ORIGINAL PAPER

COMPARISON OF RESAMPLING METHODS IN MULTIPLE LINEAR REGRESSION

TOLGA ZAMAN[1]*, KAMIL ALAKUS[2]

Manuscript received: 07.02.2018; Accepted paper: 14.11.2018;
Published online: 30.03.2019.

Abstract. *In order to estimate model parameters in multiple regression models, resampling methods of bootstrap and jackknife are used. Resampling methods are used as an alternative readjustment method to the least squares method (OLS) especially when assumptions belonging to error term in regression analysis are not met. Data used in the study are taken from 25 advertisements in Sahibinden.com website and the price of beetle car brand is accepted as dependent variable for multiple linear regression models. It is aimed that price variable taken is tried to be explained with the help of variables of fuel, case type, salesman, sunroof, wind shield, upholstery, age and engine size. When we examined the variables, it is seen that categorical variables are in question and dummy variable must be used. Firstly, model parameters of this obtained data are estimated using OLS and significances of parameters are tested, then, model parameters, significances of estimated parameters, coefficient of determination, and standard error of the model and % 90 confidence intervals are estimated using one of the resampling methods, bootstrap and jackknife method and results belonging to these three methods are compared. Also, generalization condition, which is to the population, of parameter estimation results belonging to explanatory variables used in this study are reviewed with the help of jackknife resampling method ve It has been seen that the salesman and upholstery independent variables have a considerable effect at a significance level of .10 on the dependent variable of price dependence of decision making (p < .10) and Jackknife have confirmed these generalization.*

Keywords: *Bootstrap method, jackknife method, multiple linear regressions, generalization.*

1. INTRODUCTION

Regression analysis is one of the most used methods in determinating functional relation among variables and is a statistical method which provides parameter estimation by determinating relation among variables [1]. Variables in the regression model consist of one dependent and one or more explanatory variables. Models with more than one explanatory variables are called multiple linear regression models. OLS is used in solution of this model.

There are various studies belonging to bootstrap and jackknife method which are one of the resampling methods used in the study. Some of these, bootstrap method were

[1] Cankiri Karatekin University, Department of Statistics, 18100 Cankiri, Turkey.
*Correspondig author E-mail: tolgazaman@karatekin.edu.tr
[2] Ondokuz Mayis University, Department of Statistics, 55139 Samsun, Turkey. E-mail: kamilal@omu.edu.tr

introduced first time by Efron [2] in regression. Later, performance of bootstrap method in regression is examined in most studies. Freedman's study [3] is a culmination of Bickel and Freedman [4] study and it explored asymptotic theory of bootstrap regression. Stine [5], in his study, showed that to form the distribution of residual values in linear regression model, experimental distribution is estimated through giving $1/n$ probability to residual values found. Shao and Tu [6] investigated that bootstrap and jackknife estimators in regression analysis are not affected by deviations which are one of the assumptions. Fox [7], in his study, assumed that because bootstrap method is added to the regression model estimation obtained by OLS through sampling repeating samples from residual terms, residual terms are normally distributed and by forming the experimental distribution of values sample, pointed at the need to determine algorithm of bootstrap applied. In most studies, it is showed that bootstrap approach is valid for OLS parameter estimation distribution. Freedman's study [3] is improved by Wu [8]. Freedman and Peter [9] presented the usage of bootstrap regression method to meet the energy demand in industry. Also, for bootstrap method, Saurbrei and Schumacher [10]; Grist et al. [11]; Chen and George [12]; Desli and Ray [13]; Hardle and Marron [14]; Wu [8] can be explored. Mooney and Duval [15] explained stages of the process and noted that to form sampling distribution, bootstrap method forms repetitions in great number instead of assumptions and analytic formulas and is a nonparametric method developed to remove the inefficacy of sampling method. Sahinler and Topuz [16] examined the algorithm of bootstrap and jackknife resampling to estimate regression parameters. Liu [17], in his study, introduced bootstrap procedures under some independent and not identically distributed models. Rasheed and Algamal [18] studied on linear regression model using jackknife and bootstrap methods. Zaman and Alakuş [19] investigated bootstrap and jackknife methods, which are used as a correction term when assumptions of the error in simple linear regressions are not met in detail. Also, various studies are made on generalizability and repeatability of jackknife method [20-22]. Akpanta and Okorie [23], in their study, examined the significance of correlation coefficient using jackknife estimations.

2. MATERIALS AND METHODS

Multiple linear regression model which is going to be used is defined using following matrix notation;

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{2.1}$$

In this equation

$\underline{Y}$ is a $n \times 1$ sized column vector of dependent variable Y.

X is a $n \times p$ sized matrix of known constants (explanatory variables).

$\underline{\beta}$ is a $p \times 1$ sized column vector of unknown regression coefficents.

$\underline{\varepsilon}$ is a $n \times 1$ sized column vector of random errors.

$Var(\underline{\varepsilon}) = \sigma^2 I_n$, which vector $\underline{\varepsilon}$ is independent and identically distributed, is assumed. Here I_n is a $n \times n$ sized unitary matrix and a constant where $\sigma^2 > 0$ [24]. Regression model parameter estimations here are obtained through OLS with equation

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y} \tag{2.2}$$

Covariance matrix of regression coefficients estimated through OLS are given with

$$Var\left(\underline{\hat{\beta}}\right) = \hat{\sigma}^2(X'X)^{-1}_{p\times p} \tag{2.3}$$

Variances of estimated regression coefficients are base diagonal units of equation (2.3) and values outside of base diagonal are covariances.

2.1. BOOTSTRAP METHOD

The term "bootstrapping" literally means self-starting without the aid of others, and derives from the idiom "pull oneself up by one's own bootstraps". For the first time, it is suggested as an alternative to jackknife method by Bradley Efron [2] while being indicated as easier and safer than jackknife method. It is improved by Efron and Tibshiranni [25]. It is resampling method from original data series [26]. Underlying train of thought of the method is based on forming new data series through replacing and resampling any sized observations in current data series depending on luck. Thus, it is possible to get information as much as possible from the current data set. In other words, firstly, a bootstrap sample is formed through placement method about the size of the data set on hand. Then, a large number of bootstrap sample can be formed in this way [13, 26]. Fundamental aim of bootstrap method is to get the sampling distribution of prediction value and to evaluate the uncertainty of unknown population parameter value based on this distribution [27]. Number of bootstrap sample, which will be formed since replacing method is valid, cannot be more than n^n.

Let us come to the point of usage of bootstrap method in linear regression. There are two approach in bootstrap regression usage. First of these, which is used in the study, is the expression that when regression coefficients are fixed, it uses the resampling of bootstrap error term.

1. Regression model on original data set is estimated and residuals (e_i) are calculated.
2. n sized B bootstrap subsamples are formed by given $\frac{1}{n}$ probability to (e_i) values obtained and bootstrap error term mean is calculated as follows:

$$\bar{\hat{\varepsilon}}_i^* = \frac{\sum_{b=1}^{B} \hat{\varepsilon}_{bi}}{B} \tag{2.4}$$

Here, $\bar{\hat{\varepsilon}}_i^*$is ith bootstrap error estimator and $\hat{\varepsilon}_{bi}$ is ith error estimator beloning to bth bootstrap sample [5, 6, 8, 28].

3. Through putting calculated $\bar{\hat{\varepsilon}}_i^*$ values into the place of e_is in the model formed,

$$Y_i^* = \hat{\beta}X + \bar{\hat{\varepsilon}}_i^* \tag{2.5}$$

bootstrap Y_{boot}^* values are calculated.

4. Based on Y^*_{boot} and Xs, using least squares method, bootstrap estimator of $\underline{\beta}$ is calculated as follows [17]:

$$\hat{\beta}^* = (X'X)^{-1}X'Y^* \tag{2.6}$$

Variance of bootstrap method is calculated as follows:

$$Var\left(\underline{\hat{\beta}}^{*k}\right) = \frac{1}{B-1}\sum_{k=1}^{B}\left(\underline{\hat{\beta}}^{*k} - \underline{\bar{\hat{\beta}}}^{*k}\right)\left(\underline{\hat{\beta}}^{*k} - \underline{\bar{\hat{\beta}}}^{*k}\right)' \tag{2.7}$$

where, it is $\underline{\bar{\hat{\beta}}}^{*k} = \frac{1}{B}\sum_{k=1}^{B}\underline{\hat{\beta}}^{*k}$.

Confidence interval belonging to Bootstrap method based on normal approach is calculated as follows:

$$\underline{\hat{\beta}}^{*k} - t_{n-p-n,\frac{\alpha}{2}} * \widehat{se}_{boot}\left(\underline{\hat{\beta}}^{*k}\right) < \underline{\beta} < \underline{\hat{\beta}}^{*k} + t_{n-p-n,\frac{\alpha}{2}} * \widehat{se}_{boot}\left(\underline{\hat{\beta}}^{*k}\right) \tag{2.8}$$

2.2. JACKKNIFE METHOD

In recent years, when it is not viable to evaluate data with parametric methods, resampling methods are started to be used. One of these methods is Jackknife method. Jackknife method is developed with the intention of minimizing the sampling error in concern with getting narrow confidence intervals in estimating population parameters. This method is also called Pocket-knife Method. Pocket-knife is a hand tool which is to use on various problems. This method is also, like a pocket-knife, a method which is usable to carry out various problems. In this respect, name similarity is really fitting. Also, one of the works which can be done by pocket-knife is whittling. In jackknife method, there is also a similar process to whittling [22]. Jackknife is the first of computer-based methods for estimating bias and standard error. Jackknife resampling method was used with the aim of removing statistical biases in by Maurice Quenouille [29]. Later, it was extended to form hypothesis tests and confidence intervals in by John Tukey and is called as "Jackknife" [30].

In statistics, the jackknife is a resampling technique especially useful for variance and bias estimation. The jackknife predates other common resampling methods such as the bootstrap. The jackknife estimator of an estimator is found by systematically leaving out each observation from a data set and calculating the estimate and then finding the average of these calculations. Given a sample of size n, the jackknife estimate is found by aggregating the estimates of each $n-1$ estimate in the sample.

Let $\emptyset_n(X) = \emptyset_n(X_1, \dots, X_n)$ be an estimator defined for samples $X = (X_1, \dots, X_n)$. i^{th} pseudo-value of $\emptyset_n(X)$ is

$$ps_i(X) = n * \emptyset_n(X_1, \dots, X_n) - (-1)\big(\emptyset_{n-1}(X_1, \dots, X_n)\big)_{[i]} \tag{2.9}$$

In (1), $X_{[i]}$ means the sample $X = (X_1, \dots, X_n)$ with the i^{th} value X_i deleted from the sample, so that $X_{[i]}$ is a sample of size $(n-1)$. Note

$$ps_i(X) = \emptyset_n(X) + (n-1)(\emptyset_n(X) - \emptyset_{n-1}(X_{[i]}) \tag{2.10}$$

So that, $ps_i(X)$ can be viewed as a bias-corrected version of $\emptyset_n(X)$ determined by the trend in the estimators $\emptyset_n(X)$ from $\emptyset_{n-1}(X_{[i]})$ to $\emptyset_n(X)$.

The basic jackknife recipe is to treat the pseudo values $ps_i(X)$ as if they were independent random variables with mean θ. One can then obtain confidence intervals and carry out statistical test using the Central Limit Theorem. Specially, let

$$V_{ps}(X) = \frac{1}{n-1}\sum_{i=1}^{n}(ps_i(X) - ps(X))^2 \tag{2.11}$$

In this equation, $ps(X) = \frac{1}{n}\sum_{i=1}^{n} ps_i(X)$ and jackknife estimation of standard error is given as:

$$se_{Jackk} = \frac{\left\{\frac{\left[\sum_{i=1}^{n}\left(ps_i(X)-ps(X)\right)^2\right]}{n-1}\right\}^{1/2}}{\sqrt{n}} \tag{2.12}$$

and Confidence interval belonging to jackknife method is the following statement:

$$P\left(ps(X) - t_{n-p-n,\frac{\alpha}{2}} * se_{Jackk},\ ps(X) + t_{n-p-n,\frac{\alpha}{2}} * se_{Jackk}\right) = 1 - \alpha \tag{2.13}$$

Also, generalizability states of explanatory variables to population can be explored with jackknife method. There are various studies on this in literature [20-22].

The parameters of regression methods using jackknife method are found as follow:

1. Firstly, $(n-1)$ sized n different subsamples are formed by removing the observations from the data one by one.
2. Regression coefficients belonging to regression methods in interest are estimated for each formed subsample. This regression coefficients are called deleted slope coefficients and is indicated by $\beta_{\mathrm{j}(-i)}$, $\mathrm{i} = 1,2,\dots,\mathrm{n}$; $j = 1,2,\dots,p$
3. Lastly, in order to obtain Jackknife estimator of intercept parameter value, mean of values $(y_i - (\hat{\beta}_{1(-i)})\mathrm{x}_1 + (\hat{\beta}_{2(-i)})\mathrm{x}_2 + \cdots + (\hat{\beta}_{\mathrm{p}(-i)})\mathrm{x}_\mathrm{p})$ is estimated as estimate $\hat{\beta}_{0(-i)}$ as below:

$$\hat{\beta}_{0(-i)} = \frac{\sum_{i=1}^{n}(y_i - (\hat{\beta}_{\mathrm{j}(-i)})\mathrm{x}_1 + (\hat{\beta}_{\mathrm{j}(-i)})\mathrm{x}_2 + \cdots + (\hat{\beta}_{\mathrm{p}(-i)})\mathrm{x}_\mathrm{p})}{n-1} \tag{2.14}$$

4. Mean value of these coefficients that obtained for each subsample are Jackknife estimators and expressed as below,

$$\hat{\beta}_{\mathrm{j}}^{\mathrm{ç}*} = \frac{\sum_{i=1}^{n}(\hat{\beta}_{\mathrm{j}(-i)})}{n};\ \mathrm{j} = 1,2,\dots,\mathrm{p};\ \mathrm{i} = 1,2,\dots,\mathrm{n};\quad \hat{\beta}_0^{\mathrm{ç}*} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}}\hat{\beta}_{0(-\mathrm{i})}}{\mathrm{n}} \tag{2.15}$$

3. REAL DATA APPLICATION

Data used in the study are compiled from 25 advertisings in Sahibinden.com website [31]. Independent variables being used and their levels are given in Table 1. When we examined the variables, it is seen that categorical variables are in question and dummy variable must be used. While dummy variable is being used, one of the levels of categorical variable is kept outside of the model and new variables amounting level count minus one is defined [32].

Table 1. Some independent variables and their levels.

Variable Name	**Code**	**Levels**
Fuel	Fuel	Gasoline*, LPG
Case Type	Type	Cabriolet*, Sedan
Seller	Seller	Sahibinden*, Gallery
Sunroof	Sun	Existent*, Nonexistent
Front window	Gl	Camber*, Flat
Upholstery	Uph	Fabric*, Leather
Age	Quantitive Variable	
Engine Size	1300*, 1600	
* ***Variable levels in question are out of the model to check on.***		

It is aimed that Turkey market price value of Beetle car brand is tried to be explained with the help of variables of fuel, case type, salesman, sunroof, wind shield, upholstery, age and engine size. Since fuel, case type, seller, sunroof, wind shield, upholstery and engine size are explanatory variables, their states in multiple linear regression model are explored using dummy variable.

Thus, regression model aimed to be formed is defined as:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \varepsilon_i \qquad (3.1)$$

Data set belonging to 25 observation is given in Table 2.

Table 2. Data Set (n=25, p=8).

Item Number	**Price**	**Fuel**	**Case Type**	**Seller**	**Sunroof**	**Wind Shield**	**Upholstery**	**Age**	**Engine Size**
1	8.75	0	1	0	0	0	0	42	1
2	27.5	0	0	0	1	0	0	42	1
3	26	1	1	0	0	0	1	43	1
4	23.25	1	1	0	0	0	1	42	1
5	41	0	1	0	1	1	1	55	0
6	35	0	1	0	0	1	1	52	0
7	35	0	1	1	0	1	1	53	0
8	35	1	1	0	1	1	1	51	0
9	30.5	0	1	1	0	1	1	54	0
10	28	0	1	1	0	1	0	53	0
11	14.25	1	1	0	0	0	1	42	1
12	14	0	1	0	0	0	0	42	1
13	13	0	1	0	1	0	0	42	1
14	11.45	0	1	0	0	0	1	42	1
15	37.5	0	0	0	1	0	1	43	1
16	25.5	0	1	0	0	0	0	42	1
17	25	0	1	0	0	1	1	48	1
18	40	0	1	1	0	0	1	43	1
19	28	0	1	0	0	0	1	42	1
20	17.5	0	1	1	0	0	0	43	1
21	14	0	1	0	0	0	1	39	1
22	19.99	1	1	0	0	0	1	43	1
23	16	0	1	0	0	1	1	44	1
24	46.5	0	1	1	1	0	1	43	1
25	19	0	1	0	0	1	0	51	0

Multiple linear regression analysis belonging to the data is resolved using R programming language. First, states of residual terms obtained using estimation model are interpreted in Fig. 1. Then, multiple linear regression results of data set is shown in Table 3 and obtained results are noted.

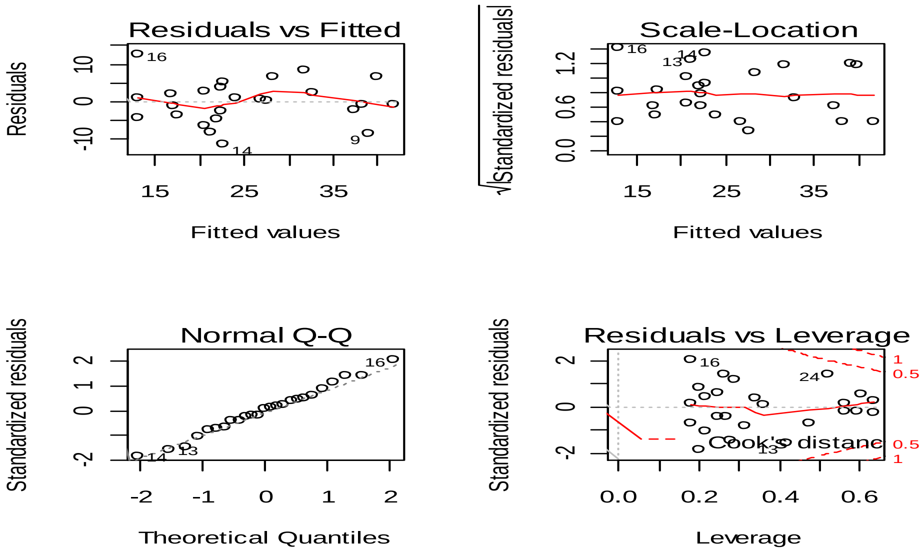

Figure 1. Graphs of residual term obtained through the use of estimation model.

When graphs of residual term in Figure 1 are looked through, it is obvious that there are deviated values in data set and assumptions of error term are not met. In this regard, especially when assumptions belonging to error term in regression analysis are not met, resampling methods are used as a correction method. Thus, it is expected that regression model is better estimated through the use of resampling methods.

Table 3. Multiple Linear Regression Results.

Variable	Non-standardized coefficients		Standardized coefficients	t-test	Significance Probability
	B	Std. Error	Beta		
Constant	-56.912	61.082		-0.932	0.365
X_1	-2.083	4.123	-0.082	-0.505	0.620
X_2	-5.598	6.306	-0.149	-0.888	0.388
X_3	7.281	3.924	0.305	1.855	0.082
X_4	8.224	3.920	0.344	2.098	0.052
X_5	-9.102	7.736	-0.428	-1.176	0.257
X_6	9.708	3.322	0.444	2.922	0.010
X_7	1.734	1.259	0.816	1.378	0.187
X_8	2.615	9.786	0.115	0.267	0.793
$R^2_{Adj} = 0.571; \hat{\sigma} = 6.824$					
Significance test of the model, F=4.89; significance probability is 0.003.					

Results belonging to explanatory variables, which are important for multiple linear regression model, are given in Table 3. When Table 3 is looked through, it is observed that explanatory variables X_3, X_4 and X_6 are statistically significant. ($p < 0.10$). Our multiple linear regression model set up through these variables are also found statistically significant ($p = 0.003 < 0.10$). For model, it is calculated as $R^2_{Adj} = 0.571$. That is to say, explanation rate of explanatory variables (X_3, X_4, X_6), which are found significant in the model, on model variance is calculated as 0.571.

Also, generalizability states of variables found significant in the model to the universe can be explored with jackknife method. Variables found significant in this sense, coefficient of determination of the model and generalizability states of standard error of the model to the universe are tested using jackknife method. In this regard, in Table 4, standardized beta coefficients, which are obtained through regression analysis by leaving out one each observation respectively, are seen.

Table 4. Beta coefficients obtained through Jackknife Method.

	$\hat{\beta}_{-1}$	$\hat{\beta}_{-2}$	$\hat{\beta}_{-3}$	$\hat{\beta}_{-4}$	$\hat{\beta}_{-5}$	$\hat{\beta}_{-6}$	$\hat{\beta}_{-7}$	$\hat{\beta}_{-8}$	R^2_{-Adj}	$\hat{\sigma}_{-}$
None excluded	-0.082	-0.149	0.305	0.344	-0.428	0.444	0.816	0.115	0.571	6.824
1. excluded	-0.098	-0.145	0.300	0.352	-0.48	0.423	0.889	0.136	0.522	6.945
2. excluded	-0.085	-0.092	0.301	0.320	-0.433	0.432	0.818	0.116	0.561	7.042
3. excluded	-0.107	-0.146	0.310	0.356	-0.386	0.439	0.737	0.070	0.572	6.962
4. excluded	-0.098	-0.146	0.302	0.350	-0.42	0.438	0.810	0.114	0.566	7.005
5. excluded	-0.096	-0.144	0.301	0.350	-0.468	0.468	0.851	0.139	0.514	7.042
6. excluded	-0.032	-0.142	0.368	0.391	-0.372	0.402	0.814	0.235	0.583	6.742
7. excluded	-0.085	-0.16	0.307	0.338	-0.416	0.458	0.773	0.085	0.548	7.014
8. excluded	-0.128	-0.169	0.295	0.269	-0.517	0.453	0.974	0.214	0.552	6.985
9. excluded	-0.095	-0.18	0.332	0.296	-0.477	0.476	0.889	0.115	0.617	6.551
10. excluded	-0.084	-0.147	0.280	0.344	-0.416	0.429	0.774	0.110	0.560	7.047
11. excluded	-0.019	-0.16	0.312	0.332	-0.441	0.454	0.821	0.117	0.571	6.801
12. excluded	-0.081	-0.155	0.316	0.353	-0.427	0.441	0.820	0.114	0.539	7.042
13. excluded	-0.110	-0.017	0.284	0.491	-0.435	0.358	0.812	0.162	0.605	6.481
14. excluded	-0.183	-0.125	0.245	0.318	-0.589	0.541	0.875	0.060	0.625	6.262
15. excluded	-0.088	-0.127	0.310	0.33	-0.446	0.462	0.847	0.120	0.534	7.042
16. excluded	-0.049	-0.184	0.360	0.372	-0.334	0.497	0.711	0.073	0.674	6.071
17. excluded	-0.080	-0.15	0.312	0.346	-0.440	0.441	0.721	0.001	0.563	7.034
18. excluded	-0.058	-0.158	0.216	0.375	-0.396	0.410	0.856	0.108	0.580	6.589
19. excluded	-0.036	-0.162	0.335	0.360	-0.355	0.402	0.792	0.141	0.581	6.878
20. excluded	-0.067	-0.147	0.339	0.341	-0.425	0.388	0.843	0.166	0.568	6.905
21. excluded	-0.123	-0.144	0.312	0.347	-0.37	0.497	0.510	-0.093	0.552	6.940
22. excluded	-0.054	-0.151	0.300	0.335	-0.449	0.442	0.863	0.143	0.561	7.014
23. excluded	-0.081	-0.153	0.318	0.352	-0.317	0.449	0.737	0.127	0.547	7.034
24. excluded	-0.060	-0.275	0.183	0.178	-0.498	0.431	1.023	0.107	0.530	6.602
25. excluded	-0.081	-0.147	0.317	0.355	-0.434	0.445	0.843	0.159	0.558	7.015

Pseudo-values belonging to explanatory variables, X_1 X_2, X_3, X_4 X_5, X_6, X_7 and X_8, are obtained as seen in Table 5 with the help of equation (2.9) through the usage of beta estimation values which are standardized in Table 4.

Table 5. Pseudo Values.

Item Number	$\hat{\beta}_1^*$	$\hat{\beta}_2^*$	$\hat{\beta}_3^*$	$\hat{\beta}_4^*$	$\hat{\beta}_5^*$	$\hat{\beta}_6^*$	$\hat{\beta}_7^*$	$\hat{\beta}_8^*$	$\hat{\sigma}^*$
None excluded	-0.082	-0.149	0.305	0.344	-0.428	0.444	0.816	0.115	6.824
1. excluded	0.302	-0.245	0.425	0.152	0.820	0.948	-0.936	-0.389	3.920
2. excluded	-0.010	-1.517	0.401	0.920	-0.308	0.732	0.768	0.091	1.592
3. excluded	0.518	-0.221	0.185	0.056	-1.436	0.564	2.712	1.195	3.512
4. excluded	0.302	-0.221	0.377	0.200	-0.620	0.588	0.960	0.139	2.480
5. excluded	0.254	-0.269	0.401	0.200	0.532	-0.132	-0.024	-0.461	1.592
6. excluded	-1.282	-0.317	-1.207	-0.784	-1.772	1.452	0.864	-2.765	8.792
7. excluded	-0.010	0.115	0.257	0.488	-0.716	0.108	1.848	0.835	2.264
8. excluded	1.022	0.331	0.545	2.144	1.708	0.228	-2.976	-2.261	2.960
9. excluded	0.230	0.595	-0.343	1.496	0.748	-0.324	-0.936	0.115	13.376
10. excluded	-0.034	-0.197	0.905	0.344	-0.716	0.804	1.824	0.235	1.472
11. excluded	-1.594	0.115	0.137	0.632	-0.116	0.204	0.696	0.067	7.376
12. excluded	-0.106	-0.005	0.041	0.128	-0.452	0.516	0.720	0.139	1.592
13. excluded	0.590	-3.317	0.809	-3.184	-0.260	2.508	0.912	-1.013	15.056
14. excluded	2.342	-0.725	1.745	0.968	3.436	-1.884	-0.600	1.435	20.312
15. excluded	0.062	-0.677	0.185	0.680	0.004	0.012	0.072	-0.005	1.592
16. excluded	-0.874	0.691	-1.015	-0.328	-2.684	-0.828	3.336	1.123	24.896
17. excluded	-0.130	-0.125	0.137	0.296	-0.140	0.516	3.096	2.851	1.784
18. excluded	-0.658	0.067	2.441	-0.400	-1.196	1.26	-0.144	0.283	12.464
19. excluded	-1.186	0.163	-0.415	-0.040	-2.180	1.452	1.392	-0.509	5.528
20. excluded	-0.442	-0.197	-0.511	0.416	-0.500	1.788	0.168	-1.109	4.880
21. excluded	0.902	-0.269	0.137	0.272	-1.820	-0.828	8.160	5.107	4.040
22. excluded	-0.754	-0.101	0.425	0.560	0.076	0.492	-0.312	-0.557	2.264
23. excluded	-0.106	-0.053	-0.007	0.152	-3.092	0.324	2.712	-0.173	1.784
24. excluded	-0.610	2.875	3.233	4.328	1.252	0.756	-4.152	0.307	12.152
25. excluded	-0.106	-0.197	0.017	0.08	-0.284	0.42	0.168	-0.941	2.240
Jackk Pseudo Mean	-0.055	-0.148	0.372	0.391	-0.388	0.467	0.813	0.149	6.396
Jackk Pseudo Standard Error	0.164	0.198	0.190	0.246	0.281	0.179	0.459	0.306	1.289

Average Jackknife parameter values related to explanatory coefficient of the model, standard error values related to pseudo-values of Jackknife "t" values related to parameter distribution of Jackknife and finally confidence interval of parameter values are summarized in Table 6.

Table 6. %90 Confidence Intervals (CI) of parameter estimation values calculated through Jackknife Method.

	$\hat{\beta}_1^*$	$\hat{\beta}_2^*$	$\hat{\beta}_3^*$	$\hat{\beta}_4^*$	$\hat{\beta}_5^*$	$\hat{\beta}_6^*$	$\hat{\beta}_7^*$	$\hat{\beta}_8^*$	R^{2*}_{Adj}	$\hat{\sigma}^*$
Original Coefficient	-0.082	-0.149	0.305	0.344	-0.428	0.444	0.816	0.115	0.571	6.824
Pseudo Jackk. Mean	-0.055	-0.148	0.372	0.391	-0.388	0.467	0.813	0.149	0.659	6.396
Pseudo Standard Error Mean	0.164	0.198	0.190	0.246	0.281	0.179	0.459	0.306	0.166	1.289
Lower Bound %90 CI	-0.341	-0.493	0.040	-0.038	-0.878	0.154	0.011	-0.385	0.369	4.145
Upper Bound %90 CI	0.231	0.197	0.703	0.820	0.102	0.779	1.614	0.683	0.948	8.646
t calculation	-0.335	-0.747	1.957	1.589	-1.380	2.608	1.771	0.486	3.969	4.961

Standard error belonging to 25 pseudo values are calculated with equation (2.12). Mean of standard errors obtained later is 6.396 (mean(serror)); standard error of obtained pseudo standard error is 1.289.

In Table 6, it is seen that parameter estimation value of variable X_3 obtained through jackknife method is 0.190, it confirms the first calculated original value, 0.305 ($\left|t_{hesap}\right| = 1.957 > 1.746$) and estimation value of jackknife is in %90 confidence interval. In this regard, it is observed that variable X_3, which is found significant for the model, is not applicable only in this sample but also at the same time is generalizable to the population. It was observed that explanatory variable X_1 is not significant in 0.10 significancy level (see Table 3). This situation is tested with jackknife method and also in this method, it is seen that variable X_1 is not significant in 0.10 significancy level (calculated t value=-0.335, critical t value=1.746). Thus, if effect of explanatory variable X_1 is considered even if a little, this effect can be tought as unique to the sample. However, variable X_7 was observed as non significant in 0.10 significance level (see Table 3). This is tested with Jackknife method and in case of this method, explanatory variable X_7 is found significant in 0.10 significancy level (calculated t value 1.771, critical t value=1.746). In the same regard, evaluations to other explanatory variables can be made.

And finally, generalizability state of explanatory coefficient of the model R^2 and standart error of the model $\hat{\sigma}$ is tested with Jackknife and according to Table 6, because calculated t value (3.969) related to R^2 and calculated t value related to $\hat{\sigma}$ are bigger than critical t value (1.746), coefficient of determination of the model and standart error of the model is not unique to the sample and has the ability to generalize to the population. In other words, it can be said that in the studies that will be made in different times, obtained value will be same or close. Likewise, it can be thought that explanatory variables, which are significant on price response variable on deciding and are able to generalize to the population, can be determined as significant in other studies too.

Now, let us carry out our processes taking account of calculation algorithm of bootstrap method given in chapter (2.1). Obtained results are as follows.

Table 7. Residuals obtained from regression model.

Item Number	Price	Fuel	Case Type	Seller	Sunroof	Wind Shield	Upholstery	Age	Engine Size	e_i	$e_i/25$
1	8.75	0	1	0	0	0	0	42	1	-4.20687	-0.17
2	27.5	0	0	0	1	0	0	42	1	0.72097	0.03
3	26	1	1	0	0	0	1	43	1	3.68465	0.15
4	23.25	1	1	0	0	0	1	42	1	2.66898	0.11
5	41	0	1	0	1	1	1	55	0	-0.71804	-0.03
6	35	0	1	0	0	1	1	52	0	6.70907	0.27
7	35	0	1	1	0	1	1	53	0	-2.30602	-0.09
8	35	1	1	0	1	1	1	51	0	2.30273	0.09
9	30.5	0	1	1	0	1	1	54	0	-8.54035	-0.34
10	28	0	1	1	0	1	0	53	0	0.40159	0.02
11	14.25	1	1	0	0	0	1	42	1	-6.33102	-0.25
12	14	0	1	0	0	0	0	42	1	1.04313	0.04
13	13	0	1	0	1	0	0	42	1	-8.18100	-0.33
14	11.45	0	1	0	0	0	1	42	1	-11.2144	-0.45
15	37.5	0	0	0	1	0	1	43	1	-0.72097	-0.03
16	25.5	0	1	0	0	0	0	42	1	12.54313	0.50
17	25	0	1	0	0	1	1	48	1	1.03134	0.04
18	40	0	1	1	0	0	1	43	1	8.32043	0.33
19	28	0	1	0	0	0	1	42	1	5.33552	0.21
20	17.5	0	1	1	0	0	0	43	1	-4.47196	-0.18
21	14	0	1	0	0	0	1	39	1	-3.46149	-0.14
22	19.99	1	1	0	0	0	1	43	1	-2.32535	-0.09
23	16	0	1	0	0	1	1	44	1	-1.03134	-0.04
24	46.5	0	1	1	1	0	1	43	1	6.59630	0.26
25	19	0	1	0	0	1	0	51	0	2.15101	0.09

If Table 7 is explored, firstly, regression residual values e_i are calculated with classical OLS. Then, each obtained e_i value is given $1/25$ probability. One of the resampling methods, bootstrap method, is applied on obtained 25 $e_i/25$ values. In table 8, there are 25 sized 1000 bootstrap sample of this value.

Table 8. Samples belonging to residuals and calculated estimation values.

1.Bootst Samp.	2.Bootst Samp.	3.Bootst Samp.	…	1000.Bootst Samp.	$\bar{\hat{\varepsilon}}_i^*$	Y_i^*
-0.09	0.09	-0.34		0.26	0.00547	12.94847
-0.33	0.03	0.33		0.02	0.00089	26.76589
0.04	-0.33	0.50		-0.45	-0.00924	22.29276
-0.09	0.03	0.03		-0.03	-0.00691	20.56109
0.21	0.50	0.15		-0.09	0.00714	41.70714
0.27	-0.04	0.02		0.15	0.00306	28.27706
0.02	0.26	0.09		0.50	0.00061	37.28961
0.21	0.03	-0.03		0.04	0.00814	32.68914
0.26	0.27	-0.25		-0.18	-0.01261	39.01039
0.09	-0.03	-0.03		-0.18	0.00679	27.58779
0.15	0.04	0.21		-0.25	0.00666	20.57466
0.09	-0.14	-0.17		-0.34	0.00461	12.94761
-0.33	-0.14	-0.14		0.09	0.00256	21.16956
0.09	-0.18	-0.34		0.33	-0.00655	22.64445
0.50	0.27	-0.33		-0.14	0.00975	38.21675
0.09	-0.25	-0.03		-0.14	0.00361	12.94661
0.04	0.33	-0.14		0.02	0.00046	23.95346
0.02	-0.45	0.26		0.09	-0.00206	31.66394
0.21	-0.09	-0.18		-0.14	0.00232	22.65332
0.27	-0.33	-0.25		0.09	-0.00607	21.95193
0.26	0.09	0.33		-0.45	-0.00101	17.44799
-0.09	0.03	-0.18		0.09	-0.00551	22.29649
-0.09	0.04	0.09		0.15	-0.00116	17.01584
0.03	0.02	-0.09		-0.17	-0.0025	39.8875
0.03	0.21	0.21	…	-0.34	-0.01402	16.81798

In table 8, there are 25 sized 1000 bootstrap sample of this value. Taking account of the bootstrap sample, $\bar{\hat{\varepsilon}}_i^{\,*}$ value and with the help of this value, Y^* value is calculated. $\bar{\hat{\varepsilon}}_i^*$ value is mean of column values in Table 8. In other words, it is mean of the first values in bootstrap example.

Value $\bar{\hat{\varepsilon}}_1^* = 0.00547$ shows residual estimator and it is derived through dividing the first values in 1000 bootstrap sample with 1000. In the same way $\bar{\hat{\varepsilon}}_{25}^* = -0.0142$ value is 25. Bootstrap shows the residual estimator and it is derived through dividing sum of 25th values in 1000 bootstrap sample obtained with 1000.

Y_{boot}^*: regression model derived from OLS is stated as $+\bar{\hat{\varepsilon}}_i^*$. In other words,

$$Y_i^* = -56.902 - 2.083Fuel - 5.598Casetype + 7.281Seller + 8.224\,Sunroof - 9.102Windshield + 9.708Upholstery + 1.734Age + 2.615Enginesize + \bar{\hat{\varepsilon}}_i^*$$

Since OLS regression is applied to obtained Y_{boot}^* value and independent variables obtained values are bootstrap estimation based on resampling of error term.
If we summarize the results found with bootstrap method in Table 9.

Table 9. %90 Confidence Intervals (CI) of parameter estimation values calculated through Bootstrap Method.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
BootMean	-0.097	-0.176	0.360	0.408	-0.507	0.525	0.965	0.136
Bootserror	0.004	0.006	0.004	0.004	0.008	0.003	0.001	0.100
Lower Bound %90 CI	-2.092	-5.610	7.271	8.222	-9.116	9.701	1.732	2.598
Upper Bound %90 CI	-2.077	-5.587	7.285	8.236	-9.088	9.713	1.736	2.633
$\hat{\sigma} = 0.0070$								

When Table 9 is explored, it is seen that all regression coefficients, which are estimated in multiple regression through bootstrap method, are significant for the model. Standard error of formed regression model is 0.0070 and is a very small value. And this shows the success of results obtained through bootstrap estimation.

4. CONCLUSIONS

In the study, comparison of bootstrap and jackknife method, which are resampling methods, and OLS, which is one of the classical methods, is made with the help of results derived from real life data. Usage of resampling methods, bootstrap and jackknife on multiple linear regression is introduced in detail. Coefficient of determination belonging to these methods, standard error of the model, statistical significance of derived model parameters and confidence intervals are calculated.

In consequence of calculations made considering resampling method of Jackknife, standard error mean belonging to the model is calculated as 6.396. Explanatory variables X_3, X_6 and X_8 are found significant for the model. Also, generalizabilition states of explanatory variables used in the study are explored with jackknife method and are evaluated. Standard error estimation belonging to the model with resampling method of bootstrap is minimized to the 0.007. Also, in all explanatory variables in the study, it is found that beetle car brand is significant in the price of Turkey market. Number of samples used for bootstrap method is 1000. When examined, even if finding result through working on 1000 sample is pretty hard, by means of progressing computer technology, this can be overcome. R program is used in bootstrap calculations. It is especially showed that researchers investigate the generalization of their findings with the Jackknife parameter estimator for experimental methods where, for due to some limitations, the number of samples is small. In this way, it is thought that the findings obtained can shed light on subsequent studies.

As a result, in the data structure of this study, it is seen that bootstrap method, which is one of the resampling methods used as a correction method in situations that assumptions belonging to error term are not met, shows results better than jackknife and OLS. In the application part, since the usage of bootstrap and jackknife methods in multiple linear regression is given in detail, it can be used as a reference for similar studies.

REFERENCES

[1] David, A., *Introduction to Statistics,*West Publishing Company, St. Paul, 1981.
[2] Efron. B., *The Annals of Statistics*, **7**, 1, 1979.
[3] Freedman, D.A., *Annals of Statistics*, **1**, 1218, 1981.
[4] Bickel, P., Freedman, D.A., *The Annals of Statistics*,1196, 1981.
[5] Stine, R., *J. Amer. Statist. Assoc.*, **80**, 1026, 1985.
[6] Shao, J., Tu, D., *The Jackknife and Bootstrap*, Springer- Verlag, New York, 1995.
[7] Fox, J., *Applied Regression Analysis, Linear Models and Related Methods,* Sage, 1977
[8] Wu, C.F.J., *Annals of Statistics*, **14**, 1261, 1986.
[9] Freedman, D.A., Peters, S.C., *Journal of American Statistical Association*, **79**, 97, 1984
[10] Sauerbrei, W., Schumacher, M.A., *Statistics in Medicine*, **11**, 2093, 1992.
[11] Grist, E.P.M, Leung, K.M.Y., Wheeler, J.R., Crane, M., *Enviromental Toxicology and Chemistry*, **21**, 1515, 2002.
[12] Chen, C.H., George, S.L., *Statistics in Medicine*. **4**, 39, 1985.
[13] Desli, E., Ray, S.C., *Indian Economic Review*. **39**, 89, 2004.
[14] Hardle, W., Marron, J.S., *The Annals of Statistics* 1991, **19**, 778, 1991.
[15] Mooney, C.Z., Duval, R.D., *Bootstrapping: A nonparametric Approach to Statistical Inference*. Sage Uni Papers, 1993
[16] Sahinler, S., Topuz, D., *Journal of Applied Quantitative Methods, 2*, 188, 2007
[17] Liu, Y.R., *Ann. of Stat*., **16**, 1696, 1988.
[18] Algamal, Z.Y., Rasheed, K.B., *Iraqi Journal of Statistical Science*, **18**, 59, 2010.
[19] Zaman, T., Alakus K., *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, **30**, 63, 2016.
[20] Bekiroğlu, N., Konyalıoğlu, R., ve Karahan, N., *Marmara Medical Journal*. **26**, 63, 2013.
[21] Kayri, M., Büyüköztürk, Ş., *Educational Science: Theory & Practice,* **9**, 1751, 2009.
[22] Zaman, T., Alakus K., *Open Journal of Statistics*, **5**, 645, 2015.
[23] Akpanda, A., Okorie, I., *International Journal of Sciences: Basic and Applied Research (IJSBAR),* **22**, 441, 2015.
[24] Dallah, H., *International Journal of Applied Science and Technologh*, **2**, 114, 2012.
[25] Efron, B. and Tibshiranni, R., *An introduction to The Bootstrap*, Chapman & Hall, New-York, 1993
[26] Chernick, M. R., *Bootstrap Methods: A Guide for Practitioners and Researchers*. 2. ed., John Wiley & Sons, Inc., New Jersey, 2008.
[27] Tercan, A. E., *Scientific Mining Journal*, **41**, 35, 2002.
[28] Stine, R., *Modern Methods of Data Analysis*; Edit: by John Fox, Scotland Sage Pub, 1990.
[29] Quenouille, M. H., *Biometrika*, **61**, 353, 1956.
[30] Tukey, J. W., *Ann Math Stat*, **29**, 614, 1958.
[31] https://www.sahibinden.com/
[32] Gujarati, N. D., Porter, D. C., *Temel Ekonometri*. Çeviri: Ümit Şenesen, Gülay Günlük Şenesen, Literatür Yayıncılık, Istanbul, 2009