Eugene Giles,[1] *Ph.D. and Linda L. Klepinger,*[1] *Ph.D.*

Confidence Intervals for Estimates Based on Linear Regression in Forensic Anthropology

REFERENCE: Giles, E. and Klepinger, L. L., **"Confidence Intervals for Estimates Based on Linear Regression in Forensic Anthropology,"** *Journal of Forensic Sciences.* Vol. 33, No. 5, Sept. 1988, pp. 1218-1222.

ABSTRACT: Forensic anthropologists commonly use simple linear regression to estimate the value of a dependent variable, such as stature, for a single specimen where the value of the independent variable, such as humerus length, is known. Published studies providing regression equations for such use almost invariably include the standard error of estimate. Unfortunately, it is exceptional for forensic anthropologists to use the standard error to calculate correctly the confidence limits for their single predicted value. We attempt to show why this may be and provide explicit guidelines for the proper construction of confidence intervals in such circumstances.

KEYWORDS: physical anthropology, human identification, statistical analyses, linear regression, standard error of estimate, stature

Linear regression is widely used in forensic anthropology to provide an estimate of a measurement, or variable, unobtainable directly. Linear regression formulas have been employed in estimating age at death, but perhaps the most common usage is the determination of the living stature of an individual now represented by one or more long bones. Such formulas take the form $Y = a + bX$, where X is the long-bone measurement, or the independent variable, and Y is stature, or the dependent variable (dependent because its relationship to the long-bone lengths is what is to be determined, thus the regression of Y on X, which is not the same as the regression of X on Y). The intercept and the slope of the regression line (a and b, respectively) are determined by the method known as least squares. Published regression formulas almost invariably include a plus-minus number known as the sample standard error of estimate. This statistic is no more than simply the square root of the average of the squared errors (error being defined as the observed value minus the predicted value) for the sample. (Technically, an unbiased estimate requires the average to be the result of dividing by the sample size $N - 2$ rather than N.)

Recently, the forensic science community has shown concern over uncritical use of statistics, not only in forensic anthropology [*1,2*] but more generally. Zerr [*3*] and Aitken [*4*] have examined the use of regression analysis, the latter noting that generally in the measurement of variation, "too often have I seen, for example, confidence intervals wrongly constructed or wrongly interpreted" [*4*].

In forensic anthropology, the misinterpretation of confidence intervals for individual estimates based on regression equations may be traceable to an article by Professor Edward

Received for publication 14 Sept. 1987; revised manuscript received 21 Dec. 1987; accepted for publication 6 Jan. 1988.

[1]Professor, Department of Anthropology, and associate professor, Departments of Anthropology and of Cell and Structural Biology, respectively, University of Illinois at Urbana-Champaign, Urbana, IL.

Keen [5], a surgeon, on stature estimation. Keen stated that "the error in using regression formulae for reconstruction of stature from long bones is statistically expressed as the 'standard error of the estimate.' This is the measurement of the error likely to be made in reconstructing the stature of an individual known to belong to a population similar to that from which the regression formula was calculated" [5]. His example was that if an individual's stature had been estimated as 180 cm from a known femur length, and the standard error (SE) of estimate was ±3.2 cm, then an interval of ±2 SE around the estimate (that is, 173.6 to 186.4 cm) would not include the correct stature only 1 out of 22 times by chance. Now, if both variables (stature and femur length) are normally distributed, we *can* say that our errors should not exceed ±6.4 cm (±2 SE) more than 1 out of 22 times, or approximately 5% of the time. Unfortunately, we *cannot* say this, as Keen did, for a single, specific predicted value for stature, but only as a general proposition if the method of prediction is applied a great number of times to normally distributed data. For a single predicted value, a modified standard error, described below, must be calculated.

Regrettably, the example which displayed Keen's misunderstanding of this particular use of the standard error of estimate was picked up in its entirety in the first American text on forensic anthropology, Professor Wilton Krogman's *The Human Skeleton in Forensic Medicine* [6]. Krogman's well-earned authority in forensic anthropology undoubtedly gave credibility and widespread dissemination to Keen's statistical misstep. Krogman's reprinting of Keen's example was further flawed by a typo not in the original (limits of ±3 SE exclude 3 of 1000 cases, not 1 of 1000). Unfortunately, both the erroneous example and the typo are included in the second edition [7].

The standard error of the estimate is a measure of how much the individual observations of the original data base vary from the regression line. But in forensic anthropology it is more appropriate to determine the confidence interval for a single predicted value of the dependent variable. As it turns out, confidence intervals for predictions from regression equations are not parallel bands on either side of the regression line, but are the arms of a hyperbola (Fig. 1). The point at which the arms are closest together is the mean of the *X* values; the further the observed individual *X* is from the mean, the greater the confidence interval for the predicted *Y* value. This arrangement is quite understandable, since to make these estimates in the first place it is assumed that the distributions of *X* and *Y* are bivariate normal or approach it. Clearly then, the security of the prediction is greater near the mean value of

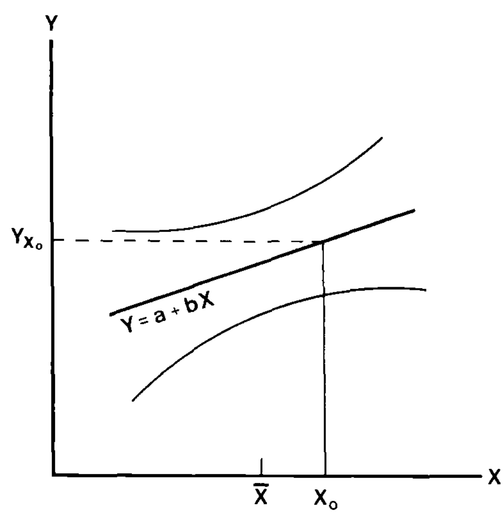

FIG. 1—*Confidence intervals for single predicted value of* Y *for given* X.

the distribution, and less at either end. Properly constructed confidence intervals should reflect this.

The paper by Maples and Rice [*8*] is one of the few by forensic anthropologists to recognize and discuss the nonlinear nature of the confidence interval band for predicted values from regression equations. The authors cite Blalock [*9*], a standard social science statistics text, in noting that "computations are tedious and rarely used." What Maples and Rice are interested in, and what Blalock finds tedious, however, is plotting the entire confidence interval band for a regression line as shown in Fig. 1. Fortunately, there is no need to do this extensive computation when one is concerned with a confidence interval for a *single* predicted value of X, rather than for *all* values, that is, a band.

The method for obtaining confidence intervals for a single predicted value of Y obtained from a known X by simple linear regression is presented in many statistics texts [*10–12*], but the format found in the lucid presentation of regression analysis, particularly appropriate for physical anthropologists, by Simpson, Roe, and Lewontin [*13*] is used here. If Y_{X_0} is the predicted value of Y for a given X_0, then the confidence interval for Y_{X_0} is

$$Y_{X_0} \pm t s_{YX} \sqrt{1 + \frac{1}{N} + \frac{(X_0 - \bar{X})^2}{(N-1)s_X^2}}$$

where

s_{YX} = sample standard error of estimate for the regression of Y on X,
X_0 = known individual X,
$\bar{X}$ = mean of the sample values of X,
s_X^2 = variance of the sample values of X,
N = sample size, and
t = t-distribution value at the desired probability level with $N - 2$ degrees of freedom.

Data from a paper by Snow and Luke [*14*] (reprinted in Ref *15*) can be used to illustrate the proper construction of confidence intervals for a single predicted value. Snow and Luke were faced with the problem of estimating the stature of a white female for whose humerus and femur they could obtain maximum length measurements. To use the humerus, they selected from the stature-on-length-of-long-bone regression formulas published by Trotter and Gleser [*16,17*] the one for white females, $Y = 3.36X + 57.97$, where Y = stature and X = humerus length, and entered their value for X, 19.8 cm. This yields a predicted stature Y of 124.5 cm. Trotter and Gleser's papers provide the necessary information for the calculation of the confidence interval for a single estimate: for the white female sample $N = 63$, the mean humerus length is 30.43 cm with a standard deviation of 1.728 cm, and the standard error of estimate for the regression is ±4.45 cm. Entering these figures in the formula given above for confidence intervals, and choosing the 95% level of confidence, t_{975}, in a t-distribution table at 61 degrees of freedom, we have

$$124.5 \pm (2.00)(4.45)\sqrt{1 + \frac{1}{63} + \frac{(19.8 - 30.43)^2}{(63-1)(1.728)^2}}$$

$$124.5 \pm (8.9)\sqrt{1.016 + (112.997/185.131)}$$

$$124.5 \pm (8.9)\sqrt{1.626}$$

$$124.5 \pm 11.349$$

The confidence interval of ± 11.35 cm is approximately 2 in. (5 cm) larger than the ± 8.9-cm interval (which is twice the published value of the standard error of estimate) offered by Snow and Luke. The difference does not affect Snow and Luke's conclusions in this case, but it does illustrate that toward the extremes of the distribution, confidence intervals calculated correctly are going to be larger than the ones customarily cited by forensic anthropologists.

Fortunately, since they are so widely used, the regression formulas of Trotter and Gleser were published with the necessary data to calculate confidence intervals: standard error of estimate, mean and standard deviation of the independent variable, and the sample size. Not all researchers have done so in the past, for example, Ref *18*, but such information should be included in future publications involving the predictive use of linear regression. The calculation of correct confidence intervals is not a tiresome procedure with even simple modern calculators. With the expanding use among forensic anthropologists of data disks for personal computers [*19*], including on them the equation to calculate confidence intervals for single predicted values from regression formulas could make its computation truly routine. The correct determination of confidence intervals is especially important in increasing the power of expert testimony offered in the courtroom.

References

[*1*] Tuttle, R. H., "Review: L. M. Robbins' *Footprints: Collection, Analysis, and Interpretation,*" *American Anthropologist,* Vol. 88, No. 4, Dec. 1986, pp. 1000-1002.

[*2*] Kahane, D. and Thornton, J., "Discussion of 'Estimating Height and Weight from Size of Footprints'," *Journal of Forensic Sciences,* Vol. 32, No. 1, Jan. 1987, pp. 9-10.

[*3*] Zerr, D., "Abstract: The Use of Regression Analysis and ANOVA in the Forensic Sciences," *Canadian Society of Forensic Science Journal,* Vol. 20, No. 3, Aug. 1987, pp. 283-284.

[*4*] Aitken, C. G. G., "Statistics and Forensic Science—A Fruitful Partnership," *Journal of the Forensic Science Society,* Vol. 23, No. 1, Jan. 1983, pp. 3-4.

[*5*] Keen, E. N., "Estimation of Stature from the Long Bones," *Journal of Forensic Medicine,* Vol. 1, No. 1, July/Sept. 1953, pp. 46-51.

[*6*] Krogman, W. M., *The Human Skeleton in Forensic Medicine,* Charles C Thomas, Springfield, IL, 1962, p. 184.

[*7*] Krogman, W. M. and İşcan, M. Y., *The Human Skeleton in Forensic Medicine,* 2d ed., Charles C Thomas, Springfield, IL, 1986, pp. 345-346.

[*8*] Maples, W. R. and Rice, P. M., "Some Difficulties in the Gustafson Dental Age Estimations," *Journal of Forensic Sciences,* Vol. 24, No. 1, Jan. 1979, pp. 168-172.

[*9*] Blalock, H. M., Jr., *Social Statistics,* McGraw-Hill, New York, 1960.

[*10*] Huntsberger, D. V. and Billingsley, P., *Elements of Statistical Inference,* 5th ed., Allyn and Bacon, Boston, 1981.

[*11*] Ostle, B. and Mensing, R. W., *Statistics in Research: Basic Concepts and Techniques for Research Workers,* 3d ed., Iowa State University Press, Ames, 1975.

[*12*] Neter, J., Wasserman, W., and Kutner, M. H., *Applied Linear Statistical Models,* 2d ed., Richard D. Irwin, Homewood, IL, 1985.

[*13*] Simpson, G. G., Roe, A., and Lewontin, R. C., *Quantitative Zoology,* rev. ed., Harcourt, Brace, New York, 1960.

[*14*] Snow, C. C. and Luke, J. L., "The Oklahoma City Child Disappearances of 1967: Forensic Anthropology in the Identification of Skeletal Remains," *Journal of Forensic Sciences,* Vol. 15, No. 2, April 1970, pp. 125-153.

[*15*] *Human Identification: Case Studies in Forensic Anthropology,* T. A. Rathbun and J. E. Buikstra, Eds., Charles C Thomas, Springfield, IL, 1984, pp. 253-277.

[*16*] Trotter, M. and Gleser, G. C., "Estimation of Stature from Long Bones of American Whites and Negroes," *American Journal of Physical Anthropology,* Vol. 10, No. 4, Dec. 1952, pp. 463-514.

[*17*] Trotter, M. and Gleser, G. C., "Corrigenda to 'Estimation of Stature from Long Bones of American Whites and Negroes,' American Journal of Physical Anthropology (1952)," *American Journal of Physical Anthropology,* Vol. 47, No. 2, Sept. 1977, pp. 355-356.

[*18*] Tibbetts, G. L., "Estimation of Stature from the Vertebral Column in American Blacks," *Journal of Forensic Sciences,* Vol. 26, No. 4, Oct. 1981, pp. 715-723.

[19] Maples, W. R., "The Microcomputer in Forensic Anthropology," *Forensic Osteology: Advances in the Identification of Human Remains,* K. J. Reichs, Ed., Charles C Thomas, Springfield, IL, 1986, pp. 308-319.

Address requests for reprints or additional information to
Eugene Giles, Ph.D.
Department of Anthropology
109 Davenport Hall
University of Illinois at Urbana-Champaign
607 South Mathews
Urbana, IL 61801