Journal of Applied Biological Sciences 7 (3): 46-50, 2013
ISSN: 1307-1130, E-ISSN: 2146-0108, www.nobel.gen.tr

A comparison of weighted least square estimation and ordinary least square estimation for analysing weight-length relationship of Por's Goatfish *(Upeneus pori* Ben-Tuvia & Golani, 1989) in Iskenderun bay

Mustafa AKAR[1] Sedat GÜNDOĞDU[1*]
[1]Fisheries Faculty, Cukurova University, 01330, Adana/ Turkey

[*]Corresponding Author
E-mail: sgundogdu@cu.edu.tr

Received: October 30, 2013
Accepted: December 25, 2013

Abstract

A weighted least square (WLS) and ordinary least square (OLS) approach for analysing fish length-weight relationships, was applied. We reviewed and compared two regression methods used in statistics to reconstruct length-weight relationship from Por's Goatfish (*Upeneus pori* Ben-Tuvia & Golani, 1989) length-weight data in Iskenderun Bay. This comparison relies on a suite of regression model verification statistics to validate the accuracy of the length-weight relationship produced by the best model. The results indicate WLS may perform slightly better in heteroscedastic data; conversely, OLS may perform slightly better when the data is homoscedastic.

Key-Words: Weighted Least Square, Ordinary Least Square, *Upeneus pori,* Population Dynamics

INTRODUCTION

When to confirming quantitative analysis methods, it is basic to know a function of the response, i.e., the relationship between response (y) and the independent variable (x) in the sample within an application range. This relationship is obtained by means of a simple linear regression, whose parameters (slope and intercept) are estimated through the least squares. In fisheries sciences, linear regression is used to perform the length-weight method, generally using the length as the independent variable. Commonly used linear regression method is the least squares (LS). There are two types of LS: OLS (ordinary least squares-data having constant uncertainty) and WLS (weighted least squares-data having varying uncertainty). OLS, can only be used if the following conditions are supplied: (i) relationship between y and x variables is linear; (ii) x is without error or less than one-tenth of the error in y; (iii) errors in y are normally distributed; (iv) error in y is homoscedastic (constant variance across the entire response range); and (v) errors associated with different observations are independent (Mosteller and Tukey, 1977; Myers, 1990; Draper and Smith, 1998; Rawlings et al., 1998; Rao and Toutenburg, 1999; Freund et al., 2006; Nascimento et al., 2010).

In fisheries science, weight and length data may exhibit a linear pattern yet have non constant variance. In these cases, variance assumption of OLS is violated and estimator of β is biased (Draper and Smith, 1998; Rawlings et al., 1998; Rao and Toutenburg, 1999; Freund et al., 2006). The most common approach adopted to solve this disagreement is variable transformation such as logarithmic transformation. After transformation of the variables, OLS can be used for assessment of parameter estimation of length-weight relationships. In case of heterogeneity of residual variance (heteroscedasticity), even though the data is transformed, the estimators lose their best linear unbiased estimation (BLUE) feature and OLS can no longer be applied (Knight, 2000; Helsel and Hirsch, 2002; Cruz and Branco, 2009; Ramachandran and Tskos, 2009). In this situation, the restrictions of OLS can be overcome by means of a weighting procedure (Myers, 1990; Draper and Smith, 1998; Rawlings et al., 1998; Rao and Toutenburg, 1999; Freund et al., 2006).

In fisheries science, growth of fish is included in population dynamics models that are used to derive sustainable harvest levels for managing stocks (Ricker, 1975). In these cases, choosing a correct statistical analysis becomes more important. In this study, we compared the use of OLS and WLS to fit the length-weight relationship line and estimated regression parameters in best fit.

MATERIAL AND METHODS

Methodology of the Weighted Least Squares Procedure

The following methodology is an adaptation for studying the condition the weighted least squares procedure was developed from; Moser (1996), Draper and Smith (1998), Davidson and MacKinnon (1999), O'Neill and Mathews (2000), Kariya and Kurata (2004), Kim and Timm (2007) and Myers et al. (2010).

The model under consideration is;

$$Y = X\beta + \varepsilon,$$

$E(\varepsilon) = 0,\ Var(\varepsilon) = W\sigma^2$ and $\varepsilon \sim N(0, W\sigma^2)$

where Y is an $n \times 1$ dependent variables vector, k is the number of independent variables, X is an $n \times (k+1)$ constant matrix, β is a $(k+1) \times 1$ parameter vector, ε is an $n \times 1$ error vector, W is an $n \times n$ diagonal matrix. It sometimes happens that some of the observations used in a regression analysis are less credible than others. What this usually means is that the variances of the observations are not all equal; in other words the non-singular matrix Var (ε) is not of the form $I\sigma^2$ but is diagonal with unequal diagonal elements. It may also happen, that the off-diagonal elements of Var (ε) are not zero, that is, the observations are correlated. When either or both of these events occur, the OLS estimation formula $\hat{\beta} = (X'X)^{-1}X'Y$ does not apply and it is necessary to apply the WLS.

The sequences of the essential steps employed in this study are:

(1) Test for heteroscedasticity. Heteroscedasticity arises in many applications. White (1980), Draper and Smith (1998), Green (2003) and Gujarati (2004) referred to graphic detection method and many other tests (White's test or with any other heteroscedasticity test).

(2) To form groups, each with a common observation (fish weight) and to calculate the variance of the observations in each group. These variances should be plotted against length for each group.

(3) To find the best fit from step (2) and calculate weights as shown as Table 1.

(4) Use statistical computer software (SPSS, Minitab, SAS etc.) for running WLS. We used SPSS v19 for running WLS.

Data Used in the Case Study

Data analyzed in this study were collected with trawl nets during March-April 2012 from Iskenderun Bay in Mediterranean Sea. The fish were measured for total length (to the nearest 0.1 cm) and weighed (to the nearest 0.01 g). Total weight and lengths were by transformed natural logarithm and then the weight-length relationship of *U. pori* was examined by WLS and OLS.

RESULTS

A total of 289 individuals were collected, ranging in size from 7.2-11.50 cm fork length. The length-frequency distribution is given in Figure. As it can be shown in the Figure, the 9-10 cm length group was the most common one. Total weight of the sampled individuals ranged from 5.55-23.51 g. Overall mean ± standard deviation total length and weight were calculated as 9.35 ± 0.77 cm and 13.47 ± 3.28 g, respectively. Before detection of the violation of homoscedasticity, length and weight for both of genders were compared by using t-test and the result of test was found as nonsignificant (P> 0.05). As a result of the test genders are not significantly different for length and weight. Therefore, we combined data and statistically analyzed.

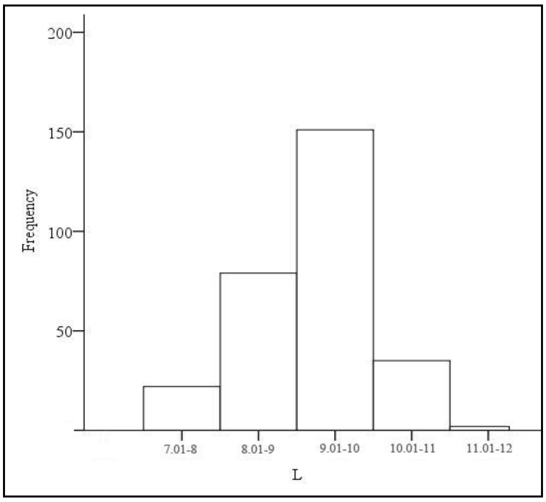

Figure Length frequency distribution of *Upeneus pori*

Table 1. Formulas for Estimation of Weights

Model	Linear	Logarithmic	Quadratic	Exponential	Growth
Weights(W_i)	$\frac{1}{\hat{\beta}_0 + \hat{\beta}_1 X}$	$\frac{1}{\hat{\beta}_0 X_i^{\hat{\beta}_1}}$	$\frac{1}{\hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2}$	$\frac{1}{\hat{\beta}_0 (e^{\hat{\beta}_1 X_i})}$	$\frac{1}{(e^{\hat{\beta}_0 + \hat{\beta}_1 X_i})}$

Data were analysed for heteroscedasticity by White's General Heteroscedasticity Test. The estimated value for the quadratic model was 0.881, which was higher than the value estimated using the White Test statistics (Table 2) showing us the data is heteroscedastic (P< 0.01). This key result of this study shows that the main assumption of OLS is violated and therefore, WLS can be applied (Halbrendet et al., 1992; Gispert and Amich, 2001). When we considered the determination constant (R^2) and the significance level of the model for choosing the best model using the fish weight groups versus their group variances, growth and exponential models were shown to have the highest R^2 (0.238) and significance levels, and were shown to be usable for weight calculation (P< 0.01, Table 3). Weights, calculated from both models were the same.

Table 2. White Test Statistic Value of Models

Source	R^2	df1	df2	χ^2_{k-1}
Linear	.880**	1	287	254.32
Logarithmic	.880**	1	287	254.32
Quadratic	.881**	2	286	254.61

** P< 0.01

All cases presented in Table 4 show results of OLS and WLS. The conclusions from the two methods coincide in all cases, and show that the standard errors and 95% confidence intervals of parameters, to be estimated with WLS, are smaller than OLS (Table 4).

DISCUSSION

The choice of least square methods is very important for estimation of fish growth parameters. In fisheries literature, OLS is the most common analysis. Most scientists are applying transformations to use the OLS method. In contrast with common applications, in this study we employed WLS analysis for estimating regression parameters instead of OLS method. Results have shown that OLS is not suitable for estimating the regression parameters of the weight-length relationship of this species, and that this depends on the rejection of the main assumption of OLS. If the main assumption of OLS (homoscedasticity) is violated, estimators lose their BLUE attribution, and this affects the values of the regression parameters ($\hat{\beta}_0, \hat{\beta}_1$ ve $\hat{Y}$) (Helsel and Hirsch, 1992). Consequently, estimation of growth parameters ($L_\infty, t_0, W_\infty, K$) and fisheries management rules may be affected (Sparre et al., 1999).

If we compare the parameter estimates of OLS and WLS (Table 4), they show us that the confidence intervals and standard errors of estimates from WLS are smaller than OLS estimates. This suggests that the choice to use WLS analyses is the correct course of action for this species. According to Draper and Smith (1998), Shields (1978), and Rawlings et al. (1998), the main assumption of OLS requires that the error variances of the Y (fish weight) groups be constant across different length values (homoscedasticity). In these cases, violation of this main assumption makes OLS insufficient.

Table 3. Determination Constant, Significance Level and Parameter Estimates of Fish Length Groups Versus Fish Weight

Source	R^2	df1*	df2**	p	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
Linear	0.106	1	27	0.084	- 8.327	1.146	
Logarithmic	0.104	1	27	0.087	- 20.866	10.434	
Quadratic	0.106	2	26	0.232	- 5.369	0.501	0.035
Growth	0.238	1	27	0.007	- 4.812	0.538	
Exponential	0.238	1	27	0.007	0.008	0.538	

*degree of freedom of model **degree of freedom of groups

Table 4. Comparison of Parameter Estimates with OLS and WLS

	WLS				OLS			
			95% CI				95% CI	
Parameters	Estimates	SE	Lower	Upper	Estimates	SE	Lower	Upper
$\hat{\beta}_0$	- 3.952*	.130	- 4.209	- 3.695	- 3.864*	.140	- 4.140	- 3.588
$\hat{\beta}_1$	2.923	.059	2.806	3.039	2.883	.063	2.759	3.006
R^2	.894	.085			.880	.088		

* This isn't retransformed. CI: Confidence Interval SE: Standard Error

In literature, researchers have calculated different values of parameters for the weight-length relationship via OLS. For instance, for this species, Taskavak and Bilecenoglu (2001) presented $\hat{\beta}_0$ and $\hat{\beta}_1$ are - 12.786 and 3.256 (n=102), Çiçek and Avsar (2011) presented - 4.537 and 2.9487 (n=274), and Erguden et al. (2009) presented - 4.154 and 2.816 (n=210). These differences are due to differences in the data collection periods and methods between these three studies, and our study as well. According to Bagenal and Tesch (1978), Sparre et al. (1989), Dulcic and Kraljevic (1996) and Gonçalves et al. (1997) $\hat{\beta}_1$ generally does not vary significantly throughout the year, unlike parameter $\hat{\beta}_0$, which may vary seasonally, daily, and between habitats. They also didn't use any test for OLS assumptions in these studies, so we are unable to compare the results of these studies to our study.

The estimation that is done with WLS for $\hat{\beta}_1$ (2.923) is found to be better than the estimation with OLS (2.883) because the growth of *Upeneus pori* is isometric (Çiçek and Avşar, 2011). According to Ricker (1975), the $\hat{\beta}_1$ value of fish species that grow isometrically should be around three.

Estimation parameters of the relationship between weight and length of a fish species in a given geographical area are an easy way to determine growth characteristics of fish species. In these cases the choice of parameter estimation methods is important. Heteroscedastic data, as in this study, requires checking via some tests. This study has shown that prior to performing OLS it is necessary to check whether the variance in length data is constant or not. After that the researcher can decide whether or not OLS is suitable method for the data set.

Acknowledgement
Data for this study was collected via a scientific grant supported by the Scientific Project Department of Çukurova University (SÜF2011BAP7).

REFERENCES

[1] Albayrak AS (2008). Weighted regression analysis alternative to the least squares technique in the presence of none constant variance and an application. Afyon Kocatepe Üniversitesi İİBF Dergisi 10(2): 111-134 (article in Turkish with an abstract in English).

[2] Bagenal TB, Tesch FW (1978). Age and growth. in: methods for assessment of fish production in fresh waters. 3rd ed. T. Bagenal (Ed.). IBP Handbook No. 3. Oxford: Blackwell Science Publications, pp 101-136.

[3] Cicek E and Avsar D (2011). Age, Growth and Mortality of *Upeneus pori* (Ben-Tuvia and Golani, 1989) off the Karatas Coasts of Iskenderun Bay. J Anim Vet Adv. 10(7): 878-882.

[4] Cruz RD, Branco MD (2009). Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves. Biometrical Journal 51(4): 588-609.

[5] Davidson R, MacKinnon JG (1999). Econometric Theory and Methods. London, UK: Oxford University Press.

[6] Draper N, Smith H (1998). Applied Regression Analysis. 3rd ed. New York, USA: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

[7] Dulcic J, Kraljevic M (1996). Weight - length relationships for 40 fish species in the eastern Adriatic (Croatian Waters). Fish Res 28: 243-251.

[8] Dytham C (2000). Choosing and Using Statistics. Oxford, UK: Blackwell Science.

[9] Erguden D, Turan C and Gurlek M (2009). Weight–length relationships for 20 lessepsian fish species caught by bottom trawl on the coast of Iskenderun Bay (NE Mediterranean Sea, Turkey). J Appl Ichthyol 25: 133-135.

[10] Freund RJ, Wilson WJ and Sa P (2006). Regression Analysis. 2rd ed. USA: Academic Press, Elsevier.

[11] Gisperts AV, Amich RM (2001). Fish condition analysis by a weighted least squares procedure: Testing geographical differences of an endangered Iberian Cypriodontid. J Fish Bio 58: 1658-1666.

[12] Gonçalves JMS, Bentes L, Lino PG, Ribeiro J, Canario AVM and Erzini K (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.

[13] Gujarati DN (2004). Basic Econometrics. 4rd ed. New Jersey, USA:McGraw-Hill.

[14] Halbrendt C, Bacon JR (1992). Weighted least squares analysis for conjoint studies: The case of hybrid striped bass. Agribusiness 8(2): 188-198.

[15] Helsel DR, Hirsch RM (1992). Statistical Methods in Water Resources. New York, USA: Elsevier.

[16] Kariya T, Kurata H (2004). Generalized Least Squares. UK: John Wiley & Sons, Ltd.

[17] Kim K, Timm NH (2007). Univariate and Multivariate General Linear Models: Theory and Applications with SAS. Boca Raton , FL: Chapman & Hall/CRC.

[18] Knight K (2000). Mathematical Statistics. USA: Chapman & Hall/CRC.

[19] Moser WR (1996). Linear Models: A Mean Model Approach. USA: Academical Press Inc.

[20] Mosteller F, Tukey JW (1977). Data Analysis and Regression. A Second Course in Statistics. Addison-Wesley Series in Behavioral Science.

[21] Nascimentoa RS, Froes RES, Silva NOC Naveira RLP, Mendes DBC, Neto WB and Silva JBB (2010). Comparison between ordinary least squares regression and weighted least squares regression in the calibration of metals present in human milk determined by ICP-OES. Talanta 80: 1102–1109.

[22] O'Neill ME, Mathews KY (2000). A weighted least squares approach to Levene's test of homogenity of variance. Aus N Z J Stat 42(1): 81-100.

[23] Ramachandran KM, Tsokos CP (2009). Mathematical Statistics with Applications. USA: Elsevier.

[24] Rao CR, Toutenburg H (1999). Linear Models: Least Squares and Alternatives. New York, USA: Springer Series in Statistics.

[25] Rawlings DJ, Pantula GS Dickey DA (1998). Applied Regression Analysis A Research Tool, 2rd ed. North Carolina USA: Spinger Text in Statistics.

[26] Ricker WE (1975). Computation and interpretation of biological statistics of fish populations. Bull Fish Res Board Can 191: 1-382.

[27] Shields JL (1978). An empirical investigation of the effect of heteroscedasticity and heterogeneity of variance on the analysis of covariance and the Johnson–Neyman technique. Alexandria, Virginia, USA: Army Research Institute for the Behavioral and Social Sciences Technical Paper.

[28] Sparre P, Ursin E and Venema SC (1989). Introduction to Tropical Fish Stock Assessment, Part-I Manual. Rome, Italy: FAO Fisheries and Technical Paper.

[29] Taskavak E, Bilecenoglu M (2001). Length–weight relationships for 18 lessepsian (Red Sea) immigrant fish species from the eastern mediterranean coasts of Turkey. J Mar Biol Assoc UK 81: 895– 896.

[30] White H (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48: 817–838.