http://www.pjbs.org

PJBS

ISSN 1028-8880

Pakistan Journal of Biological Sciences

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Pakistan Journal of Biological Sciences 9 (15): 2919-2921, 2006
ISSN 1028-8880

A Comparative Study of Least Squares Method and Least Absolute Deviation Method for Parameters Estimation in Length-weight Relationship for White Grouper (*Ephinephelus aeneus*)

[1]Levent Sangun, [2]Soner Cankaya, [1]Gökhan Gokce and [3]Hülya Atil
[1]Faculty of Fisheries, Cukurova University, Adana, Türkiye
[2]Department of Animal Science, Ondokuz Mayis University, Samsun, Türkiye
[3]Department of Animal Science, Faculty of Agriculture, Ege University, Izmir, Türkiye

Abstract: In fisheries science, by means of linear regression models, length-weight relationships are often estimated to determine weight and biomass when only length measurements are taken from fish. Least Squares method (LS) is commonly used to determinate the relationship between weight and length for fish, when the error term, e_i, is assumed to be normally distributed. If it can be observed to degenerate the structure of data in the y-direction, LS method is not completely perform to estimate the regression parameters. And than LS method explains minimum levels to the total variation of the model. In this situation, one of the linear regression methods often recommended as robust or outlier-resistant alternatives to LS is Least Absolute Deviation (LAD). The aim of this study is to investigate for comparing the least squares method and the LAD method by means of drawing conclusions and to determine the model which is optimal for displaying the relationship between length and weight for *Ephinephelus aeneus* in the presence of outliers. Mean square error and R^2 are used to evaluate estimator performance.

Key words: Least squares method, least absolute deviation method, *Ephinephelus aeneus*

INTRODUCTION

The Serranidae is one of the most economically important families, containing species such as *Epinephelus aeneus* (white grouper), *Epinephelus guaza*, *Epinephelus alencentrinus* and *Epinephelus marginatus* in commercial fisheries in southern part of Turkey. Due to high economical value, importance of the commercial and recreational fishing of the grouper is crucial. However, it is considered as endangered species and in red list of International Union for Conservation of Nature and Natural Resources (IUCN) since 1996. Only a few studies were found in Iskenderun Bay such as weight-length relationship (Can *et al*., 2002) and gonad histology and spawning pattern (Gokce *et al*., 2002).

Length-weight relationships allow the shift of growth-in-length equations to growth-in-weight for use in stock assessment models, the estimation of biomass from length observations, an estimate of the condition of the fish and this relationship is useful for between region comparisons of life histories of certain species (Goncalves *et al*., 1997; Froese and Pauly, 1998; Moutopoulos and Stergiou, 2002). Relationships between different types of lengths, for which little information seems to be available for Mediterranean species, are also very important for comparative growth studies (Froese and Pauly, 1998).

In the field of fisheries science, much papers has been devoted to study of length-weight relationship with linear regression method (e.g., Chen and Jackson, 2000; Harrison, 2001; Morey *et al*., 2003; Sinovčić *et al*., 2004). Linear regression analysis consists of a collection of techniques used to explore relationships between an independent variable X (total length) and a dependent variable Y (weight) (Weisberg, 1980). Least Squares Regression methods (LSR) generally are used to determine the relationship among the variables. Thus, one of the objectives of the experiment is to determine whether the Y values are related to the X values. There are some problems in fisheries and ecological studies as heterogeneity of error variance, non-normality of error and existence of outliers and etc. because of large measurement and process error associated with the studies (Chen and Jackson, 2000). Therefore, to overcome these problems, the researchers have recently investigated the alternative regression method which is Least Absolute Deviations method (LAD). LAD method was developed by Roger Joseph Boscovich in 1757 (Birkes and Dodge, 1993) to reconcile inconsistent measurement for the purpose of estimating the weight for *E. aeneus*. The idea of LAD regression is actually older than that of LSR, nearly 50 years, but until the development of high-speed computers, it is then too

Corresponding Author: Levent Sangun, Faculty of Fisheries, Cukurova University, Adana, Türkiye

cumbersome to have wide applicability. Its application was discussed by Ellis (1998), Allen and DeGaetano (2001), Zhao (2001) and Caner (2002).

However, computing LAD estimates is more difficult than computing LS estimates due to the fact that algorithmic method must be employed to calculate LAD estimates. Therefore, LAD estimation method is not popular as LS estimation method.

The aim of this study is to introduce LAD method for parameter under the simple linear regression model in situation that the underlying assumptions of the least squares estimation are untenable. Mean square error and R^2 are used to evaluate estimator performance for the linear regression models.

MATERIALS AND METHODS

Samples were collected between September 2002 and August 2004 at Iskenderun Bay in east of Mediterranean Sea on board the commercial bottom trawlers. A total of 52 hauls was made and all samples were measured (total length) and weighted nearest 0.1 mm and 0.1 g, respectively.

LSR estimation method performs well if the distribution of errors is thought to have a normal distribution. But, it is not recommend that LS method is used to estimate the regression parameters, due to the fact that violation of normality assumption is often manifested by the presence of outliers in the observed data. For the best fit of regression line we compare LSR and LAD methods by using raw data (for LSR and LAD) and logarithmic transformed data (for LSR) of the length-weight relationship. In the research model, total length (TL) is independent variable (X); body weight (BW) for the fish is dependent variable (*Y*).

LSR method

$$yij = \beta_0+\beta_1X+eij,\ eij \sim N(0, \sigma^2)$$

Where, y is the vector of responses, X is the design matrix and b is a vector of unknown parameters estimated by minimizing the sum of squares of e, the error or residual.

$$\sum_{i=1}^{n} e_i^2$$

Where,

$$e_i = y_i - \hat{y}_i$$

Least absolute deviation regression finds the estimate â that minimizes the sum of the absolute values of the residuals;

$$\sum_{i=1}^{n} |e_i|$$

The concept of LAD estimation is not more difficult than the concept of LS estimation; in fact, it is simpler in so far as absolute-value of e_i is a more straightforward measure of the size of a residual than e_i^2 (Birkes and Dodge, 1993).

LAD regression estimates are obtainable from the function l1fit in the computer language S-PLUS and from the robust regression package ROBSYS (Birkes and Dodge, 1993).

RESULTS AND DISCUSSION

Table 1 showed that the coefficient of variations (CV) for total length and body weight are high, 52 and 161.66%, respectively. Therefore, the data has non-normal distribution and outlier data-case numbers are 4, 8, 11, 14, 16 and 20, which has been determined using normality test of Kolmogorow-Smirnov and Box-plot. By means of transformation, although the CV for TL is turned into normal, the CV for BW is high more than 30% which is biological limit for credibility of the measured data. According to this result, it can be said that the outlier exist in the body weight-direction.

Table 1: Descriptive statistics for original and log-transformed data belonging to TL and BW

	Original data		Transformed data	
Variables	BW	TL	ln BW	ln TL
Mean	997.74	21.45	5.36	2.93
Standard deviation	1612.96	11.19	2.08	0.53
CV (%)	161.66	52.17	38.81	18.09

BW, Body weight; TL, Total Length; CV, Confidence Intervals

Table 2: Test of the Significant of the Mean Squares for the Models in Linear Regression

Source	DF	Mean Square (MS)	F	R^2
Least squares model for original data				
Regression	1	42978866.559	23.956**	0.324
Residual	50	1794091.795		
Least squares model for transformed data				
Regression	1	104.918	44.988**	0.474
Residual	50	2.332		
Least absolute deviation modelsource				
Regression	1	221.4802	203.025**	0.754
Residual	50	1.0909		

**: p<0.01

Table 3: Test of the significant of the coefficients for the models in linear regression

	Variables	The coef. of b_i	SE	t-value
LS	Constant	-760.868	404.478	-1.881
	TL	82.001	16.754	4.894**
LS[1]	Constant	-2.551	1.198	-2.129*
	TL	2.700	0.403	6.707**
LAD	Constant	-4.319	0.169	25.556**
	TL	2.974	0.24	12.392**

[1]Least square model for transformed data; *: p<0.05; **: p<0.01

Table 2 shows that least absolute deviation method was affected less than least squares methods, which were performed the original and transformed data, against the outliers. Moreover, total variation of the model explained by LAD was more than explained by LS.

Table 3 shows that the standard errors of regression coefficients for least squares method were high enough to compare with the errors for least absolute deviation method because of the presence of outliers. When the original data have been taken into consideration, while the constant for LS method is non-significant ($p>0.05$), the value for LAD method is significant at 1% level. Hence, it is recommended to researchers that Lad method is performed to determine the relationship of length-weight for white grouper under the simple linear regression model.

In conclusion, the usage of LAD analysis on length-weight relationship provide more realistic results when the lack of data of endangered species such as white grouper. It was found that the LAD estimator was more efficient than the LS estimator because of R^2 and the significant tests of their coefficients. Therefore, LAD method may reduce the danger from outliers in simple linear regression model.

REFERENCES

Allen, R.J. and A.T. De Gaetano, 2001. Estimating missing dally temperature extremes using an optimized regression approach. Intl. J. Climatol, 21: 1305-1319.

Birkes, D. and Y. Dodge, 1993. Alternative Methods of Regression. John Willey and Sons, Inc., New York, pp: 228.

Can, M.F., N. Başusta and M. Ve Çekiç,, 2002. Weight-lenght relationships for selected fish species of the small-scale fisheries of the south coast of the Iskenderun Bay. Turk. J. Vet. Animal Sci., 26: 1181-1183.

Caner, M., 2002. A note on least absolute derivation estimation of a threshold model. Econ. Theory, 18: 800-814.

Chen, Y. and D.A. Jackson, 2000. An empirical study on estimation for linear regression analyses in fisheries and ecology. Fish. Res., 49:193-206.

Ellis, S.P., 1998. Instability of least squares, least absolute derivation and least median of squares linear regression. Stat. Sci., 13: 337-350.

Froese, R. and D. Pauly, 1998. FishBase 1998: Concepts, design and data sources. Manila, ICLARM, pp: 293.

Gokce, M.A., I. Cengizler and A.A. Ozak, 2002. Gonad histology and spawning pattern of the white grouper (*Ephinephelus aeneus*) from Iskenderun Bay (Turkey). Turk. J. Vet. Anim. Sci., 27: 957-964.

Goncalves, J.M.S., L. Bentes, P.G. Lino, J. Ribeiro, A.V.M. Canario and K. Erzini, 1997. Weight-length relationships for selected fish species of the small-scale demersal fisheries of the South and South-West Coast of Portugal. Fish. Res., 30: 253-256.

Harrison, T.D., 2001. Lenght-weight relationships of fishes from South African estuaries. J. Applied Ichthyol., 17: 46-48.

Morey, G., J. Moranta, E. Massuti, A. Grau, M. Linde, F. Riera and B. Morales-Nin, 2003. Weight-length relationships of littoral to lower slope fishes from the Western Mediterranean. Fish. Res., 62: 89-96.

Moutopoulos, D.K. and K.I. Stergiou, 2002. Weight-length and length-length relationships for 40 fish species of the Aegean Sea (Hellas). J. Applied Ichthyol., 16: 200-203.

Sinovčić, G., Franièeviæ, M., Zorica, B and V. Čikeš-Keč, 2004. Length-weight and length-length relationships for 10 pelagic fish species from the Adriatic Sea (Croatia). J. Applied Ichthyol., 20: 156-158.

Weisberg, S., 1980. Applied Linear Regression. John Wiley and Sons, New York, pp: 283.

Zhao, Q., 2001. Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form. Econ. Theory, 17: 765-784.