PACIFIC ISLANDS FISHERIES SCIENCE CENTER

Fitting Length-Weight Relationships with Linear Regression Using the Log-Transformed Allometric Model with Bias-Correction

Jon Brodziak

August 2012

Administrative Report H-12-03

About this report

Administrative Reports may be cited as follows:

Brodziak, J.
2012. Fitting length-weight relationships with linear regression using the log-transformed allometric model with bias-correction. Pacific Islands Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, Honolulu, HI 96822-2396. Pacific Islands Fish. Sci. Cent. Admin. Rep. H-12-03, 4 p.

For further information direct inquiries to

Chief, Scientific Information Services
Pacific Islands Fisheries Science Center
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
U.S. Department of Commerce
2570 Dole Street
Honolulu, Hawaii 96822-2396

Phone: 808-983-5386
Fax: 808-983-2902

Pacific Islands Fisheries Science Center
Administrative Report H-12-03

Fitting Length-Weight Relationships with Linear Regression Using the Log-Transformed Allometric Model with Bias-Correction

Jon Brodziak

Pacific Islands Fisheries Science Center
Fisheries Research and Monitoring Division
National Marine Fisheries Service, NOAA
2570 Dole Street, Honolulu, Hawaii 96822-2396

August 2012

INTRODUCTION

The purpose of this report is to provide the information needed to calculate unbiased estimates of the parameters of a length-weight relationship for a given sample of length-weight data from a fish species using the method of maximum likelihood.

MATERIALS AND METHODS

Here is the standard allometric equation to predict fish weight (*W*) at length (*L*).

(1.1) $$W = A \cdot L^B$$

In equation (1.1), the parameters *A* and *B* are to be estimated with the available length-weight data. The parameter *A* is a scaling coefficient for the weight at length of the fish species. The parameter B is a shape parameter for the body form of the fish species. In theory, one might expect that the exponent B would have a value of roughly $B = 3$ because the volume of a 3-dimensional object is roughly proportional to the cube of length for a regularly shaped solid. For example, the volume (*V*) of a square box with sides of length *L* is $V = L^3$. In practice, fish that have thin elongated bodies will tend to have values of B that are less than 3 while fish that have thicker bodies will tend to have values of B that are greater than 3.

It is assumed that the length-weight data for the fish species consist of a total of *n* length and weight measurements from individual fish. That is, the length-weight data set (*D*) consists of the weight-length measurements $D = \{(W_1, L_1), (W_2, L_2), ..., (W_n, L_n)\}$ where W_k is the weight of the k^{th} fish and L_k is the weight of the k^{th} fish.

Note that equation (1.1) is nonlinear and there is no direct solution for the parameters *A* and *B* that produced an observed data set *D*. However, if one transforms the allometric equation by applying the natural logarithm to both sides of equation (1.1), then a linear regression equation to predict the logarithm of weight as a function of the logarithm of length and the transformed parameters can be derived:

(1.2) $$\log W = \log A + B \cdot \log L \equiv b0 + b1 \cdot \log L + \varepsilon$$

The log-transformed equation (1.2) is a linear regression model with an intercept parameter *b*0 and slope parameter *b*1 along with a normally distributed error term ε that has an expected value of zero and a constant variance. In particular, note that the transformed parameters are $b0 = \log A$ and $b1 = B$. The linear regression model in equation (1.2) can be fit to the observed length-weight data using the method of maximum likelihood to obtain maximum likelihood estimates (MLEs) of the parameters *b0* and *b*1, where each data point is fit with a residual error ε that represents the difference between the observed weight value and the predicted weight using the estimated regression parameters.

The definition of the residual error for the kth fish (ε_k), which is equal to the logarithm of the observed fish weight minus the predicted logarithm of weight of the k^{th} fish, is

(1.3) $$\varepsilon_k \; = \; \log W_k \; - \; \left(b0 \; + \; b1 \cdot \log L_k\right)$$

If the linear regression model is fit to the length-weight data using the method of maximum likelihood by solving the normal equations (see, for example, Larsen and Marx, 1981), analytical estimates of the parameters $b0$ and $b1$ can be derived. In particular, if the errors in the predicted log-transformed weight from the linear model are normally distributed with constant variance, i.e.,

(1.4) $$\log W_k \; \sim \; N\left(b0 \; + \; b1 \cdot \log L_k \, , \, \sigma^2 \right)$$

then the maximum likelihood estimates of $b1$, $b0$, and σ^2 have exact solutions.

To express the exact solutions for the MLEs succinctly, denote the expected values of the log-transformed observed fish weights ($E[\log W]$) and lengths ($E[\log L]$) as

(1.5) $$E\left[\log W\right] \; = \; \frac{1}{n}\sum_{k=1}^{n}\log W_k \quad and \quad E\left[\log L\right] \; = \; \frac{1}{n}\sum_{k=1}^{n}\log L_k$$

Given these definitions, the maximum likelihood estimate of $b1$ is

(1.6) $$\widehat{b1} \; = \; \frac{\sum_{k=1}^{n}\left(\log L_k - E\left[\log L\right]\right)\cdot\left(\log W_k - E\left[\log W\right]\right)}{\sum_{j=1}^{n}\left(\log L_j - E\left[\log L\right]\right)^2}$$

and the maximum likelihood estimate of $b0$ is

(1.7) $$\widehat{b0} \; = \; E\left[\log W\right] - \widehat{b1}\cdot E\left[\log L\right]$$

and the bias-corrected maximum likelihood estimate of σ^2 is

(1.8) $$\widehat{\sigma^2} \; = \; \frac{1}{n-2}\sum_{k=1}^{n}\varepsilon_k^2$$

Maximum likelihood estimates of the variances of the parameters $b1$ and $b0$ can also be derived and are functions of σ^2. The variance of the slope parameter $b1$ is

(1.9) $$\widehat{VAR[b1]} = \frac{\sigma^2}{\sum_{k=1}^{n}\left(\log L_k - E[\log L]\right)^2}$$

and the variance of the intercept parameter *b*0 is

(1.10) $$\widehat{VAR[b0]} = \frac{\sigma^2 \cdot \sum_{k=1}^{n}\left(\log L_i\right)^2}{n \cdot \sum_{j=1}^{n}\left(\log L_i - E[\log L]\right)^2}$$

These variances can be used to construct confidence intervals for the parameters *b*0 and *b*1, using the standard deviations of *b*0 and *b*1 where

(1.11) $$\widehat{STDEV[b0]} = \sqrt{\widehat{VAR[b0]}} \;\; and \;\; \widehat{STDEV[b1]} = \sqrt{\widehat{VAR[b1]}}$$

Given the MLEs of the regression parameters *b*0 and *b*1, the MLE of the exponent parameter *B* for the original allometric equation is simply

(1.12) $$\hat{B} = \widehat{b1}$$

The standard deviation of *B* is simply equal to the standard deviation of *b*1. That is

(1.13) $$\widehat{STDEV[B]} = \widehat{STDEV[b1]}$$

The MLE of the parameter *A* needs to be back-transformed from the logarithmic scale to obtain the parameter value in the original scale. The naive estimate of *A* is *A* = exp(*b*0). It can be shown that this estimate has a negative bias (Hayes et al., 1995). That is, the expected value of the naive estimate is less than the true value of *A*. This negative bias results from the fact that the regression was based on log-transformed (Miller, 1984). In particular, the basis for the linear regression model changes from the arithmetic mean in the original data units to the geometric mean in log-transformed units. The negative bias can be approximately corrected by multiplying the *A* parameter by $\exp(0.5\sigma^2)$, where σ is the estimated residual variance of the regression model fit (Hayes et al., 1995). Thus, the bias-corrected *A* parameter is

(1.14) $$\hat{A} = \exp\left(\widehat{b0}\right)\exp\left(\frac{\widehat{\sigma^2}}{2}\right)$$

The bias-corrected standard deviation of *A* (STDEV[*A*]) can also be approximated in a similar manner as

(1.15)
$$\widehat{STDEV\left[A\right]} = \exp\left(\sqrt{\widehat{VAR\left[b0\right]}}\right)\exp\left(\frac{\widehat{\sigma^2}}{2}\right)$$

The value of the coefficient of determination for the regression analysis (R^2) can be derived from the residuals of the regression fit as

(1.16)
$$R^2 = 1 - \frac{\sum_{k=1}^{n} \varepsilon_k^2}{\sum_{j=1}^{n} \left(\log W_j - E\left[\log W\right]\right)^2}$$

The R^2 value provides a measure of the goodness-of-fit of the linear regression model to the length-weight data with higher R^2 values indicating better fits to the observed data.

SUMMARY

The allometric equation (equation 1.1) is a commonly used model to predict fish weights from fish lengths. Maximum likelihood estimates of the scale (*A*) and shape (*B*) parameters of the allometric equation can be calculated from a linear regression on log-transformed fish weight and length data. The analytical formulas for the maximum likelihood estimates of the scale parameter *A* and its variance have been provided (equations 1.6, 1.9, 1.10, 1.13, and 1.14), where a bias-correction factor has been included to adjust for transformation bias of the scale parameter. Similarly, the formulas for maximum likelihood estimates of the shape parameter *B* and its variance have also been provided (equations 1.5, 1.8, 1.11, and 1.12), and it is noted that no bias correction factor is needed for the shape parameter.

REFERENCES

Hayes, D. B., J. K. T. Brodziak, and J. B. O'Gorman.
1995. Efficiency and bias of estimators and sampling designs for determining the length-weight relationships of fish. Can. J. Fish. Aquat. Sci. 52:84-92.

Larsen, R. J., and M. L. Marx.
1981. An introduction of mathematical statistics and its applications. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 536 p.

Miller, D. M.
1984. Reducing transformation bias in curve fitting. Am. Stat. 38:124-126.