mals: their productivity and population dynamics, IBP. Volume 5. Cambridge University Press, London, England.
Mahmud, F. S. 1980. Alary polymorphism in the small brown planthopper *Laodelphax striatellus* (Homoptera: Delphacidae). Entomologia Experimentalis et Applicata **28**: 47–53.
McCoy, E. D., and J. R. Rey. 1981. Alary polymorphism among the salt marsh Delphacidae (Homoptera: Fulgoroidea) of northwest Florida. Ecological Entomology **6**:285–291.
Rey, J. R. 1981. Ecological biogeography of arthropods on *Spartina* islands in northwest Florida. Ecological Monographs **51**:237–265.
Stiling, P. D., and D. R. Strong. 1982. The parasitoids of the planthopper *Prokelisia marginata* (Homoptera: Delphacidae). Florida Entomologist **65**:191–192.
Varley, G. C., G. R. Gradwell, and M. P. Hassell. 1973. Insect population ecology. Blackwell, Oxford, England.
Watt, A. D., and A. F. G. Dixon. 1981. The role of cereal growth stages and crowding in the induction of alatae in *Sitobion avenae* and its consequences for population growth. Ecological Entomology **6**:441–447.

[1] *Manuscript received 21 September 1981; revised 24 May 1982; accepted 25 May 1982.*
[2] *Department of Biological Science, Florida State University, Tallahassee, Florida 32306 USA.*

*Ecology*, 64(1), 1983, pp. 209–210


# *CORRECTING FOR BIAS IN LOG-TRANSFORMED ALLOMETRIC EQUATIONS*[1]

D. G. Sprugel[2]

Logarithmic transformations are used routinely in dimension analysis studies to fit allometric equations to sample data. That is, in fitting to sample data an equation of the form

$$y = a \cdot x^b,$$

where $x$ is typically stem diameter and $y$ is biomass or productivity of some tissue, it is common practice to replace all values by their logarithms, so that the equation becomes:

$$\log y = \log a + b \cdot \log x.$$

This transformation greatly simplifies the calculations, since standard least-squares regression techniques can be used to fit the line. Moreover, in most cases it increases the statistical validity of the analysis by homogenizing the variance over the entire range of the sample data. Commonly, the spread of the points around the best-fit line is much greater for large values of $y$ than it is for small values; the log transformation tends to equalize the variance over the entire range of $y$-values, which satisfies one of the prerequisites for proper use of parametric regression. However, the transformation also introduces a systematic bias into the calculations, and it has now become fairly widely recognized that a correction factor is necessary to counteract this bias (Finney 1941, Baskerville 1972). Unfortunately, the formula for the correction factor is often given incorrectly, apparently due to a widespread misunderstanding of its derivation.

The nature of the bias produced by log transformation has been explained in detail by several authors (Baskerville 1972, Whittaker and Marks 1975). To eliminate the bias, the final result is usually multiplied by a correction factor which is calculated from the standard error of estimate (SEE) of the regression. SEE is in turn calculated from the formula:

$$\text{SEE} = \sqrt{\Sigma(\log y_i - \widehat{\log y_i})^2/(N - 2)},$$

where the $\log y_i$'s are the values of the dependent variable and the $\widehat{\log y_i}$'s are corresponding predicted values calculated from the equation (Snedecor and Cochran 1967:138, Steel and Torrie 1980:253). (The denominator in this equation is incorrectly given as $N - 1$ in the seminal and widely quoted article by Whittaker and Woodwell [1968]. The formula given by Whittaker and Woodwell is correct only for calculating the standard error of estimate of a linear regression through the origin, where only one parameter is being fitted. In a log-transformed allometric equation, two parameters [the coefficient and the exponent] are involved, and so the denominator in the formula must be $N - 2$. When three parameters are fitted [e.g., in equations of the form $y = a \cdot x_1^b \cdot x_2^c$, where $x_1$ and $x_2$ are different independent variables], the denominator should be $N - 3$.) What is sometimes not realized is that the value of the standard error of estimate as calculated above depends on the base to which the logarithms are taken when the values are transformed; base-10 logs give a different value for the standard error of estimate than natural (base-$e$) logs. Thus, when the correction factor is calculated according to the usual formula:

$$CF = \exp(\text{SEE}^2/2),$$

it too depends on the base to which the logarithms were taken.

To obtain the correct value for the correction factor, SEE in the formula must be based on natural logarithms. This is clear from the derivation of the formula, in which the exponential function appears not only as the inverse of the logarithmic function but also in the probability density function for the normal distribution. Thus, using a base-10 standard error in the formula does not give the correct value; moreover, the answer still comes out wrong even if "exp" in the formula is replaced by "$\text{antilog}_{10}$." The correct procedure for dealing with a base-10 SEE is to convert it to base $e$ (by multiplying by $\log_e 10 = 2.303$) and use the base-$e$ value in the formula above.

The caveat discussed here does not invalidate published values of "$E$," the estimate of relative error of the regression (Whittaker and Woodwell 1968), since that parameter is calculated by taking the antilog (to the appropriate base) of the standard error of estimate, which eliminates the effect of different bases. (The difference between this and the correction factor is that in calculating the correction factor SEE is squared before the antilog is taken.) However, it should be clearly understood that $E$ is not in any sense a standard error itself and cannot be treated as such statistically. The formula for the actual standard error of estimate (in arithmetic units) of the regression line is given by Baskerville (1972). This value too should be used with care, since the distribution on which it is based is highly skewed (in arithmetic units) and does not fit the requirements for typical parametric statistical tests.

The logarithmic correction factor is a simple and straightforward statistical tool to remove a systematic bias and should be used whenever logarithmic transformations are used to fit allometric equations to data. It is true that the errors involved are usually fairly small (generally 10% or less), but since they are known errors and easily correctable the correction should always be made. The fact that other factors may introduce larger, uncorrectable errors into dimension analysis calculations is no excuse for failing to correct one of the few sources of error which can be avoided.

*Acknowledgments.* I would like to thank Walt Westman for productive discussions which led to the writing of this paper, and W. Westman and C. W. Ramm for helpful comments on the manuscript. Michigan Agricultural Experiment Station Journal Article Number 10250.

## *Literature Cited*

Baskerville, G. L. 1972. Use of logarithmic regression in the estimation of plant biomass. Canadian Journal of Forest Research **2**:49–53.

Finney, D. I. 1941. On the distribution of a variate whose logarithm is normally distributed. Journal of the Royal Statistical Society Series B **7**:155–161.

Snedecor, G. W., and W. G. Cochran. 1967. Statistical methods. Sixth edition. Iowa State University Press, Ames, Iowa, USA.

Steel, R. G. D., and J. H. Torrie. 1980. Principles and procedures of statistics. Second edition. McGraw-Hill, New York, New York, USA.

Whittaker, R. H., and P. L. Marks. 1975. Methods of assessing primary productivity. Ch. 4 *in* H. Leith and R. H. Whittaker, editors. Primary productivity of the biosphere. Springer-Verlag, New York, New York, USA.

Whittaker, R. H., and G. M. Woodwell. 1968. Dimension and production relations of trees and shrubs in the Brookhaven forest, New York. Journal of Ecology **56**:1–25.

[1] *Manuscript received 23 February 1982; accepted 30 June 1982.*
[2] *Department of Forestry, Michigan State University, East Lansing, Michigan 48824 USA.*