THE STATISTICS OF VARIATION

Leigh Van Valen
Department of Biology
University of Chicago
1103 East 57th Street
Chicago, Illinois 60637
Received March 30, 1978

ABSTRACT: Univariate and multivariate measures of variation are reviewed, for both absolute and relative variation. Biologists almost never test variation appropriately; in particular, the *F* test should never be used. There are, however, several useful and robust tests which apply, with modifications, to all useful measures of variation, including the multivariate coefficient of variation. Some of these applications are new. A measure of the effective dimensionality of variation permits the relative variation of quantities such as volumes to be compared with that of linear measurements.

* * *

Organisms vary in their characters as well as in fitness, so natural selection occurs. However, this variation within populations blurs real distinctions among populations. For these reasons and others, we often want to estimate variation within populations and test for its equality among populations. Unfortunately, there are serious difficulties with some of the most common methods, and these difficulties are not widely known. I therefore review these difficulties and ways to avoid them, as well as the main properties of the structural statistics. Some modifications and applications are new.

My domain will be continuous variation and that discrete variation which can (as most can) be approximated by continuous distributions. Variation of more sharply discrete parameters, and of qualitative ones, is best measured by information-type statistics, which are beyond the scope of this review (but cf. Van Valen, 1974, for related suggestions).

Absolute variation: univariate case

NEVER USE AN *F*-TEST TO TEST EQUALITY OF VARIANCES. The basis for this shocking statement is well known to statisticians (Box, 1953; Pearson and Please; 1975) but is almost never communicated to biologists.

The variance, σ^2 for population and s^2 for sample, is the most widely suitable measure of variation in one dimension. Independent variances are additive; for variances of interdependent characters the amount of interdependence (the covariance) is subtracted from their sum. The standard deviation, σ or s, measures variation in the units in which measurements are taken and so is often heuristically preferable. Statistics such as the observed range, often excluding 10 or 25 percent at each end, are sometimes used. The full observed range is very sensitive to sample size and is rarely useful; truncated ranges can be useful when one is interested in only the central part of the distribution. Smith's test (see below) can be applied to ranges; see Kendall and Stuart (1963, p. 243) for the variance of estimation.

Unfortunately, the *F* distribution (the distribution of the ratio of two sample variances) is exceptionally sensitive to deviations from normality of the measured distributions. When the *F* distribution is applied in the ordinary analysis of variance, this is unimportant except in extreme cases. The ordinary analysis of variance tests equality of means, and by the central limit theorem

* * *

Evolutionary Theory 4:33-43 (September, 1978)
The editors thank F.L. Bookstein and R. Lande for help in evaluating this paper.

the distribution of means of samples from a given population is asymptotically normal. However, there is no such asymptotic normality for the variation of the character itself. It has even been half-seriously stated that the F test is a better test for normality than for equality of variances.

And one can't just test one's distributions for normality and assume them to be normal if they pass the test. Such tests just aren't sensitive enough except with very large samples. This is partly because of the difficulty in demonstrating any kind of small deviation from any distribution. It is also true because the tails of a distribution contribute disproportionately much to the variance and the tails are hardest to test. It can easily happen that a distribution can't be distinguished from normality but really is different enough to invalidate the F test. In practice, one can almost never have adequate reason to believe that a pair of distributions are each close enough to normality that the F test for equality of variances is suitable.

Almost every statistical comparison, in the biological literature, of the variances of two or more populations is wrong, often seriously so. Bartlett's test, being a generalization of the F test, is just as sensitive to nonnormality as is the F test itself. There are, however, three useful tests which are robust to nonnormality and were developed for this reason. Each has its advantages, but I have seen none of them applied more than once or twice in biology.

Levene's test: This is the simplest of the three tests. Its original description (Levene, 1960) is meant for statisticians, but the test itself is simple both conceptually and in execution. In formal terms, it tests the mean deviation rather than the variance.

For each sample, find the mean. Then calculate the absolute difference of each original datum from the mean. This is a new variable; the more varying the sample, the higher the values will be. Then find the mean and variance of the new variable. The means of the new variable for each population are then tested for equality by a t test or an analysis of variance. Symbolically, $y_i=|x_i-\bar{x}|$ and equality of the $\bar{y}$'s are tested.

Note that Levene's test isn't quite a test of equality of variances. It is, strictly speaking, a joint test of all the even moments of the distributions: variance, kurtosis, etc. However, the effect of the variance predominates.

Levine's test can be improved a little by using the median of the distribution instead of the mean as the center from which to measure deviations (Brown and Forsythe, 1974). And they say that the best procedure of all is to exclude the 10 percent of data at each end of the distribution and use the mean of the remaining 80 percent.

Smith's test: This is a very general but somewhat more complex test. It was developed by Cedric A. B. Smith and has been published, as far as I know, only in Grüneberg et al., 1966. It is a test directly of equality of variances. For a sample of size n, the variance of estimation of the variance (the square of the standard error of the variance) for the jth population is

$$s^2_{s^2_j} = \frac{\sum_i (x_i-\bar{x})^4 - s_j^2\left(\frac{n-3}{n}\right)}{(n-2)(n-3)}, \qquad (1)$$

correcting a misprint. For k samples, where k can be 2 or more, there is an approximate chi-square statistic with $(k-1)$ degrees of freedom:

$$\chi^2_{k-1} = \sum_j \frac{s_j^4}{s^2{}_{s_j}^2} - \frac{\sum_j \frac{s_j^2}{s^2{}_{s_j}^2}}{\sum_j \frac{1}{s^2{}_{s_j}^2}} . \qquad (2)$$

This test can be used to test for homogeneity for any statistic for which the standard error is available, by replacing s_j^2 with the desired statistic. Again, normality is not required. Smith's test is the only one of the three which can be used with other people's summary statistics of variation.

Jackknifing: This is the most complex but most general test. It can be used for correcting biases of estimation and for tests on almost any statistic. However, the sample size n must be moderately large (at least 20 or so). Good references are Arveson and Schmitz (1970), Miller (1974), and Bissell and Ferguson (1975). My treatment and symbols are simplified from these references.

First, divide the data into n broadly overlapping groups of size $(n\text{-}1)$ by eliminating one datum for each group, a different datum each time. Then estimate the variance for each group (s_i^2) and for the total (s^2_{total}). Let

$$\hat{\sigma}_i^2 = n s^2_{total} - (n\text{-}1) s_i^2, \qquad (3)$$

for the ith group. Then the bias-reducing estimate is

$$\hat{\sigma}^2 = \frac{1}{n} \sum \hat{\sigma}_i^2 . \qquad (4)$$

Because of the use of differences, at least 2 more significant figures than usual should be used in the calculations. The variance of estimation of $\hat{\sigma}^2$ is

$$s^2_{\hat{\sigma}^2} = \frac{\sum \left(\hat{\sigma}_i^2 - \hat{\sigma}^2 \right)^2}{n(n\text{-}1)} \qquad (5)$$

and is distributed as t with $(n\text{-}1)$ degrees of freedom. Thus ordinary t tests and analyses of variance can be performed on $\hat{\sigma}^2$. Bissell and Ferguson (1975) have given a straightforward (but necessarily even more complex) multivariate generalization. Jackknifing is reportedly somewhat more accurate than Levene's test. It is the only method which can be used to give confidence intervals for the variance.

To jackknife any other statistic, replace each s^2 in equation (3) by the appropriate sample estimate of that statistic. Order statistics such as the range and median may behave poorly, but the standard deviation can be used directly. Normality is not required for jackknifing but estimates are improved if there is a reasonable approximation to it.

Absolute variation: multivariate case

Beyond the mean there are no unique multivariate generalizations of univariate statistics. We choose which ones we want to use on the basis of exactly what it is we want to measure, and by convenience.

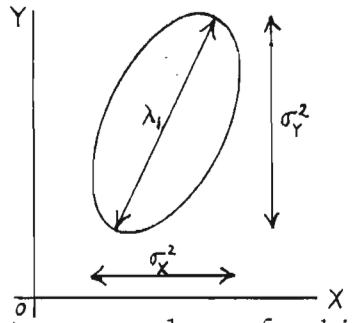

Figure 1. The ellipse represents an envelope of a bivariate distribution. The variances of X and Y are proportional to the squares of the lengths of their respective arrows; λ_1, the first eigenvalue, is also proportional to the square of the length of its arrow.

* * *

The usual scalar measure of variation in two or more dimensions called the generalized variance, $|\Sigma|$ for populations and $|S|$ for samples. It is the determinant of the covariance matrix. Geometrically, it is roughly proportional to the square of the area, volume, or hypervolume of any of the family of equi-probability envelopes of the referent distribution, just as the univariate variance is roughly proportional to the square of the length of the line that corresponds to any proportion of points in a univariate distribution. The square root of the generalized variance is an analog of the standard deviation.

In two dimensions,

$$|\Sigma| = \sigma_1^2\sigma_2^2(1-\rho^2), \tag{6}$$

where ρ is the correlation coefficient. In three dimensions,

$$|\Sigma| = \sigma_1^2\sigma_2^2\sigma_3^2(1 - \sum_{i>j}\rho_{ij}^2 + 2\prod_{i>j}\rho_{ij}). \tag{7}$$

The factor in parenthesis is the determinant of the correlation matrix, and the same general form holds for higher dimensions.

The variance of a distribution can be measured in any direction. If we measure it in the direction of the longest axis of the distribution (Figure 1) it is called the first eigenvalue. But it is still a variance. So are the other eigenvalues. They are the variances in other directions perpendicular to the first and to each other. There are as many eigenvalues as there are real dimensions, although not always as many as the axes of original measurement. A diagonal line in a plane has one real dimension and one eigenvalue but is located with respect to two original axes of measurement. An eigenvalue is conventionally represented by λ (not λ^2). The generalized variance can also be viewed as the product of the eigenvalues:

$$|\Sigma| = \prod\lambda_i \tag{8}$$

(Dempster, 1969, p. 137).

There are several serious problems with using $|\Sigma|$ to measure variation (Van Valen, 1974). For one thing, it vanishes when all the correlations are 1, and is small when most correlations are large. But in such cases there may still be appreciable real variation left on the line, or in the very elongate

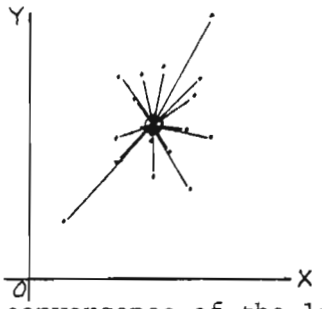

Figure 2. The circle at the convergence of the lines represents the bivariate mean of the distribution of dots. Each line is the distance from one point to the mean. These distances are squared in the calculation of the total variance.

* * *

dimension of a line-like distribution. The volume of a line is 0, and that constitutes the difficulty.

The effect of the variances is worse. $|\Sigma|$ also vanishes when any one or more of the variances is 0, and the effect of some variances being much smaller than others is even more severe than the effect of large correlations. There can still be large variation in all the other dimensions. The geometry of this difficulty is that the volume of a plane is 0, and in general the hypervolume in p dimensions of a distribution with only (p-1) dimensions is 0.

One may occasionally want to use the hypervolume of the distribution, but even then there is a problem, which is also perhaps the most severe one in using $|\Sigma|$ to measure variation. What, exactly, is the number of real dimensions, i.e. eigenvalues? In most real cases with more than three or four variables there is enough correlation among the variables that the last eigenvalues are near 0 and don't differ significantly from 0. If they are included, $|\Sigma|$ will be near 0 for the same geometrical reason as for the second difficulty. Moreover, small errors of estimation in small eigenvalues will cause proportionally large errors in $|\Sigma|$, especially when more than one eigenvalue is low. If we avoid this situation by including only those eigenvalues significantly different from 0, then the number we include will depend on the significance level we choose. Therefore our estimate of $|\Sigma|$ will also strongly depend on this quite arbitrary choice.

A better scalar measure of multivariate variation is the total variance, σ^2_p (Van Valen, 1974). It generalizes the univariate variance in a different way. The univariate variance is the sum of the squares of the distances of each point in the population (or sample) from the mean, divided by the number of points. Exactly the same wording can be applied to any number of dimensions, using Euclidean distances. Figure 2 illustrates the concept.

Algebraically, for p dimensions and n points,

$$\sigma^2_p = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{n}(x_{ij} - \mu_j)^2, \tag{9}$$

where μ_j is the population mean for variable j. The denominator should be $(n - 1)$ for samples, as in the univariate case. σ^2_p would then be called s^2_p. It happens that the total variance also equals the sum of the univariate variances:

$$\sigma^2_p = \sum \sigma^2_j. \tag{10}$$

This is quite a useful property, although a conceptually peripheral one. Because the sum of the variances in all independent directions must be the same whatever set of such directions is used, in particular

$$\sum \lambda_j = \sum \sigma_j^2 \qquad (11)$$

(Dempster, 1969, p. 137),
the total variance is also the sum of the eigenvalues. Use of the total variance requires that all variables be commensurable; e.g., lengths and angles can't be combined. This is an entirely natural stipulation for a single multivariate distribution.

The univariate standard deviation, the square root of the variance, measures variation in the scale of the variable itself. Similarly, the square root of the total variance does so for several variables jointly. This follows directly from the definition, by an exact analogy with the univariate case. σ_p is the total standard deviation. σ_p can alternatively be thought of as the average of the univariate standard deviations measured in all possible directions from the joint mean.

The covariance matrix is itself a measure of multivariate variation, although one less easy to comprehend as a whole. It measures variation in shape and direction as well as in size. Two or more covariance matrices will differ if the total variances differ, but they may differ even if the latter don't. I know of no robust (and therefore useful) tests in the latter situation.

Levene's test, Smith's test, and jackknifing are each applicable to tests on the total variance and on eigenvalues, again without requiring normal distributions.

For Levene's test, one can either first estimate the multivariate mean (mean vector) or estimate the deviation y_i of each point directly:

$$y_i = \sqrt{\sum_j (x_{ij} - \bar{x}_j)^2} \qquad (12)$$

for the jth dimension and the ith point. $\bar{y}$ is tested as before.

A test on eigenvalues does need the position of the joint mean. One must calculate the distance of each point from the joint mean. This distance should be parallel to the axis of the distribution corresponding to the eigenvalues being tested (Figure 3). Imagine that each point is collapsed perpendicularly onto the axis and then the distance y_i is calculated along the axis to the mean of the points, just as in the univariate Levene test (which this really is, just with the axis pointing in another direction.) Because eigenvalues are asymptotically independent, being measured in mutually perpendicular directions, the test can be used for eigenvalues of the same or different samples.

However, in samples the eigenvalues aren't strictly independent, because the first eigenvector defines the direction in which the sample, not necessarily the population, is most variable. This produces a small upwards bias for the first eigenvalue or so and a small downwards bias for the last. My simulations indicate that the bias isn't large, but some caution should be used when precise significance values are desired.

Smith's test is best considered for eigenvalues (λ_i) first. Calculate the distance y_i as for Levene's test. Then

$$s^2_{\lambda_j} = \frac{\sum y_i^4 - \lambda_j\left(\frac{n-3}{n}\right)}{(n-2)\quad(n-3)}. \qquad (13)$$

Because an eigenvalue is a variance, the same formula applies with appropriate changes in notation. Then replace s_j^2 in equation (2) by λ_j and solve. To prove the applicability of this and other tests to eigenvalues, we note that they are immediately applicable when the principal axes of the distribution are parallel

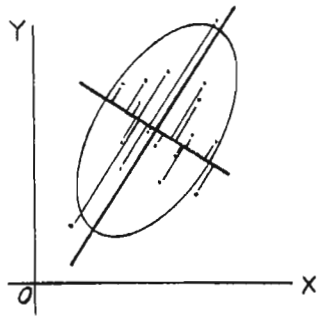

Figure 3. The distribution of Figure 2 now has an elliptical envelope, whose two axes are also represented. The distance of each point to the minor axis represents the distance of that point along the major axis. The distance shown is then squared in the calculation of the first eigenvalue. It is also the distance used as the deviation y_i in Levene's test for the first eigenvalue.

* * *

to the reference axes. The eigenvalues are then the variances of the original variables. Rotation of axes leaves the eigenvalues unchanged, so the tests are also unchanged and still apply.

Because the total variance $s_p^2 = \sum \lambda_j$ and all eigenvalues are approximately independent, the variance of estimation of s_p^2 is close to

$$s^2_{s_p^2} = \sum s^2_{\lambda_j}. \tag{14}$$

Again, s_p^2 for each sample replaces s_j^2 in Equation (2).

Jackknifing is conceptually the same as before but computation may be more complex than one can easily program.

Relative variation: univariate case

Is a mouse as variable as an elephant, in a biologically meaningful way? To see, use the coefficient of variation

$$CV = 100\,\frac{s}{\bar{x}}. \tag{15}$$

Haldane's small-sample correction (1955) is of no practical use because it is always much smaller than the expected error. The coefficient of variation should almost always be used in comparing variation when means differ.

There is an important assumption underlying use of the coefficient of variation, namely that absolute variation (s, not s^2) is proportional to the mean. This is so often true that we may tend to forget that there are cases where it isn't.

The only real class of exceptional cases I know are those in which variation starts from a more or less constant value above 0 (MacGillavry, 1965). These are cases where the structure is well canalized but can still vary a little. The number of presacral vertebrae in a population of a mammal is such a character. The number of lumbar vertebrae (presacrals posterior to the ribs) makes up most of this variation and the number of neck vertebrae is almost always constant. CV for total presacral vertebrae is less than that for lumbars, but the structural variation is pretty much the same in each case. Threshold characters have the same problem (Lande, 1977).

Despite defeatist statements still prevalent in the literature, there are more suitable ways to test relative variation than there are for absolute variation. Each of the three methods given above is appropriate directly for *CV*, and tests on a transformation are also often possible. The assumption that absolute variation is proportional to the mean is the same assumption one needs in order to use a logarithmic scale. Therefore the original data can be transformed to logarithms and the tests for absolute variation can be performed on the transformed data (Wright, 1952; Lewontin, 1966). One cannot, however, do such tests on the logarithm of statistics such as the variance. The data themselves must be transformed. The transformation is appropriate in both the univariate and the multivariate case. The scale of measurement (e.g. arithmetic or logarithmic) is important conceptually, although not with respect to the ease of making tests, and biological judgment is needed as to which scale is best.

However, logarithmic transformation is unnecessary for testing and can best be forgotten in this context except as a device to produce rather symmetrical distributions. To adapt Levene's test to *CV*, divide each deviation by the mean:

$$y_i = \frac{|x_i - \bar{x}|}{\bar{x}}. \tag{16}$$

Then proceed as before. For Smith's test, it is merely necessary to know the variance of estimate of *CV*:

$$s^2_{CV} \simeq \frac{CV^2}{2}, \tag{17}$$

when the distribution is roughly normal; in other cases Kendall and Stuart (1963, p. 233) give the appropriate value. Substitute *CV* for s^2 and s^2_{CV} for s^2_2 in Equation (2) and proceed as before. Similarly, jackknifing works by substituting the appropriate *CV* for s^2 in Equation (3).

Confidence intervals on *CV* can be found by jackknifing, as for s^2 or s. However, there is a simpler way (Miller and Kahn, 1962) when skewness is relatively low and (as usual) *CV* is less than about 30:

$$\frac{nCV^2}{\chi^2_{\frac{\alpha}{2}}(1+CV^2)-nCV} < \text{true } CV^2 < \frac{nCV^2}{\chi^2_{1-\frac{\alpha}{2}}(1+CV^2) - nCV^2}, \tag{18}$$

where α is the significance level (1 - confidence level) and χ^2 has $(n - 1)$ degrees of freedom. This method is of course also a direct test on whether the true *CV* has some specific value.

Relative variation: multivariate case

There is an immediate multivariate generalization of *CV* (Van Valen, 1974):

$$CV_p = 100 \frac{\sigma_p}{|\boldsymbol{\mu}|}, \tag{19}$$

where $|\boldsymbol{\mu}|$ is the mean vector (multivariate joint mean).

$$|\boldsymbol{\mu}| = \sqrt{\sum \mu_j^2}\ , \tag{20}$$

by Pythagoras's Theorem in p dimensions, so

$$CV_p = 100\sqrt{\frac{\sum \sigma_j^2}{\sum \mu_j^2}}\ . \tag{21}$$

Like the univariate CV, CV_p is dimensionless in the sense that multiplying all measurements by the same constant leaves CV_p unchanged. CV_p is also independent of the number of variables and so is numerically comparable to CV.

Tests on CV_p are analogous to those on the univariate CV. For Levene's test the deviation of each point i from the mean vector must be divided by the mean vector to produce y_i. Smith's test is identical to the univariate case. And jackknifing CV_p adds no new complexity to that for s_p^2.

Dimensionality of variation

As Lande (1977) has shown most clearly, the univariate CV gives too high values when applied to parameters such as volume which have dimensionality greater than 1. An adequate correction for this effect can't be derived from Lande's treatment because the latter is proportional to the correlations, yet even with perfect correlation CV for a volume can be almost the same as for one of its sides (as when a cylinder varies almost only in height). What is needed here is a measure of the effective dimensionality of variation. The equivalent number of independent variables, or amount of nonredundant information (Van Valen, 1974), is in one sense the true dimensionality of the shape of a distribution, being 1 for a line in any direction, near 1 for a line-like distribution, 2 for a circular disc, and 3 for a sphere. It measures precisely the dimensionality of variation. If variation of a volume is all, or mostly, in one dimension, the true variation is more nearly 1-dimensional than 3. Biological variation can be greater in three dimensions than it can in one, so we should have a way to adjust for differences in real dimensionality of variation.

In most cases, the dimensionality of variation D in the two-dimensional case is

$$D = (2-r^2), \tag{22}$$

where r is the correlation coefficient. In the p-dimensional case,

$$D = 1 + (p-1)(1-r_p^2), \tag{23}$$

where r_p^2 is the total correlation.

$$r_p^2 = \frac{1}{p}\sum_{i=1}^{p} R_i^2, \tag{24}$$

R_i being the multiple correlation coefficient. In the three-dimensional case,

$$R^2_{i \cdot jk} = 1 - (1-r^2_{ij})(1-r^2_{ik \cdot j}). \quad (25)$$

When the measured variables differ appreciably in variance or other self-information, estimation of D must proceed sequentially (Van Valen, 1974, p. 242).

Once one has D, then dividing the directly estimated CV (for the multidimensional measure) by D gives a new CV which is numerically comparable to a univariate CV or to CV_p. The same procedure can be applied to the standard deviation. However, estimating D in the case of something like weight may sometimes be difficult or impossible. In such a case, unless there is approximate similarity in shape or density (in which case D is close to 3), no useful comparisons can be made using CV.

When the dimensions are perfectly correlated (no shape difference), variation in volume or area is greater than when they are uncorrelated but with the same variation as before. (The approximations of Lande (1977) fail in the latter case, giving a CV of 0.) s^2_p is the same in each case. I interpret this effect as a biologically real aspect of variation rather than as an artifact of dimensionality, even though it can't occur in the one-dimensional case. It therefore needs no correction for dimensionality.

Usually the dimensions of variation will not be integers. Fractional dimensions are perhaps unfamiliar, but they occur naturally in other contexts, such as the dimensionality of a path of Brownian motion (Mandelbrot, 1977). This is something totally different from ontological imprecision, or fuzziness, which can also be measured precisely (Van Valen, 1964; Zadeh, 1965).

Acknowledgments

I thank F. L. Bookstein, M. Friedman, and R. Lande for comments.

Literature cited

Arveson, J.N., and T.H. Schmitz. 1970. Robust procedures for variance component problems using the jackknife. Biometrics 26:677-686.

Bissell, A.F., and R.A. Ferguson. 1975. The jackknife - toy, tool or two-edged weapon? The Statistician 24:79-100.

Box, G.E.P. 1953. Non-normality and tests on variances. Biometrika 40: 318-335.

Brown, M.B., and A.B. Forsythe. 1974. Robust tests for the equality of variances. Journal of the American Statistical Association 69:364-367.

Dempster, A.P. 1969. Elements of Continuous Multivariate Analysis. Reading, Massachusetts: Addison-Wesley. 388 pp.

Grüneberg, H., G.S. Bains, R.J. Berry, L. Riles, C.A.B. Smith, and R.A. Weiss. 1966. A Search for Genetic Effects of High Natural Radioactivity in South India. London: Her Majesty's Stationery Office. 59 pp.

Haldane, J.B.S. 1955. The measurement of variation. Evolution 9:484.

Kendall, M. 1975. Multivariate Analysis. London: Charles Griffin, 210 pp.

_____ and A.G. Stuart. 1963. The Advanced Theory of Statistics. Volume 1, edition 2. London: Charles Griffin. 433 pp.

Lande, R. 1977. On comparing coefficients of variation. Systematic Zoology 26:214-217.

Levene, H. 1960. Robust tests for equality of variances. In Contributions to Probability and Statistics (I. Olkin, S.G. Ghurye, W. Hoeffding, W.G. Madow, and H.B. Mann, eds.), pp. 278-292. Stanford: Stanford University Press.

Lewontin, R.C. 1966. On the measurement of relative variability. Systematic Zoology 15:141-142.

MacGillavry, H.J. 1965. Variability of larger Foraminifera. Part I: Natural position of zero values. Koninklijke Nederlandse Akademie van Wetenschappen (Amsterdam). Proceedings, Series B, vol. 68:335-355.

Mandelbrot, B.B. 1977. Fractals. San Francisco: W.H. Freeman. 365 pp.

Miller, R.G. 1974. The jackknife - a review. Biometrika 61:1-15.

Miller, R.L., and J.S. Kahn. 1962. Statistical Analysis in the Geological Sciences. New York: John Wiley and Sons. 483 pp.

Pearson, E.S., and N.W. Please. 1975. Relation between the shape of population distribution and the robustness of four simple test statistics. Biometrika 62:223-241.

Van Valen, L. 1964. An analysis of some taxonomic concepts. In Form and Strategy in Science (J.R. Gregg and F.T.C. Harris, eds.), pp. 402-415. Dordrecht: D. Reidel Publishing Co.

_____. 1974. Multivariate structural statistics in natural history. Journal of Theoretical Biology 45:235-247.

Wright, S. 1952. The genetics of quantitative variability. In Quantitative Inheritance (E.C.R. Reeve and C. Waddington, eds.), pp. 5-41. London: Her Majesty's Stationery Office.

Zadeh, L.A. 1965. Fuzzy sets. Information and Control 8:338-353.