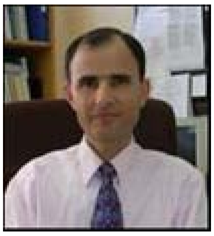
BOOTSTRAP AND JACKKNIFE RESAMPLING ALGORITHMS FOR ESTIMATION OF REGRESSION PARAMETERS

Suat SAHINLER[1]
Biometry and Genetics Unit, Department of Animal Science,
Agriculture Faculty, University of Mustafa Kemal, 31100, Hatay, Turkey

E-mail: sahinler@mku.edu.tr



Dervis TOPUZ
Vocational School of Ulukisla, University of Nigde, Nigde, Turkey

E-mail: dervis_topuz@hotmail.com

Abstract: *In this paper, the hierarchical ways for building a regression model by using bootstrap and jackknife resampling methods were presented. Bootstrap approaches based on the observations and errors resampling, and jackknife approaches based on the delete-one and delete-d observations were considered. And also we consider estimating bootstrap and jackknife bias, standard errors and confidence intervals of the regression coefficients, and comparing with the concerning estimates of ordinary least squares. Obtaining of the estimates was presented with an illustrative real numerical example. The jackknife bias, the standard errors and confidence intervals of regression coefficients are substantially larger than the bootstrap and estimated asymptotic OLS standard errors. The jackknife percentile intervals also are larger than to the bootstrap percentile intervals of the regression coefficients.*

Key words: *bootstrap; jackknife; resampling; regression*

Introduction

Regression analysis is a statistical analysis technique that characterizes the relationship between two or more variables for prediction and estimation by a mathematical model called regression model. Finding estimates of bias and variance of the estimator $\hat{\beta}$ in estimation β and constructing confidence intervals for β and prediction intervals for a future observation with explanatory variables x_i are also interested in. Let the linear regression model be $y = X\beta + \varepsilon$ with the variance $\text{var}(y) = \sigma^2$, where $y=(y_1,y_2,...y_n)`$ denotes the nx1 vector of the response, $X=(x_1,x_2,...x_n)`$ is matrix of regressors with nxp

dimension including intercept, p is the number of parameters, ε_i is an nx1 vector of uncorrelated error terms of zero mean and identical variance σ^2 (Fox,1997; Sahinler and Bek, 2006). Then the least squares estimator $\hat{\beta} = (X^{/} X)^{-1} X^{/} Y$ has variance-covariance matrix $Var(\hat{\beta}) = \sigma^2 (X^{/} X)^{-1}$ and 100(1-α) % confidence intervals $\hat{\beta}_j \pm t_{n-p,\alpha/2} * S_e(\hat{\beta}_j)$.
Traditional approaches, like ordinary least squares, rely on some major modelling assumptions strongly. Although they are provided, the conclusions are based on asymtotical or approximate properties frequently. The reliability ot the satatistical analysis depends therefore on the validity of these assumptions and on the sample size. There are several useful methods for diagnosing and treating violations of the regression assumptions. Robust estimation strategies and residual diagnostics have improved the usefulness of these tecniques (Sahinler, 2000). However, they may not be provided these assumptions by using these methods.

The observed data was considered as a representative picture of the entire population in resampling methods. Hence, the main idea to make statistical inference based on an artificial resample, which is drawn from the full sample (Friedl and Stampfer, 2002b). The ordinary sampling techniques use some assumptions related to the form of the estimator distribution, but resampling methods do not need these assumptions because the sample is thought as population. The bootstrap and jackknife are nonparametric and specific resampling techniques that purpose of deriving estimates of standard errors and confidence intervals of a population parameter like a mean, median, proportion, odds ratio, correlation coefficient or regression coefficient calculations without making distributional assumptions when those assumptions are in doubt, or where parametric inference is impossible or requires very complicated formulas for the calculation of standard errors (Efron, 1982).

This study focuses on illustration and aplication of resampling techniques in regression analysis. Some hyerarchical algoritms of concerning techniques in regression analysis are demonstrated. The basics of the bootstrap and jackknife resampling techniques and their applications to the real numerical example that can be described by linear regression model were discussed and compared the results with ordinary least squares regression results.

Materials and Methods

Material. The aim of the following study is to illustrate the bootstrap and jackknife regression parameter estimation as the methodology in method. The real data produced in the fisheries study in Mustafa Kemal University (Turkey) was used as material. Amongst others, the Total Length of fish (TL) and Otolith Length (OL) were considered as independent variables in order to explain the variation in Fish Age (FA) of n=100 fish related to a fish species (Can and Sahinler, 2005). The statistical packages S-PLUS FOR WINDOWS was used for the statistical analysis of these data.

Method. To describe the resampling methods we start with an n sized sample $w_i = (Y_i, X_{ji})'$ and assume that w_is are drawn independently and identically from a distribution of F, where $Y_i = (y_1, y_2, ..., y_n)'$ contains the responses, $X_{ji} = (x_{j1}, x_{j2}, x_{j3}..., x_{jn})'$ is a matrix of dimension n x k, where j=1,2,...k, i=1,2,3,...,n.

Bootstrapping Regression Algoritm. Here, two approachs for bootstrapping regression methods were given. The coise of either methods depends upon the regressors are fixed or random. If the regressors are fixed, the bootstrap uses resampling of the error term. If the regressors are random, the bootstrap uses resampling of observation sets w_i (Stine, 1990; Shao, 1996).

Bootstrap Based On The Resampling Observations. This approach is usually applied when the regression models built from data have regressors that are as random as the response. Let the (k+1)x1 vector $w_i = (y_i, x_{ji}')'$ denote the values associated with *i*th observation. In this case, the set of observations are the vectors $(w_1, w_2, ..., w_n)$. The bootstrap procedure based on the resampling observations is as follows.

1$^{(o)}$. Draw a n sized bootstrap sample $(w_1^{(b)}, w_2^{(b)}, ..., w_n^{(b)})$ with replacement from the observations giving 1/n probability each w_i values and label the elements of each vector $w_i^{(b)} = (y_i^{(b)}, x_{ji}^{(b)})'$, where j=1,2,...k, i=1,2,...n. From these form the vector $Y_i^{(b)} = (y_1^{(b)}, y_2^{(b)}, ..., y_n^{(b)})'$ and the matrix $X_{ji}^{(b)} = (x_{j1}^{(b)}, x_{j2}^{(b)}, ..., x_{jn}^{(b)})'$

2$^{(o)}$. Calculate the OLS coefficients from the bootstrap sample:

$$\hat{\beta}^{(b1)} = (X^{(b)'}X^{(b)})^{-1}X^{(b)'}Y^{(b)} \quad (1)$$

3$^{(o)}$. Repeat steps 1 and 2 for r=1,2,...,B, where B is the number of repetition.

4$^{(o)}$. Obtain the probability distribution ($F_{(\hat{\beta}^{(b)})}$) of bootstrap estimates $\hat{\beta}^{(b1)}, \hat{\beta}^{(b2)}, ..., \hat{\beta}^{(bB)}$ and use the ($F_{(\hat{\beta}^{(b)})}$) to estimate regression coefficients, variances and confidence intervals as follows. The bootstrap estimate of regression coefficient is the mean of the distribution $F_{(\hat{\beta}^{(b)})}$ (Fox,1997),

$$\hat{\beta}^{(b)} = \sum_{b=1}^{B} \hat{\beta}^{(br)} / B = \overline{\hat{\beta}}^{(br)} \quad (2)$$

5$^{(o)}$. Thus, the bootstrap regression equation is

$$\hat{Y} = X\hat{\beta}^{(b)} + \varepsilon \quad (3)$$

where $\hat{\beta}^{(b)}$ is unbiased estimator of β (Shao,1995).

An illustrative example that presents how the regression parameters are estimated from the bootstrap based on the the resampling observations was given in Table 1.

Bootstrap Based On The Resampling Errors. If the regressors are fixed, as in desing experiment, then the bootstrap resampling must preserve that structure. The bootstrap procedure based on the resampling errors as follows.

1$^{(e)}$. Fit the least squares regression equation for full sample.

2$^{(e)}$. Calculate the e_i values ($e_i = Y_i - \hat{Y}_i$).

3$^{(e)}$. Draw a n sized bootstrap random sample with replacement $(e_1^{(b)}, e_2^{(b)}, ..., e_n^{(b)})$ from the e_i values calculated in step 2$^{(e)}$ giving 1/n probability each e_i values(Stine, 1985; 1990; Wu,1986)

4(e). Compute the bootstrap Y values by adding resampled residuals onto the ordinary least squares regression fit, holding the regression desing fixed(Liu,1988; Leger et al,1992):

$$Y^{(b)} = X\hat{\beta} + e^{(b)} \quad (4)$$

5(e). Obtain least squares estimates from the 1th bootstrap sample:

$$\hat{\beta}^{(b1)} = (X^{/} X)^{-1} X^{/} Y^{(b)} \quad \text{(we need Y*) or} \quad (5)$$

$$= \hat{\beta} + (X^{/} X)^{-1} X^{/} e^{(b)} \quad \text{(we don not need Y*)} \quad (6)$$

6(e). Repeat steps 3(e),4(e) and 5(e) for r=1,2,...,B, and proceed as in resampling with random regressors 4(o) and 5(o).

The bootstrap bias, variance, confidence and percentile interval. The bootstrap bias equals,

$$bi\hat{a}s_b = \hat{\beta}^{(b)} - \hat{\beta} \quad (7)$$

(Further discussion are described in Efron and Tibshirani, 1993). The bootstrap variance from the distribution F$_{(\hat{\beta}^{(b)})}$ are calculated by (Liu, 1988; Stine 1990)

$$\text{var}(\hat{\beta}^{(b)}) = \sum_{b=1}^{B} \left[\left(\hat{\beta}^{(br)} - \hat{\beta}^{(b)}\right)\left(\hat{\beta}^{(br)} - \hat{\beta}^{(b)}\right)' \right] / (B-1), \quad r=1,2,...,B \quad (8)$$

The bootstrap confidence interval by normal approach is obtained by

$$\hat{\beta}^{(b)} - t_{n-p,\alpha/2} * S_e(\hat{\beta}^{(b)}) < \beta < \hat{\beta}^{(b)} + t_{n-p,\alpha/2} * S_e(\hat{\beta}^{(b)}) \quad (9)$$

where $t_{n-p,\alpha/2}$ is the critical value of t with probability α/2 the right for n-p degrees of freedom; and Se($\hat{\beta}^{(b)}$) is the standard error of the $\hat{\beta}^{(b)}$. If sample size is n ≥30, then Z-distribution values are used instead of t in estimation of confidence intervals (Diciccio and Tibshirani, 1987).

A nonparametric confidence interval named percentile Interval can be constructed from the quantiles of the bootstrap sampling distribution of $\hat{\beta}^{(b)}$. The (α/2)% and (1-α/2)% percentile interval is

$$\hat{\beta}^{(br)}{}_{(lower)} < \beta < \hat{\beta}^{(br)}{}_{(upper)} \quad (10)$$

where $\hat{\beta}^{(br)}$ is the ordered bootstrap estimates of regression coefficient from Equation 2 or 5, lower=(α/2)B, and upper = (1-α/2)B.

Jackknifing Regression Algoritm. Here, two algoritm for Jackknifing regression models based on the resampling observations were given. These approachs are usually applied when the regression models built from data have fixed explanatory variables. There are two cases of jackknife resampling. First of them is based on the deleting single case from the original sample (delete one jackknife), and second is based on the deleting multiple case from the original sample (delete d jackknife) sequentially (Efron and Gong, 1983; Wu, 1986; Shao and Tu, 1995). Let the px1 vector $w_i = (y_i, x_{ji}')'$, (i=1,2,...n) denote the values associated with *i*th observation. In this case, the set of observations are the vectors ($w_1, w_2, \ldots, w_n$).

Steps of The Algoritms for Delete-One Jackknife Regression. The jackknife procedure based on delete-one (do) is as follows.

1[do]. Draw n sized sample from population randomly and label the elements of the vector $w_i = (Y_i \ , X_{ji})'$ as the vector $Y_i = (y_1, y_2, ..., y_n)'$ and the matrix $X_{ji} = (x_{j1}, x_{j2}, x_{j3} ..., x_{jn})'$ where j=1,2,...k, i=1,2,3,...,n.

2[do]. Omit first row of the vector $w_i = (Y_i \ , X_{ji})'$ and label remaining n-1 sized observation sets $Y_i^{(J)} = (y_2^{(J)}, ..., y_n^{(J)})'$ and $X_{ji}^{(J)} = (x_{j2}^{(J)}, x_{j3}^{(J)} ..., x_{jn}^{(J)})'$ as delete-*one* Jackknife sample ($w_1^{(J)}$) and estimate the OLS regression coefficients $\hat{\beta}^{(J_1)}$ from ($w_1^{(J)}$). Then, omit second row of the vector $w_i = (Y_i \ , X_{ji})'$ and label remaining n-1 sized observation sets $Y_i^{(J)} = (y_1^{(J)}, y_3^{(J)} ..., y_n^{(J)})'$ and $X_{ji}^{(J)} = (x_{j1}^{(J)}, x_{j3}^{(J)}, ..., x_{jn}^{(J)})'$ as $w_2^{(J)}$ and estimate the OLS regression coefficients $\hat{\beta}^{(J_2)}$. Similarly, omit each one of the n observation sets and estimate the regression coefficients as $\hat{\beta}^{(J_i)}$ alternately, where $\hat{\beta}^{(J_i)}$ is Jackknife regression coefficient vector estimated after deleting of i^th^ observation set from w_i.

3[do]. Obtain the probability distribution $F_{(\hat{\beta}^{(J)})}$ of Jackknife estimates $\hat{\beta}^{(J_1)}, \hat{\beta}^{(J_2)}, ..., \hat{\beta}^{(J_n)}$

4[do]. Calculate the jackknife regression coefficient estimate which is the mean of the $F_{(\hat{\beta}^{(J)})}$ distribution (Fox,1997) as;

$$\hat{\beta}^{(J)} = \sum_{i=1}^{n} \hat{\beta}^{(J_i)} / n = \overline{\hat{\beta}}^{(J_i)} \tag{11}$$

5[do]. Thus, the delete-*one* Jackknife regression equation is

$$\hat{Y} = X\hat{\beta}^{(J)} + \varepsilon \tag{12}$$

An illustrative study which shows how the delete-*one* jackknife regression parameters are estimated was given in Table 2.

Steps of The Algoritms for Delete-*d* Jackknife Regression. The jackknife procedure based on delete-*d* (dd) is as follows.

1[dd]. Draw n sized sample (w_1,w_2,…,w_n) from population randomly and devide the sample into s independent groups of which size is d.

2[dd]. Omit first d observation set from full sample at a time and estimate the OLS coefficients $\hat{\beta}^{(J_1)}$ from (n-d) sized remaining observation set called delete-*d* jackknife sample (Wu, 1986).

3[dd]. Omit second d observation set from full sample at a time and estimate the OLS coefficients $\hat{\beta}^{(J_2)}$ from (n-d) sized remaining observation set.

4[dd]. Omit each d of the n observation sets and estimate the regression coefficients as $\hat{\beta}^{(J_k)}$ alternately, where $\hat{\beta}^{(J_k)}$ is jackknife regression coefficient vector estimated after deleting of k^th^ d observation set from full sample. Thus, $S = \binom{n}{d}$ delete-*d* jackknife samples are obtained, k=1,2,…,s.

6(dd). Obtain the probability distribution $F_{(\hat{\beta}^{(J)})}$ of delete-*d* jackknife estimates $\hat{\beta}^{(J_1)}, \hat{\beta}^{(J_2)}, ..., \hat{\beta}^{(J_s)}$

7(dd). Calculate the jackknife regression coefficient estimate which is the mean of the $F_{(\hat{\beta}^{(J)})}$ distribution as;

$$\hat{\beta}^{(J)} = \sum_{k=1}^{s} \hat{\beta}^{(J_k)} / s = \overline{\hat{\beta}}^{(J_k)} \qquad (13)$$

8(dd). Thus, the delete-*d* Jackknife regression equation is

$$\hat{Y} = X\hat{\beta}^{(J)} + \varepsilon \qquad (14)$$

Jackknife bias, variance, confidence and percentile interval. The jackknife bias, variance and confidence intervals are estimated by using the following equations from $F_{(\hat{\beta}^{(J)})}$ distribution (Miller, 1974).

The jackknife bias equals,

$$bi\hat{a}s_J(\hat{\beta}) = (n-1)(\hat{\beta}^{(J)} - \hat{\beta}) \qquad (15)$$

The jackknife variance equals,

$$\text{var}(\hat{\beta}^{(J)}) = \frac{(n-1)}{n} \sum_{i=1}^{n} \left(\hat{\beta}^{(Ji)} - \hat{\beta}^{(J)}\right)\left(\hat{\beta}^{(Ji)} - \hat{\beta}^{(J)}\right)' \qquad (16)$$

where $\hat{\beta}_j^{(Ji)}$ is the estimate produced from the replicate with i^th observation set or j^th group deleted (Friedl and Stampfer, 2002a).

Jackknife (1-α) 100 % confidence interval equals (Efron and Tibshirani, 1993).

$$\hat{\beta}^{(J)} - t_{n-p,\alpha/2} * S_e(\hat{\beta}^{(J)}) < \beta < \hat{\beta}^{(J)} + t_{n-p,\alpha/2} * S_e(\hat{\beta}^{(J)}) \qquad (17)$$

where $t_{n-p,\ \alpha/2}$ is the critical value of t with probability α/2 the right for n-p degrees of freedom; and Se($\hat{\beta}^{(J)}$) is the standard error of the $\hat{\beta}^{(J)}$.

The jackknife percentile Interval can be constructed from the quantiles of the jackknife sampling distribution of $\hat{\beta}^{(J)}$. The (α/2)% and (1-α/2)% percentile interval is

$$\hat{\beta}^{(J)}_{(lower)} < \beta < \hat{\beta}^{(J)}_{(upper)} \qquad (18)$$

where $\hat{\beta}^{(J)}$ is the ordered jackknife estimates of regression coefficient from Equation 11 or 13, lower=(α/2)n, and upper = (1-α/2)n.

Results

First, the ordinary least squares regression model was fitted to data given in Figure 1 and the results of the ordinary least squares regression was summarized in Table 1. The regression of FA on TL and OL is significant as result of variance analysis (P<0.01**). According to the t-tests for significance of regression coefficients, all of the regression coefficients are significant (P<0.01).

Table 1. The summary statistics of regression coefficients for OLS regression

Variables	$\hat{\beta}$	S.E.($\hat{\beta}$)	t	Sig.	95 % Confidence Interval
Constant	-2.16133	0.178	-12.126	0.000	-2.4538, -1.8682
TL	0.08336	0.034	2.421	0.017	0.0271, 0.1389
OL	0.49573	0.084	5.913	0.000	0.3578, 0.6342
R^2=0.867, N=100, s^2=0.233, SSE=31.491, F=442.3**					

The data and fitted line was given in Figure 1.

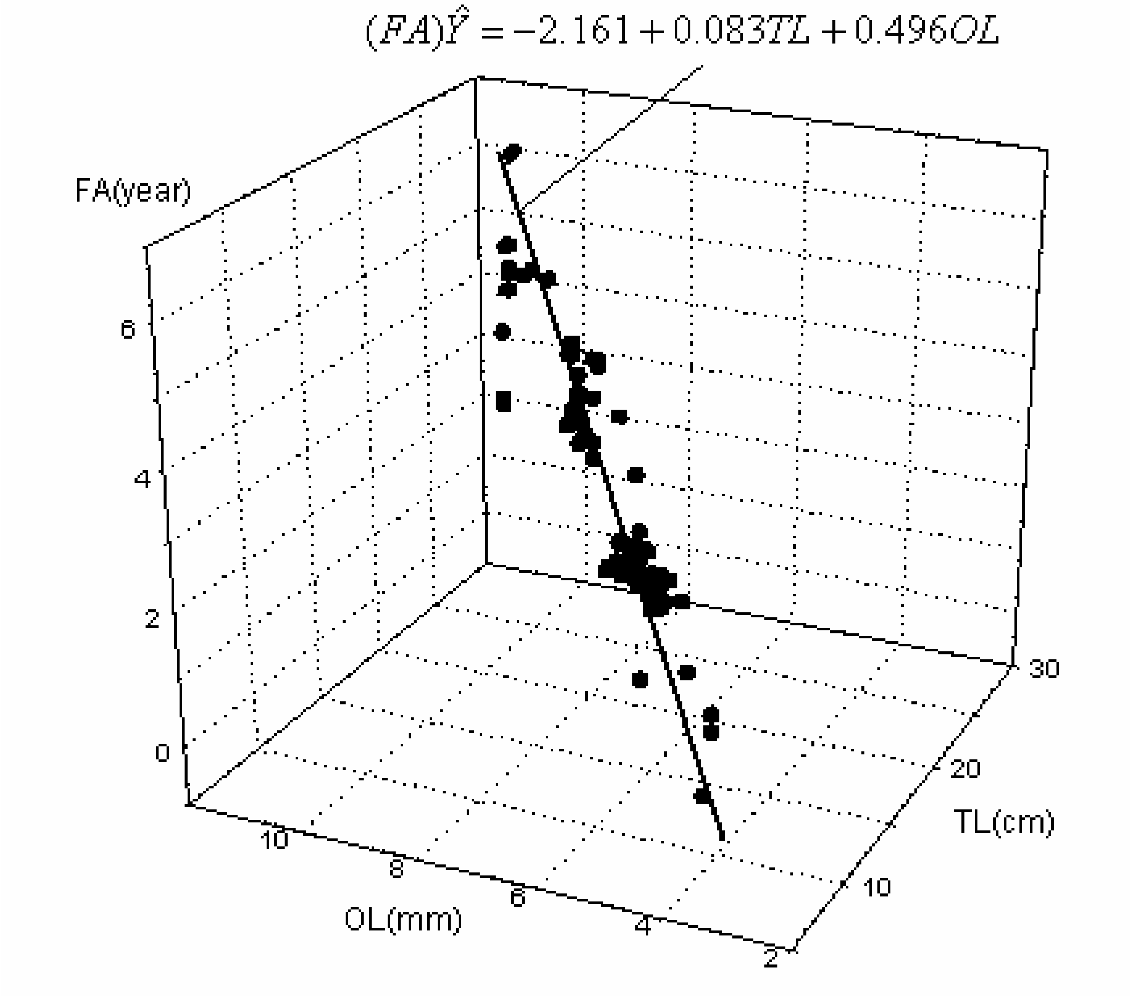


Figure 1. The data and fitted OLS regression line

The illustration of the bootstrap (B=10000 bootstrap samples, each of size n=100) and the jackknife (jackknife samples, each of size n-1=100-1=99) regression procedure, from the data given in Figure 1, calculating the bootstrap and jackknife estimates of the regression parameters for each sample are shown in Table 2 and 3.

Table 2. The illustration of the bootstrap (B=10000 bootstrap samples, each of size n=100) regression procedure from the data given in Figure 1, calculating the bootstrap estimates of the regression parameters for each sample for fish age model

r	Variables	$w_1^{(b)}$,	$w_2^{(b)}$,	$w_3^{(b)}$,	…	$W_{100}^{(b)}$,	$\hat{\beta}_o^{(b)}$	$\hat{\beta}_1^{(b)}$	$\hat{\beta}_2^{(b)}$
1	FA(year)(Y) TL(cm) (X_1) OL(mm) (X_2)	1.16 10.00 4.10	1.84 13.90 5.70	0.92 10.00 4.10	… … …	3.41 19.7 8.10	-2.183	0.083	0.487
2	FA(year)(Y) TL(cm) (X_1) OL(mm) (X_2)	5.08 22.10 9.10	0.92 10.00 4.10	2.25 15.90 6.50	… … …	5.08 22.10 9.10	-2.179	0.081	0.495
3	FA(year)(Y) TL(cm) (X_1) OL(mm) (X_2)	3.16 20.70 8.50	2.08 13.00 5.40	0.08 9.30 4.10	… … …	4.25 25.90 10.90	-2.191	0.080	0.491
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
10000	FA(year)(Y) TL(cm) (X_1) OL(mm) (X_2)	0.08 9.30 4.10	4.16 21.20 8.50	5.08 22.10 9.10	… … …	0.92 10.00 4.10	-2.162	0.084	0.498
$\hat{\beta}^{(b)} = \sum_{b=1}^{B} \hat{\beta}^{(br)} / B = \overline{\hat{\beta}}^{(br)}$							-2.1589	0.0834	0.4954

Table 3. The illustration of the jackknife (jackknife samples, each of size n-1=100-1=99) regression procedure from the data given in Figure 1, calculating the jackknife estimates of the regression parameters for each sample for fish age model

$w_i^{(J)}$	Variables	Observation sets					$\hat{\beta}_o^{(J)}$	$\hat{\beta}_1^{(J)}$	$\hat{\beta}_2^{(J)}$
		1	2	3	…	100			
1	FA(year)(Y) TL(cm) (X_1) OL(mm) (X_2)	omitted	0.92 10.00 4.10	0.08 9.30 4.10	… … …	5.08 22.10 9.10	-2.192	0.084	0.497
2	FA(year)(Y) TL(cm) (X_1) OL(mm) (X_2)	1.16 10.00 4.10	omitted	0.08 9.30 4.10	… … …	5.08 22.10 9.10	-2.176	0.084	0.497
3	FA(year)(Y) TL(cm) (X_1) OL(mm) (X_2)	1.16 10.00 4.10	0.92 10.00 4.10	omitted	… … …	5.08 22.10 9.10	-2.122	0.080	0.498
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
100	FA(year)(Y) TL(cm) (X_1) OL(mm) (X_2)	1.16 10.00 4.10	0.92 10.00 4.10	0.08 9.30 4.10	… … …	omitted	-2.141	0.083	0.493
$\hat{\beta}^{(J_o)} = \sum_{i=1}^{n} \hat{\beta}^{(J_i)} / 100$							-2.1613	0.0834	0.4957

The summaries of the some bootstrap and jackknife values of regression coefficients are presented in Table 4.

Table 4. The summary statistics of the regression coefficients for bootstrap and jackknife regression (n=100, B=10000)

	Variables	Observed	Average	S.E.	Bias	95% Confidence Interval	5%, 95% Persentile Interval
Bootstrap	Constant	-2.16133	-2.1589	0.18273	0.002457	-2.459, -1.858	-2.460, -1.850
	TL	0.08336	0.0834	0.04229	0.000069	0.0138, 0.1529	0.0137, 0.153
	OL	0.49573	0.4954	0.10250	-0.000333	0.3267 ,0.6640	0.3290, 0.663
Jackknife	Constant	-2.16133	-2.16132	0.18733	0.0007837	-2.469, -1.853	-2.19, -2.13
	TL	0.08336	0.08335	0.04326	-0.0005290	0.0122, 0.1545	0.078, 0.089
	OL	0.49573	0.49574	0.10488	0.0013688	0.3232, 0.6683	0.483, 0.506

B=10000 bootstrap samples are generated randomly to reflect the exact behavior of the bootstrap procedure and the distributions of bootstrap regression parameter estimations ($\hat{\beta}^{(b)}$) are graphed in Figure 2(a), 2(b), 2(c). The histograms of the bootstrap estimates conform quite well to the limiting normal distribution for all regression coefficients. Hence, the confidence intervals should therefore be based on that distribution, where B is sufficiently large(B=10000). And jackknife samples are generated omitting each one of the n observation sets and estimated the regression coefficients as $\hat{\beta}^{(J_i)}$. To reflect the exact behavior of the jackknife procedure and the distributions of jackknife regression parameter estimations ($\hat{\beta}^{(J_i)}$) are graphed in Figure 2(d), 2(e), 2(f). The histograms of the jackknife estimates conform quite atypical to the limiting normal distribution for all regression coefficients.

The bootstrap standard errors of the TL and OL coefficients are substantially larger than the estimated asymptotic OLS standard errors, because of the inadequacy of the bootstrap in small samples (Fox, 1997, Karlis, 2004). The confidence intervals based on the bootstrap standard errors are very similar to the percentile intervals of the TL and OL coefficients; however, the confidence intervals based on the OLS standard errors are quite different from the percentile and confidence intervals based on the bootstrap standard errors. Comparing the bootstrap coefficients averages $\overline{\hat{\beta}}_o^{(br)}$, $\overline{\hat{\beta}}_1^{(br)}$ and $\overline{\hat{\beta}}_2^{(br)}$ with the corresponding OLS estimates $\hat{\beta}_o$, $\hat{\beta}_1$ and $\hat{\beta}_2$ shows that there is a little bias in the bootstrap coefficients.

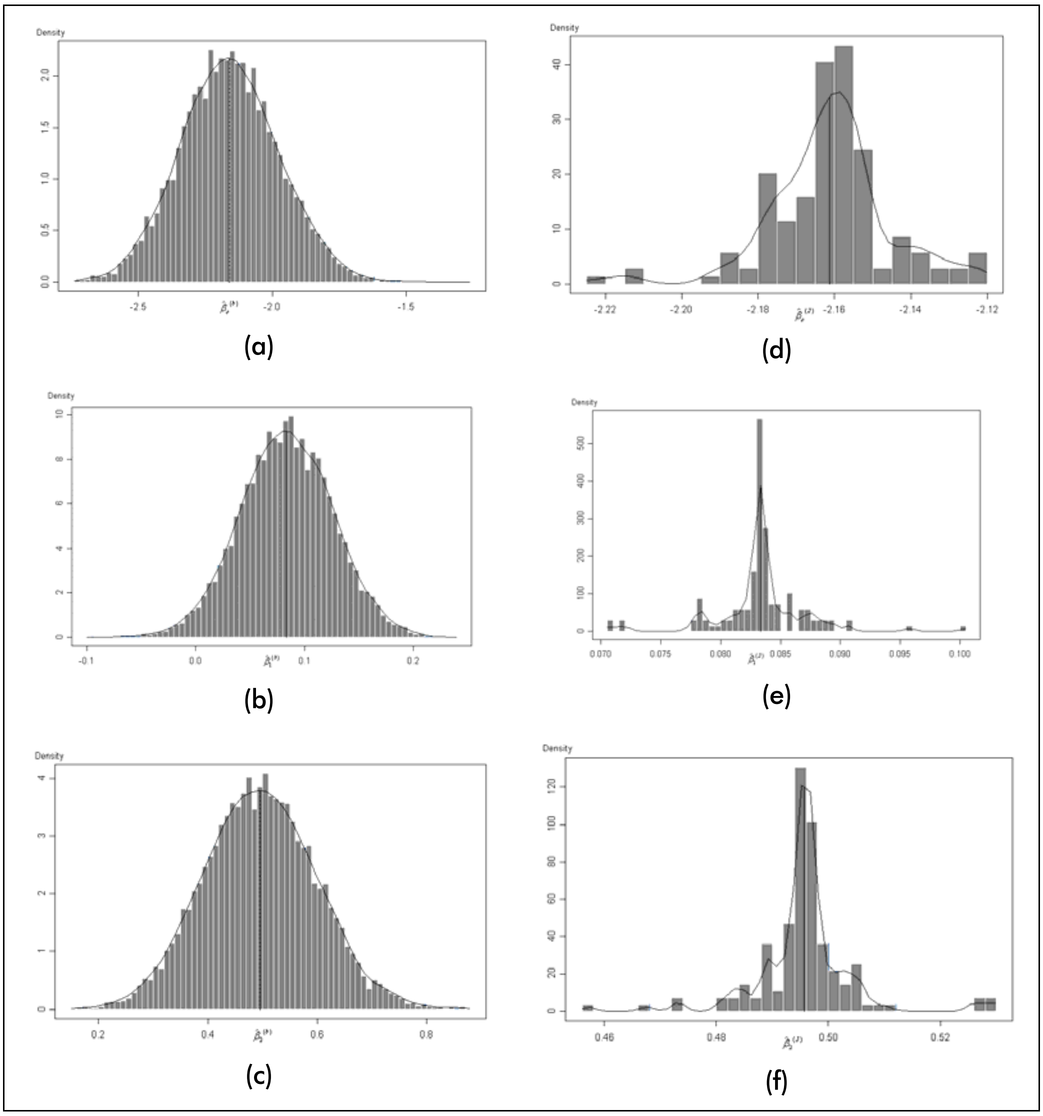


Figure 2. Histogram of bootstrap (B=10000, (a), (b), (c)) and jackknife ((d), (e), (f)) regression parameter estimates.

The shape of these graphs show that a histogram of the replicates with an overlaid smooth density estimate and the skewness of the distribution of regression parameter estimate from the bootstrap and jackknife replicate. A solid vertical live is plotted at the observed parameter value, and a dashed vertical line at the mean of the replicates

Discussion and conclusions

It is known from the statistical theory of the bootstrap that a finite total of n^n possible bootstrap samples exist. If it was computed the parameter estimates for each of these n^n samples, it would obtain the true bootstrap estimates of parameters but such extreme computation is wasteful and unnecessary (Stine, 1990). By making B large enough, it is seek to ensure that the bootstrap estimates of the regression parameter is close to the true bootstrap estimates of parameters which based on the all n^n bootstrap samples (Fox, 1997). It was suggested the bootstrap replications sufficient to be for estimating of variance $50 \le$ B

≤100, B≅1000 for estimating of standard errors, perhaps it is not enough for confidence intervals, (Leger et al, 1992; Efron, 1990, Karlis, 2004). The number of bootstrap replications B depends on the application and size of sample and computer availability. Disadvantages of bootstrap method are; i) $\hat{F}$ (bootstrap distribution of $\hat{\beta}$) is not a good approximation of F in case of small data sets and existing of outliers in the sample, ii) so bootstrap is based on the independent assumption that it is not suggested for dependence structures like time series models, iii) bootstrap based on the error procedure assumes the fitted regression model is correct and the errors are identically distributed but is preferable to the bootstrap based on the resampling of observation, for violating the assumption for constant design matrix (Karlis, 2004). In addition, the most important advantages of the bootstrap regression method are to need smaller sample than ordinary least squares method and its practical performance is frequently much better but this is not guaranteed (Hawkins and Olive, 2002). Because of this it is a mistake to hope that bootstrap regression method always gives confident results. The confidence depends on the structure of the data and distribution function.

Fan and Wang, (1995) stated that due to the fact that sample size does impose a limit on the number of resamples, the jackknife may not be appropriate for small samples, but when the sample size is large, the bootstrap and jackknife would give similar results. Heltshe and Forrester, (1985) also reported that not only sample size but also the total number of individuals in the sample is important in improving the jackknife estimators. Hence, the jackknife bias, the standard errors and confidence intervals of the TL and OL coefficients based on the distribution $F_{(\hat{\beta}^{(J)})}$ are substantially larger than the bootstrap and estimated asymptotic OLS standard errors. The jackknife percentile intervals also are larger than to the bootstrap percentile intervals of the TL and OL coefficients.

The bootstrap and jackknife methods estimate the variation of a statistic from the variation of that statistic between sub samples, rather than from parametric assumptions and may yield similar results in many situations. In addition, they provide a way of decreasing bias and obtaining standard errors in situations where the standard methods might be expected to be inappropriate. But when bootstrap is used to estimate the standard error of a statistic, it gives very little different results when repeated on the same data, whereas the jackknife gives exactly the same result each time. The bootstrap is a more general technique and preferred to the jackknife. However the jackknife is easier to apply to complex sampling schemes than the bootstrap. Application of both techniques depends on development of computer technologies and would also more frequently use if statistical computer packages featured these analyses.

As a conclusion, bootstrap method is preferable in linear regression because of some theoretical properties like having any distributional assumptions on the residuals and hence allows for inference even if the errors do not follow normal distribution.

References

1. Can, M.F., Sahinler, S., **Age Estimation of Fish Using Otolith and Fish Measurements in a Multi-species Fishery: A Case Study for Pagellus erythrinus (L., 1758) from Iskenderun Bay (NE Mediterranean Sea),** Pakistan Journal of Biological Sciences, 8(3), pp. 498-500, 2005
2. DiCiccio, T., Tibshirani, R., **Bootstrap Confıdence Intervals and Bootstrap Approximations**, J. Amer. Statist. Assoc., 82, pp. 161–169, 1987

3. Edward. J.O., **Modern Mathematical Statistics**; John Wiley and Sons inc, New York, 1988
4. Efron, B., **The Jackknife, The Bootstrap and Other Resampling Plans,** CBMS-NSF Regional Conference Series in Applied Mathematics Monograph 38, SIAM, Philadelpia, 1982
5. Efron B., More **Efficient Bootstrap Computations**, J. Amer. Statist. Assoc., 86, pp. 79–89, 1990
6. Efron, B., Gong, G., **A leisurely look at the bootstrap, the jackknife, and cross-validation,** Amer. Statist., 37, pp. 36-48, 1983
7. Efron, B., Tibshirani, R.J., **An Introduction to the Bootstrap**; Chapman & Hall, New York, 1993
8. Fan, X., Wang, L., **How comparable are the jackknife and bootstrap results: An investigation for a case of canonical correlation analysis**, The annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 387 509), 1995
9. Friedl, H., Stampfer, E., **Jackknife Resampling,** Encyclopedia of Environmetrics, 2, Eds.: A. El-Shaarawi, W. Piegorsch, Wiley: Chichester, pp. 1089-1098, 2002a
10. Friedl, H., Stampfer, E. **Resampling Methods**, Encyclopedia of Environmetrics, 3, Eds.: A. El-Shaarawi, W. Piegorsch, Wiley:Chichester, pp. 1768-1770, 2002b
11. Fox, J., **Applied Regression Analysis, Linear Models and Related Methods**; Sage, 1997
12. Hawkins D.M., Olive D.J., **Inconsistency of resampling algorithms for high-breakdown regression estimators and a new algoritm**, J. Amer. Statist, Assoc., 97, pp. 136-159, 2002
13. Heltshe, J.F., Forrester, N.E., **Statistical Evaluation of the Jackknife Estimate of Diversity when Using Quadrat Samples**, Ecology, 66, pp. 107-111, 1985
14. Karlis, D., **An introduction to Bootstrap Methods**, 17th conference of Greek Statistical Society, Greece, 2004
15. Leger, C., Politis, D. N., Romano, J.P., **Bootstrap Technology and Applications**, Technometrics, 34, pp. 378-397, 1992
16. Liu, Y. R., **Bootstrap Procedures under Some Non-i.i.d. models**, Ann. of Stat., 16, pp. 1696-1708, 1988
17. Miller, R.G., **The jackknife- a review**, Biometrika, 61, pp. 1-15. 1974
18. Pedhazur, E. J., **Multiple Regression in Behavioral Research;** 3rd Edition, Orlando, FL:Harcourt Brace, 1997
19. Sahinler, S., **The Basic Principles of Fitting Linear Regression Model By Least Squares Method,** Journal of Agricultural Faculty, Mustafa Kemal University, 5(19), pp. 57-73, 2000
20. Sahinler, S., Bek, Y., **A Comparison of Some Statistics On The Determination of Collinearity In Animal Science Data**, Journal of Animal and Veterinary Advances, 1(3), pp. 116-119, 2002
21. Sahinler, S., Bek, Y., **A Classification of Single Influential Observation Statistics In Regression Analysis**, Journal of Applied Science, Kırgızistan-Türkiye Manas University, 7, pp. 1-18, 2006
22. Shao, J., **Bootstrap Model Selection**, J. Amer. Statist. Assoc, 91, pp. 655-665, 1996
23. Shao, J., Tu, D., **The Jackknife and Bootstrap;** Springer, New York, 1995
24. Stine, R., **Bootstrap prediction intervals for regression**, J. Amer. Statist. Assoc, 80, pp. 1026-1031, 1985
25. Stine, R., **Modern Methods of Data Analysis**; Edit: by John Fox, pp. 325-373, Scotland Sage Pub., 1990
26. Wu, C.F.J., **Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis**, Annals of Statistics, 14, pp. 1343-1350, 1986

[1] (Corresponding Author)