

Methods for Analysis of Unbalanced, Longitudinal, Growth Data

LINCOLN E. MOSES,^{1,2} LYNN C. GALE,² JEANNE ALTMANN^{3,4,5}

¹*Department of Statistics, Stanford University, Stanford, California;* ²*Center for Advanced Study in the Behavioral Sciences, Stanford, California;* ³*Department of Ecology and Evolution, University of Chicago, Chicago Illinois;* ⁴*Department of Conservation Biology, Chicago Zoological Society, Brookfield, Illinois;* ⁵*Institute of Primate Research, National Museums of Kenya, Karen, Nairobi*

We describe an approach to analysis of growth that does not depend on assumptions about the underlying functional growth pattern and that allows for multiple observations arising from individual-specific, irregularly spaced data. We produce estimated growth curves for predefined subject groups by using LOWESS, a nonparametric smoothing algorithm. We describe how statistical significance of curve features may be evaluated by using the "jackknife," a sample re-use method; this technique can be used to assess differences between subject groups. We then obtain residuals at each data point by reference to the estimated curve. Consistency of residuals is evaluated as a characteristic of individual subjects, and in the presence of individual consistency, relative size-for-age is then scored by the average residual for each individual. This allows study of relationships between relative size and other individual characteristics such as birth order, dominance rank, or age of maturation. Finally, we indicate flexibility of these methods and alternatives, propose uses related to other questions about growth, and suggest potential applications to variables other than body size. Appendices demonstrate application of the LOWESS and jackknife algorithms to the problem of testing sex differences in growth. © 1992 Wiley-Liss, Inc.

Key words: irregular growth data, unbalanced data, nonparametric curve-fitting, jackknife, components of variance

INTRODUCTION

Despite the centrality of growth to many topics in biology, both theoretical and empirical progress have been impeded by the paucity of comparative data and by the limited analyses to which available data have been subjected. Longitudinal data on the growth of individuals have been virtually unavailable from free-living, unprovisioned mammals and relatively rare even from provisioned or captive animals. Moreover, the irregularities and unknown distributional properties that are

Received for publication July 16, 1991; revision accepted February 2, 1992.

Address reprint requests to Jeanne Altmann, Department of Ecology and Evolution, University of Chicago, 940 E. 57th St., Chicago, IL 60637.

characteristic of available data have constrained existing treatments [e.g., for nonhuman primates see Gavan, 1971; Glassman et al., 1984; Altmann & Alberts, 1987; Strum, 1991].

An emerging body of longitudinal growth data is now being obtained from repeated measurement of individual mammals under field conditions [e.g., Altmann & Alberts, 1987; Creel & Creel, 1989; Hofer & East, 1989 and in preparation], and these data sets promise to enhance greatly our understanding of growth in wild populations. Nonetheless, in these situations, as well as in many samples from captive animals, the data points are not only irregularly spaced, but variable in number and in spacing for different individuals. Richer methodological options will be needed if we are to realize the full potential from these longitudinal data.

In addition to problems of data irregularity, appropriate analysis is made more difficult because the underlying functional form of growth is usually unknown for animals in either captive or field situations. When the underlying functional form is not known, it is unattractive to "assume" such a form and then produce an estimated curve that is required to be a straight line, or, say, a quadratic or exponential function, depending on the assumption imposed. We prefer a method that allows the data to express itself in a trend not constrained by an initial assumption introduced as a mathematical convenience (see also Koziol et al. [1981] and Guo et al. [1987] for nonparametric analyses with incomplete but not irregular growth data, and Schluter [1988] and Cheverud et al. [1992] for nonparametric analyses of cross-sectional, single value data).

Therefore, our goal here is to present a computational approach that uses data that are both irregular and of unknown functional form in order to serve two ends:

1. To estimate a growth curve that embodies the age/body-mass data for all of the animals or for discrete subsets, such as males and females, to evaluate characteristics of such curves, and to compare different curves, assessing statistical significance.
2. To evaluate each individual's data points as referenced to the estimated curve, thereby displaying information, individual by individual, that can illuminate questions such as these: Is being small-for-age or large-for-age a stable characteristic of individuals during ontogeny? Can we identify environmental, parental, or individual predictors of an individual's relative size? Is relative size, that is, being small-for-age or large-for-age, predictive of important biological sequelae such as high mortality or precocial reproductive maturation?

ANALYSES

The analysis approach that we describe below was motivated and is illustrated by analysis of growth in free-living baboons, *Papio cynocephalus*, for which longitudinal data were collected during animals' voluntary, opportunistic use of an unbaited scale placed near their sleeping groves. The method of data collection for the baboon data has been described in more detail previously [Altmann & Alberts, 1987; Altmann & Samuels, 1989].

The two stages of analysis that we outline correspond to the two sets of questions above. First, we sought to characterize average size across ages, and to identify major classes of individuals, such as males and females, that differed greatly in growth patterns. Second, we sought a metric that we could use for each individual to describe within-class variability in growth.

Method for Constructing the Average-Size Curve and Analyzing Features

Four subsets of animals, characterized by sex and by feeding conditions (wild-foraging vs. partially garbage-feeding) were treated separately as follows. All the

points from every juvenile in the subset were plotted on one graph, with the horizontal coordinate being age and the vertical coordinate being body mass at that age. Age was recorded in months, and where several weighings were made of an individual in one month, the average was used for that month. For both feeding conditions, data were included only up to the age of adulthood for wild-feeding animals (six years for females, eight years for males). A plot of body mass vs. age for these data showed a swarm of points that rather closely followed an upward sloping curve.

Within each subset we then applied the LOWESS algorithm (with “window” $f = 0.5$) [see, e.g., Cleveland & Devlin, 1988]. This algorithm produces an estimated average mass at every age found in the data set. A curve for the graph shows these estimated points, connected, each to the next, by short (imperceptibly) straight line segments. LOWESS stands for *locally weighted regression scatter plot smoothing*. The LOWESS procedure was favored over parametric smoothers such as polynomial regression because it does not depend on some underlying functional form for the curve as a whole or for large sections of it; rather, linearity is assumed only over very short, local age spans. In addition, the procedure makes use of all the data and is a “resistant” one, in that it “down-weights” or is not sensitive to outlier values (a similar, alternative nonparametric smoother would be a spline; see, e.g., Cheverud et al. [1992] for physical growth, Schluter [1988] for fitness). Smoothing functions are available in several microcomputer statistical graphics packages. We chose a package that would not only plot the smoothed function but that would also allow access to the values for our subsequent manipulations to investigate individual variability within subsets. We used AxumTM software on a DOS-based microcomputer to calculate and plot the LOWESS curves (Fig. 1). For details see the Axum manual [TriMetrix, 1990]; for background and more information on the mathematics of the technique, see Cleveland [1979], Chambers et al. [1983], and Cleveland and Devlin [1988].

Figure 1a–d was produced in this way. Each panel presents the data points for one subset of animals defined by sex and feeding condition; those points are seen in relation to the LOWESS curve based on them. Also shown in the figure for each subset are the LOWESS curves (but not the points) for the other three subsets. The figure shows good fit of each set of points to their respective curves, though some straggling points occur among the older garbage-feeding animals. The curves show that feeding condition is associated with larger growth differences than is sex and that among the older juveniles and subadults in both feeding conditions males are larger for age than females.

Despite not depending on a parametric form or on independence among points, the estimated curves allow testing for significance of differences between any two of these curves, for example between wild-feeding males and females, as follows. We ask if the average height of the males’ curve significantly exceeds the average height of the curve belonging to the females. This is equivalent to asking whether the area under the one curve significantly exceeds the area under the other. We address that question by applying the “jackknife” [e.g., Mosteller & Tukey, 1977] to each area separately, which provides two estimated areas $A_{J,1}$ and $A_{J,2}$ and standard errors for them, $se_{J,1}$ and $se_{J,2}$. Then the difference between the two area estimates is compared to the estimated standard error of that difference, which is, as usual, the square root of the sum of the two squared standard errors. Details of the jackknife’s application to this problem appear in Appendix A. The computer implementation for the example of wild-feeding males and females is available from the authors. We note here that the jackknife should be applied using each individual, including all its data points, as the statistical unit.

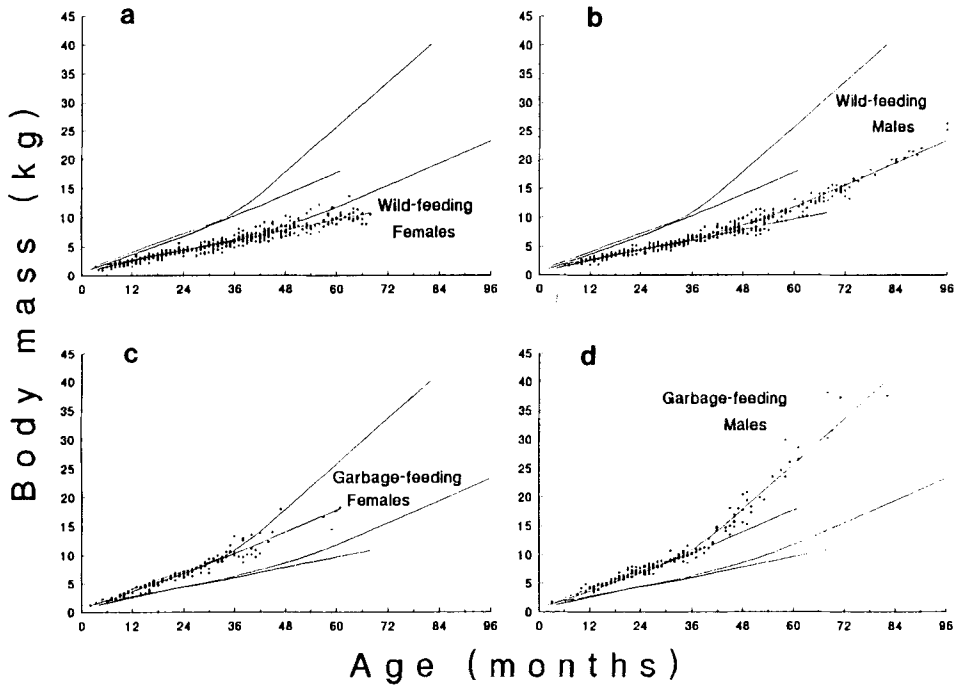


Fig. 1. Body mass as a function of age for free-living baboons, plotted separately by sex and feeding condition. Each point represents the mean of measurements for an individual for that month. Individuals contributed from 1 to 23 points to the data set. Lines represent the LOWESS-smoothed curves for each of the subsets (see text for details).

Estimating the area under a LOWESS curve can be very usefully generalized. Any interesting numerical property of the curve (a “functional”) can be estimated from it and then the jackknife will produce a less biased estimate with a standard error for that estimate. The area under the curve is a functional; another is the apparent increase in slope seen in some of our LOWESS curves. Is the appearance of this upward bowing (i.e., “growth spurt”) in the curve for wild-feeding males statistically significant or might it be a chance phenomenon? At issue here is the reliability of the tendency for the middle of the curve to lie below a straight line joining a point from the early part and one from the late part of the curve. We could proceed in this way: choose an age belonging to the middle (to illustrate, we chose 48 months), and then choose two points at equal distance from that value (e.g., 24 and 72 months). Now let M_{24} , M_{48} , and M_{72} be the body mass read from the curve at those ages, and construct the difference:

$$D = \frac{1}{2} (M_{24} + M_{72}) - M_{48}.$$

This D is our functional. If the true trend were a straight line, D would have expected value zero, and positive and negative values of D would be about equally likely. Applying the jackknife will give a jackknifed average D_j and a standard error for that estimate. And if D_j exceeds zero by more than 2 or 3 standard errors, the upward bowing would be statistically significant. More formally, the ratio of D_j to its estimated standard error can be compared to significance points of Stu-

dent's t-distribution with degrees of freedom equal to one less than the number of pseudo-values used in calculating D_j .

The jackknife is a "sample re-use" procedure and is widely applicable to non-standard problems [see discussions in Mosteller & Tukey, 1977; Sokal & Rohlf, 1981]. Another such sample re-use procedure is the "bootstrap" [e.g., Diaconis & Efron, 1983]. It would be a satisfactory alternative to the jackknife for our purposes, though it might require more computation, which is why we have applied the jackknife. (Again, note that the unit of re-sampling must be the entire record of an individual.)

Method for Constructing Age-Relative Sizes for Individuals: Residual Analysis

The LOWESS curve can be thought of as showing average body mass as a function of age for the set of individuals whose data it summarizes. This curve then affords a way to take individual body mass measurements at different ages and render them comparable: each can be expressed relative to the average body mass at that age. We use the natural logarithm of the ratio of the observed mass (at age t_i) to the average body mass given by the LOWESS curve at that age; these values are "residuals." An observed mass lying exactly on the curve would result in a residual of zero. Individuals large for their age produce positive residuals; those that are small produce negative residuals.

Most animals in our example had several weighings, and thus several residuals. Figure 2 shows the residuals for eight animals. The residuals for different individuals occur in somewhat different parts of the age range, are variable in spread across the range, and vary in numbers of points. Most animals show a clear persistence over time. Pepo, for example, is always large for age; Jitu is always small. For each exemplar animal, except perhaps Pombe, the size of its residuals is quite similar across time. Of those animals with three or more observations in the full data set, 80% showed no significant (.05) time trend up or down, and only a very few crossed from small to large, or vice versa, like Pombe. Consequently, we use the mean of an individual's residuals as an informative measure of its size-for-age. (Appendix B demonstrates that little efficiency is lost for these data by this choice.) The average of a juvenile's residuals is then usable in studying questions that involve relative size and other variables, such as the effect of mother's rank, or the juvenile's age at maturation, by use of standard statistical methods, such as regression, chi-square analysis, and the like (as in Altmann, ms.).

DISCUSSION

The method we have described has several features: 1) It provides an average-size curve, and the form the curve takes is found in the data rather than in any initial assumption about the shape. 2) The LOWESS algorithm is a resistant one; that is, it down-weights extreme observations, which we think is desirable in estimating a trend for average size. 3) The jackknife can produce approximate confidence intervals for any numerical property of the curve (any functional). We have given two illustrations of jackknifing such functionals: first, for the difference between the curves for two groups, and second, for assessing whether an apparent upward bowing is statistically significant. 4) By means of their average residual, individual animals are quantitatively scored on relative size-for-age, which we have found to be a persistent individual characteristic. Such individual scores are then usable as predictors of aspects of future development and as outcomes to be related to antecedents such as mother's rank, birth order, etc.

We now point to other uses of the method for analyzing other aspects of growth.

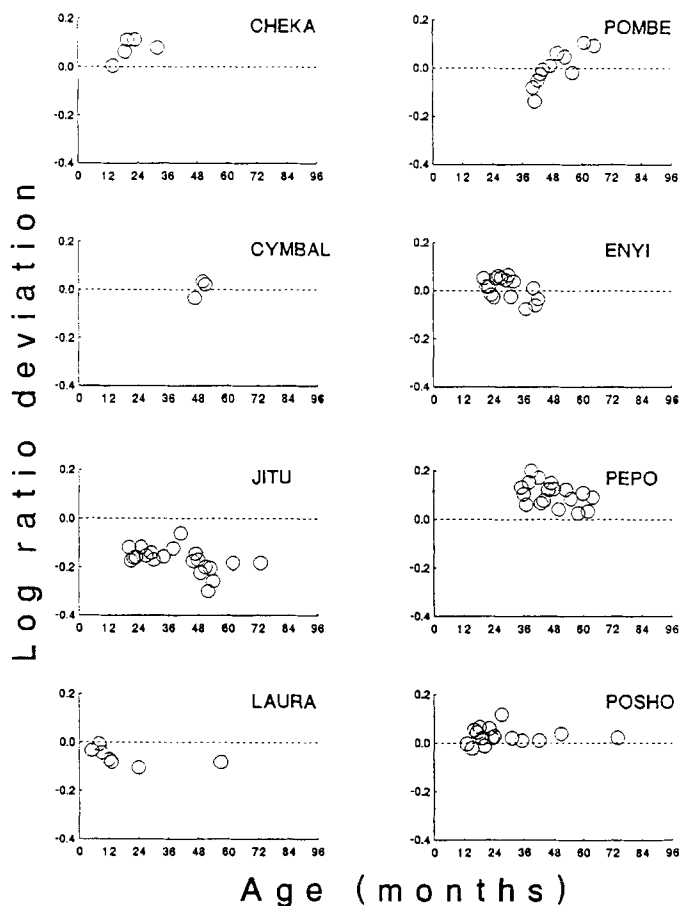


Fig. 2. Body-size residuals as a function of age for eight representative baboons (see text for details).

First, comparing some other large set of baboon growth data to ours can be done in the same manner as was used to compare two of our groups. However, if only a few records are available from another source (rather than a large set of data) the set of average size-for-age residuals of the new individuals (using the Amboseli curve) can be referred to the distribution of average residuals of the Amboseli individuals, by means of the Wilcoxon two-sample test (for instance). Likewise, if two small data sets are available, one might construct a single curve and the set of average size-for-age residuals of one set of individuals can be compared to the distribution of average residuals of the other.

Second, the approach can be extended to estimation algorithms other than LOWESS. A piecewise linear (or "hockey-stick") regression function might be used as a model for juvenile growth, in order to identify an adolescent spurt. Where every data point is statistically independent of every other, classical statistical methods allow estimation and significance testing. But where each individual furnishes several statistically dependent points, such significance tests are invalid. Nonetheless the estimated function, which is obliged to lie near the data, may still be useful, and its statistical uncertainty can be validly assessed by applying the

jackknife, omitting, in turn, each individual, with all its data points. One result of such an analysis would be a confidence interval for “the time T^* at which the spurt begins.” We used quotation marks to indicate that though we can come up with an estimate, and a confidence interval for it, we are in fact dealing with an idealized construct; we do not possess sufficiently detailed data or theory to rule out, for example, an interval of rapid acceleration from slow growth to rapid growth, rather than a single point in time, T^* . We have thus far suggested two approaches to assessing a growth spurt. A third approach, less model-dependent than reliance on the construct T^* , would exhibit the first differences of the LOWESS curve (approximating its derivative) and choose some functional of that curve, then assess the statistical uncertainty of that functional by use of the jackknife.

We remark that the key ideas, really only three, each allow some latitude regarding particulars. The first idea, estimation of an average descriptive curve that does not depend on some imposed model, can be done with LOWESS, but other data analytic smoothing techniques such as spline, moving averages, running medians, and their cousins might also be used. The second idea, assessment of statistical significance of features of the estimated average curves by means of sample re-use methods, is done here with the jackknife, but the bootstrap might be a satisfactory alternative, and would apply as well to situations in which a parametric growth curve was justifiable but in which individuals contribute several, statistically dependent, data points. The third idea, construction of residuals by referring observed data points to the smoothed values, can be done in more than one way, including (signed) arithmetic differences, ratios, or as we have preferred here, the logarithms of ratios.

We point to these areas of flexibility partly to highlight what is essential to the methods and what is not, partly to acknowledge that some data sets may call for modifications of some of the particulars that we have used. We are not suggesting that several variants should be tried on a given data set in some sort of quest for statistical significance.

Finally, we believe the methods set forth have applicability to variables other than body size that interest primatologists or biologists more generally. Thus reproductive success might be approached as follows. Begin by defining $A(t)$, reproductive accomplishment by age t , of a female in some way, perhaps number of live offspring by age t , or number of weaned offspring by then, or number of her offspring reaching maturity by her age t . Now each female in the group furnishes several points with coordinates t_i (horizontal) and $A(t_i)$ (vertical). If a 14-year-old female has had three offspring reach maturity, at her ages 11, 12½, and 13½, then her data set would comprise three points, having elevations 1, 2, and 3, respectively, at those time points. A LOWESS smooth over the points of all individuals in the set provides an idea of average reproductive accomplishment by age t , and individuals' residuals from that trend can be studied as predictors or outcomes, much as we have done with the body-mass data from juvenile baboons. Likewise, the approach could be used with, say, ontogenetic changes in dental scores or in behavioral measures of independence. We believe the method should be useful in at least some applications to non-monotonic functions such as in ontogenetic processes that increase, reach a peak, and then decline.

CONCLUSIONS

1. Sporadic growth data, with varying times and number of observations from different individuals, contain information on both the average-size curve and individual deviations from the average; but the data imbalance makes it difficult to estimate the two components.

2. The LOWESS smoothing algorithm is a useful way to estimate the average-size curve.

3. Each data point ($t_i, M(t_i)$) can be converted to a residual which is the natural logarithm of the ratio of body mass $M(t_i)$ at time t_i to the values of the LOWESS estimate at time t_i ; residuals are then positive or negative as points lie above or below the smoothed curve. If the residuals of any individual show marked consistency, the average of an individual's residuals is a useful indicator of size-for-age.

4. Numerical properties of the LOWESS curve, functionals of the curve, can be estimated (with valid standard errors) by means of the jackknife in which the entire data set of each individual is taken as the statistical unit. Examples of functionals: average height, extent of upward bowing of the estimated growth curve.

5. The jackknife with individual as the statistical unit can be used to give valid confidence intervals for parameters estimated by other methods (such as piecewise linear regression) where usual significance tests, based on independence of all data points, are not valid.

6. The methods proposed are very general, admitting of variation in the problems addressed, the smoothing procedure, the residual measure, and the sample re-use procedure employed.

ACKNOWLEDGMENTS

This project was conducted while J.A. was a Fellow at the Center for Advanced Study in the Behavioral Sciences where financial support was provided by the John D. and Catherine T. MacArthur Foundation. Virginia Heaton assisted with manuscript preparation. Tom Cooper of TriMetrix Technical Support provided useful comments on computer implementation. Susan Alberts, Stuart Altmann, James Cheverud, Sharon Emerson, Netzin Gerald, Gaea Leinhardt, Jane Phillips-Conroy, Carel van Schaik, Joan Silk, and three anonymous reviewers provided helpful comments on an earlier draft, and Anthony Coelho suggested several references.

REFERENCES

- Altmann, J. Ecological, demographic, and social factors as sources of intrapopulation variability in growth. 1992 (ms).
- Altmann, J.; Alberts, S. Body mass and growth rates in a wild primate population. *OECOLOGIA* 72:15–20, 1987.
- Altmann, J.; Samuels, A. Upscale baboons. *NATURAL HISTORY* 5/89:60–63, 1989.
- Bloch, D.A.; Moses, L.E. Nonoptimally weighted least squares. *THE AMERICAN STATISTICIAN* 42:50–53, 1988.
- Chambers, J.M.; Cleveland, W.S.; Kleiner, B.; Tukey, P.A. *GRAPHICAL METHODS FOR DATA ANALYSIS*. Boston, Duxbury Press, 1983.
- Cheverud, J.M.; Dittus, W.P.J.; Wilson, P. Primate population studies at Polonaruwa. III. Somatic growth in a natural population of toque macaques (*Macaca sinica*) *JOURNAL OF HUMAN EVOLUTION* (in press), 1992.
- Cleveland, W.S. Robust locally weighted regression and smoothing scatterplots. *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION* 74:829–836, 1979.
- Cleveland, W.S.; Devlin, S.J. Locally weighted regression: An approach to regression analysis by local fitting. *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION* 83:596–610, 1988.
- Creel, S.R.; Creel, N.M. Body weight in the dwarf mongoose: Relationships with environmental, behavioral, endocrine and life-history variables. Fifth International Theoretical Congress, Rome; 2:860, 1989.
- Diaconis, P.; Efron, B. Computer-intensive methods in statistics. *SCIENTIFIC AMERICAN* May 1983.
- Gavan, J.A. Longitudinal postnatal growth in chimpanzee. Pp 47–102 in *THE CHIMPANZEE*, Vol. 4. G. Bourne, ed. Basel, Springer, 1971.
- Glassman, D.; Coelho, A.; Carey, K.; Bramblett, C. Weight growth in savannah baboons: A longitudinal study from birth to adulthood. *GROWTH* 48:425–433, 1984.

- Guo, S.; Simon, R.; Talmadge, J.E.; Klabin-sky, R.L. An interactive computer program for the analysis of growth curves. *COMPUTERS AND BIOMEDICAL RESEARCH* 20:37-48, 1987.
- Hofer, H.; East, M. Growth of spotted hyena cubs in relation to prey availability, litter size and sex of offspring. Fifth International Theriological Congress, Rome; 2: 865, 1989.
- Koziol, J.A.; Maxwell, D.A.; Fukushima, M.; Colmerauer, M.E.M.; Pilch, Y.H.; A distribution-free test for tumor-growth curve analyses with application to an animal tumor immunotherapy experiment. *BIO-METRICS* 37:383-390, 1981.
- Mosteller, F.; Tukey, J.W. A method of direct assessment. *DATA ANALYSIS AND REGRESSION*. Reading, MA, Addison-Wesley, 1977.
- Schluter, D. Estimating the form of natural selection on a quantitative trait. *EVOLUTION* 42:849-861, 1988.
- Searle, S.R. *LINEAR MODELS*. New York, John Wiley & Sons, Inc., 1971.
- Sokal, R.R.; Rohlf, F.J. *BIOMETRY*. New York, W.H. Freeman and Company, 1981.
- Strum, S.C. Weight and age in wild olive baboons. *AMERICAN JOURNAL OF PRIMATOLOGY* 25:219-237, 1991.
- TriMetrix, Inc. *AXUM RELEASE NOTES, VERSION 1.02*. Seattle, TriMetrix, Inc., 1990.

APPENDIX A
The Jackknife

One has n independent observations, and a statistic T that is computed from them. T is an estimate of a quantity, θ . It is desired to give an indication of the standard error of T , which may be a very complicated function of the data. In our case T might be the area under a LOWESS-derived curve, or an indicator of upward bowing of such a curve.

In the simplest case the n independent observations are all single measurements (X_1, \dots, X_n) , and we shall at first explain the method for such data.¹

To apply the jackknife to $T(X_1, X_2, \dots, X_n)$ which we shall simply call T , we calculate that same statistic omitting the first X , and call that result $T^{(1)}$; then we omit instead the second X , obtaining $T^{(2)}$; we proceed in this manner until we have calculated $T^{(1)}, T^{(2)} \dots T^{(n)}$ (as well as T , involving all the data). Then we construct n pseudo-values. The first, which we call T_1 is defined thus:

$$T_1 = nT - (n - 1)T^{(1)}$$

and

$$T_2 = nT - (n - 1)T^{(2)},$$

etc. (Note that if T has very little sampling variation we should expect it not to matter very much which piece was omitted. So variability among the pseudo-values should reflect sampling variability.)

Next, we average the pseudo-values obtaining the jackknife estimate

$$T_J = \frac{1}{n} \sum_{i=1}^n T_i.$$

(This is ordinarily a less biased estimate of θ than the statistic T with which we began.) The standard error of T_J is computed from the T_i as if they were usual independent observations. That is,

¹But in our applications here, the independent observations are the data sets of the individual juveniles. Thus for one individual the observation may comprise three data points—body mass at three ages—and for another individual the observation may comprise one, or eight, or twenty data points, depending on the particular individual.

$$s.e. (T_J) = \sqrt{\frac{1}{n} \cdot \frac{\sum(T_i - T_J)^2}{n-1}}.$$

If n is large, such as 100, or even 30, then the amount of computation can be reduced by breaking the n observations up into a smaller number, k , of subsets (between 10 and 20 is good) each with the same number of observations. This partitioning must be done with a table of random numbers; then calculations proceed much as before. T uses all the data; $T^{(i)}$ uses all the data except those observations in the i th subset; the pseudo-values are

$$T_i = kT - (k-1)T^{(i)}$$

and the jackknife estimate is

$$T_J = \frac{1}{k} \sum_{i=1}^k T_i.$$

With our data for wild-feeding males there were 34 animals. Several had as few as one data point. So we constructed our subsets in this way: We took the 11 animals with the most data points and listed them in random order (List A). We took the next 11 animals (in terms of number of data points) and placed them in random order (List B). Finally, we randomly ordered the remaining 12 animals (List C). Then our first subset was made up of the top animals from lists A, B, and C. The second subset comprised the three animals that were in second position on the three lists. The eleventh subset contained the animals in eleventh place on the three lists plus the remaining individual from List C.

Then the LOWESS curve on all the data gave T , which was the area under that curve. The LOWESS curve computed omitting all the data points of animals in the first subset gave an area $T^{(1)}$; with the data points for the first subset restored, and those for the second subset omitted, a LOWESS fit was run, which gave the area $T^{(2)}$, and so forth. From $T, T^{(1)}, T^{(2)}, \dots, T^{(11)}$ were calculated the pseudo-values T_1, T_2, \dots, T_{11} , the jackknife estimate T_J , and its standard error as already described.

APPENDIX B

Efficiency of the Mean Residual for Unbalanced Data

Reducing each animal's residuals to a one-number summary, the animal's mean residual, can appear to be arbitrary when some animals have many, and some have only a few, or even a single residual. However, concern over this apparent defect often will not be warranted. Evaluation in any particular situation can be done using well-established statistical principles and the relative size of inter- and intra-animal variance of residuals, as we show below using our baboon sample for illustration.

Let r_{ij} be the j th measured residual of animal i and let n_i be the number of residuals for animal i .

Represent r_{ij} as follows:

$$r_{ij} = d_i + e_{ij} \quad j = 1, \dots, n_i.$$

Here we regard the measured residual as being made up of two components, d_i , the animal's "true" size-for-age, and e_{ij} , a perturbation from that, affecting this j th measured value. Neither of these can be observed directly. But it is the sizes of their variances, σ_d^2 and σ_e^2 , that matter. And both of these variances can be estimated from data such as ours, where at least some animals have multiple mea-

surements, using standard variance components methods [Searle 1971, pp. 437–438].

Now, if all the r_{ij} were independent, σ_d^2 would have to be zero. In our data this is far from the case (not surprisingly). Our estimates of σ_d^2 and σ_e^2 show σ_d^2 distinctly greater than σ_e^2 for three groups, with virtual equality in the garbage-feeding males. This circumstance, that $\sigma_d^2 \geq \sigma_e^2$, carries strong consequences, as we now detail.

Clearly, the average of the ten residuals of one animal has smaller variance than the average of the two residuals of another—but not by much:

$$\text{Var} \{ \bar{r}_i (10) \} = \sigma_d^2 + \frac{\sigma_e^2}{10}$$

$$\text{Var} \{ \bar{r}_i / (2) \} = \sigma_d^2 + \frac{\sigma_e^2}{2}.$$

If $\sigma_d^2 = \sigma_e^2$ these stand in the ratio of 1.1 to 1.5. More explicitly, if $\sigma_d^2 \geq \sigma_e^2$ then

$$\frac{\text{Var} \{ \bar{r}_i (n) \}}{\text{Var} \{ \bar{r}_i / (1) \}} < 2.0.$$

This means that carefully weighting the average residuals to take account of n_i can gain only a little efficiency. Bloch and Moses [1988] show in their Example 1 that when $\sigma_d^2 \geq \sigma_e^2$ the variance of an equally weighted average (or regression coefficient) is not so much as 12.5% greater than the variance of the optimally weighted statistic. Thus, for these data (with $\sigma_d^2 \geq \sigma_e^2$) we welcome the convenience of treating equivalently all the animals' average residuals.

In many biological contexts, unequal numbers of observations are available on individuals, and the principle applied here is applicable. In particular, where $\sigma_d^2 \geq \sigma_e^2$, very little information is sacrificed by reducing each individual's measures to an average for that individual, and ignoring the inequality in number of data points for the various individuals. Thus, calculating estimates of σ_d^2 and σ_e^2 in such situations can inform good decisions about appropriate data reduction.