TURKISH JOURNAL OF SCIENCE
VOLUME 6, ISSUE 2, 71–75
ISSN: 2587–0971

http://dergipark.gov.tr/tjos

An Application of Least Squares Method in Nonlinear Models-Solid Waste Sample

Fatma Buğlem YALÇIN[a]

[a] *Department of Mathematics, Faculty of Science and Letter, Ordu University, 52200, Ordu, Turkey*

Abstract. Nowadays, the usage of nonlinear regression models is common. However, non-linear models are more difficult to teach than linear models. The ordinary least squares method is effective in teaching nonlinear regression models. This study aims to teach the subjects of non-linear regression in statistics for students. The transformed form of non-linear regression models are used for this. Therefore the ordinary least squares estimators of regression models are obtained and the comparison of these is made. Besides, an explanatory application is made on this subject.

1. Introduction

Regression analysis technique is one of the statistical analysis used to determine the relationship between variables that have a cause-effect relationship and to make predictions about this relationship [5]. In this technique, a mathematical model is first established to find the relationship between the variables, and then the validity of the established model is examined [3]. If the established regression model doesn't fit to the data, misleading results will occur in the future [7]. The classical linear regression is used to modelling the relationship between two variables such as income-education level, height-weight of people, dose-effect of a drug, or height-boiling point of water [8]. Sometimes the dependent variable can be affected by more than one independent variable. In this case, the multiple linear regression is used to modelling the relationship between the variables [8].

Nowadays, the linear models are preferred so that calculations such as hypothesis testing and parameter estimation can be made more easily. The linear model works well in many cases, but in some cases this isn't possible. It cannot be said that the relationship can best be expressed with a linear model. Therefore, it can also be decided more accurately by trying non-linear models as well as linear models.

Various estimation methods have been developed in linear and non-linear regression models. The ordinary least squares (OLS) method is one of the most frequently used methods in regression analysis [6]. Therefore, the OLS method plays an important role in teaching regression analysis. In addition, OLS method is an optimal method according to the Gauss-Markov Theorem, since it aims to minimize the sum of squares of error.

It is possible for the stochastic term to be additive or unpredictable depending on the dependent variable. Depending on the variability of the stochastic term, nonlinear models can be linear regression models with

Corresponding author: FBY: mail address: yalcinfatmabuglem@gmail.com ORCID:0000-0003-4276-1820
Received: 6 July 2021; Accepted: 21 August 2021; Published: 30 September 2021
Keywords. linear model, non-linear model, ordinary least squares, estimate
2010 *Mathematics Subject Classification*. 62J02, 62J05, 62F03
Cited this article as: Yalçın FB. An Application of Least Squares Method in Nonlinear Models-Solid Waste Sample. Turkish Journal of Science, 2021, 6(2), 71-75.

appropriate transformation in parameters (log) and nonlinear regression models. Similarly, OLS method can be used for parameter estimation in transformed models. If the OLS criterion for linear and non-linear models in the variables was applied to the initial variables, the OLS criterion for non-linear models in the parameters should also be applied to the transformed variables, for example $\ln \nu$. However, the OLS parameter estimation for the transformed models is biased. Although non-linear models can be converted to linear regression models and estimated with OLS, it is useful to be careful about the characteristics of the stochastic error term that enters these models. Otherwise, the application of the OLS for the transformed model will not produce a model with correct statistical properties.

The teaching of nonlinear models is quite complex compared to linear models. Students often have problems interpreting the transformed regression model in regression analysis [1]. The OLS method plays an important role in teaching regression analysis to students. In this study, the visualized approach to teaching nonlinear regression subjects for graduate students is presented, based on the population-based solid waste amount experiment in Ordu Metropolitan Municipality. Thus, the OLS method is applied in non-linear models. In addition, the estimation parameters of the models are also compared.

2. Main Results

Consider the regression model (1)

$$\omega_i = \theta_1 \nu_i^{\theta_2} \varepsilon_i, (i = 1, ..., n) \tag{1}$$

where ω_i is a dependent variable, ν_i is an independent variable, θ_1 and θ_2 are unknown parameters, and ε_i is an error variable.

The regression model (1) is inherently linear regression model since it can be made linear according to the parameters θ_1 and θ_2 by suitable transformations.

Then it is assumed that the error isn't additive. In other words, the error variability isn't constant for all ν_i. There is likely to be unpredictable and random fluctuation at the different levels of ν_i [2]. In this case, in [2], it is assumed that

$$\prod_{i=1}^{n} \varepsilon_i = 1 \;\; or \;\; \prod_{i=1}^{n} \ln \varepsilon_i = 0. \tag{2}$$

By taking the logarithm of both sides of the model (1), the equation (3) is obtained:

$$\ln \omega_i = \ln \theta_1 + \theta_2 \ln \nu_i + \ln \varepsilon_i. \tag{3}$$

Under the minimum criterion (4),

$$\phi_1 = \Sigma_{i=1}^{n} (\ln \omega_i - \ln \theta_1 - \theta_2 \ln \nu_i)^2 \tag{4}$$

using OLS method, differentiation with respect to θ_1 and θ_2 yields the normal equations (5):

$$\Sigma_{i=1}^{n} \ln \omega_i = n \ln \theta_1 + \theta_2, \;\; and \;\; \Sigma_{i=1}^{n} \ln \nu_i \ln \omega_i = \ln \theta_1 \Sigma_{i=1}^{n} \ln \nu_i + \theta_2 \Sigma_{i=1}^{n} (\ln \nu_i)^2. \tag{5}$$

The parameter estimates θ_1 and θ_2 are obtained when the normal equations are solved simultaneously. To use the ordinary linear regression model, $\ln \varepsilon_i \sim N(0, \sigma^2)$ is supposed. When the regression model (3) run, normality tests are applied to the error estimates obtained from this regression [4].

Now, let consider the regression model (6). This model is actually non-linear with respect to the parameters θ_1 and θ_2. The error variability is assumed to be independent of ν_i, that is, the regression model (6) is in the form:

$$\omega_i = \theta_1 \nu_i^{\theta_2} + \varepsilon_i. \tag{6}$$

Under the minimum criterion (7),

$$\phi_2 = \Sigma_{i=1}^{n}(\omega_i - \theta_1 {\nu_i}^{\theta_2})^2 \tag{7}$$

using OLS method, differentiation with respect to θ_1 and θ_2 yields the normal equations (8):

$$2\Sigma_{i=1}^{n}(\omega_i - \theta_1 {\nu_i}^{\theta_2}){\nu_i}^{\theta_2} = 0, \ \ and \ \ 2\Sigma_{i=1}^{n}(\omega_i - \theta_1 {\nu_i}^{\theta_2})\theta_1 {\nu_i}^{\theta_2} \ln \nu_i = 0. \tag{8}$$

Since it isn't possible to solve these normal equations analytically, the OLS estimation can be made iteratively by using a linearization of the model with respect to θ_1 and θ_2. The estimation of the parameters θ_1 and θ_2 that obtained from the criterion (7) is biased from the estimation of the parameters θ_1 and θ_2 that obtained from the criterion (4).

It is also necessary to be careful about the properties of the stochastic error term. For hypothesis testing, it is supposed that the stochastic error term ε_i of the regression model (6) fits the normal distribution, but the stochastic error term ε_i of the regression model (1) and its statistical counterpart (3) fits the log-normal distribution with mean $\exp(\sigma^2/2)$ and variance $\exp(\sigma^2)(\exp(\sigma^2) - 1)$ [2].

3. An Explanatory Application

In this section, the impact of the population of Ordu Metropolitan Municipality on the amount of solid waste is examined. The analysis is based on the amount of solid waste collected by the district municipalities for the winter period of 2015 in 19 different districts of Ordu province and the population in the districts. The data are given in Table 1.

Table 1: Population of districts and total average solid waste amount.

Districts	Population	Total average solid waste amount (kg)
Altınordu	202310	160000
Ünye	120720	80000
Fatsa	111072	73000
Perşembe	31094	15000
Kumru	31064	21000
Korgan	29349	9000
Gölköy	28952	18000
Aybastı	25900	17000
Akkuş	23064	13000
Ulubey	18239	6000
Mesudiye	15759	6000
İkizce	14969	3000
Gürgentepe	13821	7000
Çatalpınar	13786	7000
Çaybaşı	13127	4000
Kabataş	10604	11000
Çamaş	8594	6000
Kabadüz	8531	6000
Gülyalı	7994	4000

The following hypothesis is chosen for the regression analysis:

"The amount of solid waste depends on the population."

Let the regression model (1) be chosen for population-based solid waste modeling, where ω_i is the total average amount of solid waste in the winter period in 2015 and ν_i is the population in these municipalities. Under the minimum criterion (4), estimates of θ_1 and θ_2 are obtained by using OLS method:

$$\hat{\theta}_1\text{=0,185142} \quad \text{and} \quad \hat{\theta}_2\text{=1,102617.}$$

The regression results show that the $\ln\theta_1$ and θ_2 parameters are both significant (p-value$<$ 0.05) and the population is dependent on waste. θ_2 is the slope parameter and it measures the percentage change in ω for a given percentage change in ν (Figure 1). Now, let the model (6) be chosen for population-based solid waste modeling. Under the minimum criterion (7), estimates of θ_1 and θ_2 are obtained by iteration (with the generalized Newton-Raphson iteration):

$$\hat{\theta}_1\text{=1,212444} \quad \text{and} \quad \hat{\theta}_2\text{=0,955536.}$$

The green line indicates the amount of solid waste corresponding to the population. And the red and blue points indicate the values corresponding to the regression models (1) and (6), respectively (Firuge 1).

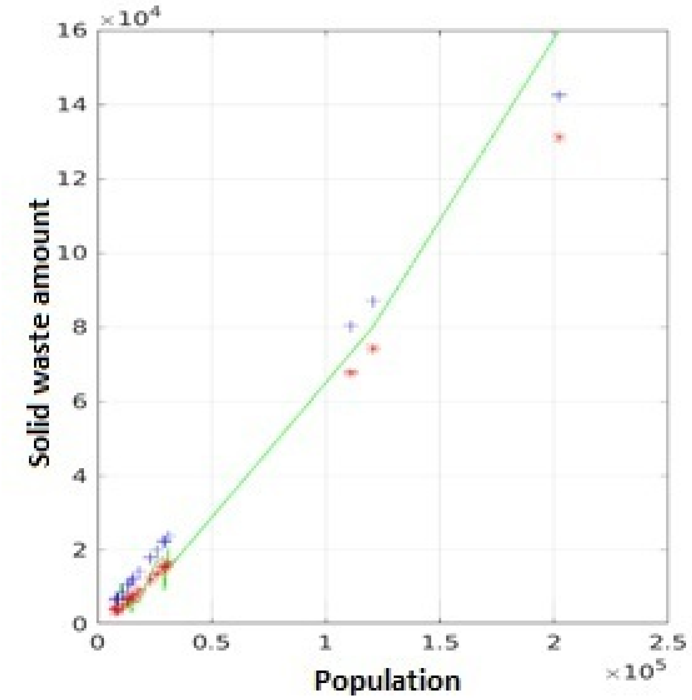

Figure 1: Relationship between population and solid waste by using models with additive and unpredictable error terms.

Both $\phi_1(0,185142;1,102617)$ and $\phi_2(1,212444;0,955536)$ minimum criterions have biased minimum solutions (Figure 2).

As a result, when the (1) regression is run, the OLS method is applied for the linear regression model according to the parameters. When the regression (6) is run, the OLS method is applied for the non-linear model according to the parameters.

(a) 2a (b) 2b

Figure 2: Comparison of $\phi_1(0,185142;1,102617)$ and $\phi_2(1,212444;0,955536)$ minimum criterions solutions.

4. Conclusion

The application of the OLS helps students to better understand the interpretation of parameters in the transformed models. It also allows them to obtain clear interpretation of the statistical models. They get explanation about the statistical properties of transformed models and the using of the OLS in the nonlinear regression. The above analysis explains the properties of the stochastic error term to students with application.

Acknowledgments

The author thank to Ordu Metropolitan Municipality for providing numeric datas.

References

[1] Arhipova I, Balina S. Problem of Choosing Statistical Hypotheses in Applied Statistics. 16th Symposium of IASC, Proceedings in Computational Statistics. 2004,629 – 636.
[2] Arhipova I, Arhipovs S. Application of Ordinary Least Square Method in Nonlinear Models. International Statistical Institue. 2007,56th Session,p.4.
[3] Birkes D, Dodge Y. Alternative Methods of Regression. John Wiley Sons. 1993,80 – 140.
[4] Gujarati DN. Basic Econometrics. McGraw-Hill, Inc. 1995, 108 – 151.
[5] Gürünlü AÖ, Vupa Ö. Regresyon Analizinde Kullanılan En Küçük Kareler ve En Küçük Medyan Kareler Yöntemlerinin Karşılaştırılması. Süleyman Demirel Üniversitesi Fen Edebiyat Fakültesi Fen Dergisi (E-Dergi). 2,2008,219 – 229.
[6] Örün S, Karatekin C. İstanbul İli Avrupa Yakası İçin Uzun Dönem Elektrik Enerjisi Yük Tahmini. Elektrik-Elektronik-Bilgisayar ve Biyomedikal Mühendisliği Sempozyumu. 2014,38 – 43.
[7] Wilcox RR. Introduction to Robust Estimation and Hypothesis Testing. Academic Press.4(4),1997,2 – 21.
[8] Rencher AC, Schaalje GB. Linear Models in Statistics. John Wiley and Sons. 2008,126 – 184.