separating the lowest 25% of the data from the highest 75%. According to this model where OLM is dependent variable; AST and TAN variables were statistically significant at 5% significance level. In addition, the  $R^2$  (correlation coefficient) value of the model is 48.1% and the probability value is 0.0010. Since this probability value is less than 0.05 significance level, the model is significant.

The second model to be established has been created by using a quantile value of 0.50. The 0.50 quantile is also called the expected value or the second quarter. According to model results, while OLM is dependent variable; only the TAN variable was statistically significant at 5% significance level. In addition,  $R^2$  (correlation coefficient) value of this model is 46.63% and the probability value is 0.0012. Since this probability value is less than 0.05 significance level, the model is significant.

In the last model, a quantile value of 0.75 was used. It analyzes and separates 25% of the highest value data in the 0.75 quantile data set. According to model results, while OLM is dependent variable; only the

**Table 3.** Correlation Between Variables

	OLM	AST	DEP	DIY	FIZ	MEYV	SOL	TAN
OLM	1.000	-0.675	-0.130	-0.064	0.284	0.427	0.005	0.722
AST	-0.675	1.000	0.420	0.084	-0.467	-0.275	0.054	-0.485
DEP	-0.130	0.420	1.000	0.057	0.123	0.048	0.256	-0.178
DIY	-0.064	0.084	0.057	1.000	0.341	-0.184	0.254	0.210
FIZ	0.284	-0.467	0.123	0.341	1.000	0.158	0.212	0.139
MEYV	0.427	-0.275	0.048	-0.184	0.158	1.000	-0.102	0.279
SOL	0.005	0.054	0.256	0.254	0.212	-0.102	1.000	0.137
TAN	0.722	-0.485	-0.178	0.210	0.139	0.279	0.137	1.000

**Table 4.** Regression with LSM

Dependent Variable: OLM					
Variable	Coefficient	Standard Error	t-statistic	p-value	
AST	-0.472023	0.216784	-2.177392	0.0400*	
DEP	1.126236	1.046958	1.075722	0.2932	
DIY	-0.187719	0.206751	-0.907948	0.3733	
FIZ	0.003869	0.021287	0.181757	0.8574	
MEYV	0.027364	0.028251	0.968608	0.3428	
SOL	-0.073512	0.192508	-0.381866	0.7061	
TAN	0.256787	0.069202	3.710669	0.0012*	
C	6.311919	2.420523	2.607668	0.0157*	
R2	0.7215	F-Statistic		8.5137	
Verified R2	0.6367	p-value		0.0000*	

Note: The values marked with \* are statistically significant at 5% significance level

**Table 5.** Changing Variance Test

Breusch-Pagan-Godfrey Test Results			
F-statistic	1.054586	Probability Value	0.4225

TAN variable was statistically significant at 5% significance level. This variable will be included in the model to be installed. In addition, the R<sup>2</sup> (correlation coefficient) value of this model is 62.62% and the probability value is 0.0000. Since this probability value is less than 0.05 significance level, the model is significant. The results of regression models constructed by quantile values of 0.25, 0.50 and 0.75 are given in Table 6 respectively.

The equations of the models obtained from these three quantile values are given below and used for interpretation in the Results and Conclusion section.

LSM  $\Rightarrow$  OLM = - 0,472023\*AST + 0,256787\*TAN + 6,311919

0,25'lik Quantile Regression  $\Rightarrow$  OLM = - 0,767576\*AST + 0,267048\*TAN

0,50'lik Quantile Regression  $\Rightarrow$  OLM = 0,311914\*TAN

0,75'lik Quantile Regression  $\Rightarrow$  OLM = 0,294173\*TAN

**Results and Discussion:** In order to determine the variables that affect the “General Mortality (OLM)” variable, a number of analyzes were performed in this study. Firstly, a regression model was established by LSM, where “General Mortality (OLM)” is a dependent variable. In this model, the “Rate of Asthma People (AST)” variable has a coefficient of -0.472023 and negatively affects the “General Mortality (OLM)” variable. The “Rate of People with Blood Pressure (TAN)” variable and C constant coefficient positively affect the “General Mortality (OLM)” variable with coefficients of 0,256787 and 6,311919, respectively. However, in order to use this regression model established with LSM, some assumptions must be provided. For this purpose, on the model, the assumptions of conformity to normal distribution, not having varying variance and not

**Table 6.** The Regression Models Constructed by 0.25, 0.50 and 0.75 Quantiles

Dependent Variable: OLM				
Quantile Regression (tau = 0,25)				
Bandwidth method: Hall-Sheather, bw=0,2142				
Variables	Coefficient	Std. Error	t-statistic	Probability
AST	-0.767576	0.350074	-2.192610	0.0387*
DEP	0.713929	1.430681	0.499013	0.6225
DIY	-0.308649	0.348847	-0.884770	0.3854
FIZ	0.005674	0.026244	0.216210	0.8307
MEYV	0.056875	0.046246	1.229835	0.2312
SOL	0.022449	0.293189	0.076567	0.9396
TAN	0.267048	0.101509	2.630794	0.0149*
C	6.266865	4.395009	1.425905	0.1673
Pseudo R2	0.4810		LR statistics	24.154
Verified R2	0.3230		Probability	0.0010*
Quantile Regression (tau = 0,50; median)				
Bandwidth method: Hall-Sheather, bw=0,30928				
Variables	Coefficient	Std. Error	t-statistic	Probability
AST	-0.275815	0.502941	-0.548405	0.5887
DEP	1.143399	1.474953	0.775210	0.4461
DIY	-0.615288	0.473026	-1.300750	0.2062
FIZ	0.029370	0.042560	0.690073	0.4971
MEYV	-0.004219	0.041757	-0.101032	0.9204
SOL	-0.203164	0.334343	-0.607651	0.5494
TAN	0.311914	0.123299	2.529743	0.0187*
C	7.587692	4.155969	1.825734	0.0809
Pseudo R2	0.4663		LR statistic	23.860
Verified R2	0.3039		Probability	0.0012*
Quantile Regression (tau = 0,75)				
Bandwidth method: Hall-Sheather, bw=0,2142				
Variables	Coefficient	Std. Error	t-statistic	Probability
AST	-0.181351	0.540649	-0.335432	0.7403
DEP	0.700017	1.458311	0.480019	0.6357
DIY	-0.423271	0.409184	-1.034427	0.3117
FIZ	0.038200	0.047945	0.796743	0.4337
MEYV	0.013316	0.039772	0.334805	0.7408
SOL	-0.242260	0.458236	-0.528678	0.6021
TAN	0.294173	0.114992	2.558204	0.0176*
C	6.067630	5.029990	1.206291	0.2400
Pseudo R2	0.6262		LR Statistics	38.297
Verified R2	0.5125		Probability	0.0000*

Note: The values marked with \* are statistically significant at 5% significance level

having contradictory value were examined respectively. This regression model established by LSM has normal distribution and does not contain changing variance. However, according to DFFITS and COVRATIO statistics, it was found to have outlier values. For this reason, we have switched to

the Quantile Regression Method, which does not require any assumption.

Quantile Regression was applied on different quantile values of 0.25, 0.50 and 0.75. The equations obtained from these regression models are given above. Accordingly, for the model established with a quantile



value of 0.25; the variable “Rate of People with Asthma (AST)” has a negative coefficient of -0.767576 and “Rate of People with Blood Pressure (TAN)” has a positive coefficient of 0.267048 on the “General Mortality (OLM)” variable. For the model established with a median value of 0.50; only the “Rate of People with Blood Pressure (TAN)” variable has a positive coefficient of 0.311914 on the “General Mortality (OLM)” variable. The last quantile value was determined as 0.75, and again, only the “Rate of People with Blood Pressure (TAN)” variable has a positive coefficient of 0.294173 on the “General Mortality (OLM)” variable. While the positive coefficient variables in these equations directly affect the the “General Mortality (OLM)” variable, the negative coefficient variables affect inversely proportionally. In addition, the Pseudo R<sup>2</sup> (correlation coefficients) values of 3 quantile regression models were; 48.10%; 46.63% and 62.62%. These correlation coefficients indicate the consistency of the variables in the regression model. For the Quantile Regression models established with 7 independent variables, these correlation coefficients are consistent and acceptable. Moreover, the correlation coefficient of the 3<sup>rd</sup> model established with a quantile value of 0.75 is higher than the others. This result shows that this model is more meaningful than the others.

As a result, interpretations for these 4 models can only be made for models constructed using the Quantile Regression values. In the first model with a quantile value of 0.25 established by quantile regression; the 100 units increase in the “Rate of People with Asthma” reduces “General Mortality” by 76.75 units. The increase of 100 units in the “Rate of People with Blood Pressure” increases the “General Mortality” by 26.70 units. In the second model with a value of 0.50 quantile; the 100 units increase in the “Rate of People with Blood Pressure” increases the “General Mortality” by 31.19 units. In the last model, which has a value of 0.75 quantile, the increase of 100 units in the “Rate of People with Blood Pressure” increases the “General Mortality” by 29.41 units. The general conclusion to be drawn from these models, which were created by using 3 different quantile values, is as follows: while individuals with asthma have a decreasing effect on mortality, people with blood pressure have an increasing effect on mortality. For example through a community; the participation of individuals with asthma in the community will reduce deaths in that community. But if people with blood pressure participate, deaths will increase in that community. According to the results obtained from this study, it is suggested that blood pressure individuals and asthma patients should be included in the study as a factor.

## References

1. Abuzaid A, Mohamed İ, Hussin AG, Rambli A. A COVRATIO Statistic for Simple Circular Regression Model. *Chiang Mai J. Sci* 2011; 38: 321-330.
2. Acarlar I. Comparison of Diagnostic Methods for Detecting an Influential Observation in Regression. *Anadolu University Journal of Science and Technology* 2011; 1: 105-116.
3. Afshin A, Forouzanfar M, Reitsma M, vd. Health Effects of Overweight and Obesity in 195 Countries Over 25 Years. *New England Journal of Medicine* 2017; 377: 13-27.
4. Alma ÖG, Vupa Ö. The Comparison of Least Squares and Least Median Squares Estimation Methods Which Are Used in Linear Regression Analysis. *Süleyman Demirel Üniversitesi Fen Edebiyat Fakültesi Fen Dergisi* 2008; 3: 319-329.
5. Belsley DA, Kuh E, Welsch RE. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley Series in Probability and Mathematical Statistics 1980; 6-84.
6. Büberkökü Ö. Large-Scale Turkish Deposit Banks’ Exposure to Short-and Long-Term Interest Rate Risk: A Quantile Regression Analysis. *Finansal Araştırmalar ve Çalışmalar Dergisi* 2018; 10: 243-261.
7. Çelik O, Selim S. Analysis of Public-Private Sector Wage Differentials in Turkey By Quantile Regression Approach. *Yönetim ve Ekonomi* 2014; 21: 205-232.
8. Dirican E. The Investigation of Total Cholesterol, Ldl, Hdl and Triglycerides Levels with Respect to Age by Different Regression Models. *İnönü Üniversitesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Yüksek Lisans Tezi* 2012, Malatya.
9. Dobson AJ. Simple Approximations Forthe Von Mises Concentration Statistic. *Applied Statistics* 1978; 27: 345-347.
10. Çınar UK. A Alternative to Ordinary Least Squares Regression: Quantile Regression. *Avrasya Uluslararası Araştırmalar Dergisi* 2019; 7: 57-71.
11. Horn, RA, Johnson CR. (1985) *Matrix Analysis*, Cambridge University Press.
12. Hussin AG, Fieller NRJ, Stillman EC. Linear Regression for Circular Variables with Application to Directional Data. *Journal of Applied Science & Technology* 2004; 8: 1-6.
13. Gaziano TA, Bitton A, Anand S. Growing Epidemic of Coronary Heart Disease in Low- and Middle-Income Countries. *Current Problems in Cardiology* 2010; 35: 72-115.
14. Gürler ÖK, Birecikli ŞÜ, Eryavuz AK. Investigation of Household Consumption and Food Expenditures by Quantile Regression Method in Turkey. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 18. EYİ Özel Sayısı 2018; 219-238.

15. Keys A, Menotti A, Aravanis C. The Seven Countries Study: 2,289 Deaths in 15 Years. *Preventive Medicine* 1984; 13: 141-154.
16. Keys A, Menotti A, Karvonen MJ, et al. The Diet and 15-Year Death Rate in the Seven Countries Study. *American Journal of Epidemiology* 2017; 185: 1130-1142.
17. Mozaffarian D, Fahimi S, Singh G, et al. Global Sodium Consumption and Death from Cardiovascular Causes. *New England Journal of Medicine* 2014; 371: 624-634.
18. Roth G, Abate D, Abete K, et al. Global, Regional, and National Age-Sex-Specific Mortality for 282 Causes of Death in 195 Countries and Territories, 1980–2017: a Systematic Analysis for the Global Burden of Disease Study 2017. *The Lancet* 2018; 392: 1736-1788.
19. Yavuz AA, Aşık EG. Quantile Regression. *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi* 2017; 9: 138-146.
20. Topbaş F, Unat E. The Stability of the Income and Consumption Relationship: Evidence From The Quantile Regression Model According to Expenditure Groups and Time. *Izmir Democracy University Social Sciences Journal* 2018; 1: 103-126.
21. [https://europa.eu/european-union/about-eu/countries\\_en](https://europa.eu/european-union/about-eu/countries_en)
22. <https://ec.europa.eu/eurostat/data/database>
23. <https://ec.europa.eu/eurostat/databrowser/view/tps00029/default/map?lang=en>