On the Measurement of Relative Variability

R. C. LEWONTIN

Systematists and quantitative biologists in general are often interested in something they call "relative variation" or "intrinsic variation" of some character. What is meant by this is explained by Simpson, Roe and Lewontin as follows:

> The fact that elephants may have a standard deviation of 50 mm for some linear dimension and shrews a standard deviation of .5 mm for the same dimension does not necessarily mean that the elephants are more variable, in the essential zoological sense, than shrews. The elephants are one hundred times the size of shrews in any case, and we should expect the absolute variation also to be about a hundred times as great without any essential difference in functional variability.

They then go on to discuss the usual solution to this problem, the coefficient of variability, V, which is the standard deviation divided by the mean (and then usually multiplied by 100). This measure of relative variability is used by nearly all biologists, but it suffers from a grave disadvantage. There is little use in having an estimate of a quantity unless one can ask questions about it. In particular we want to know whether elephants and shrews have the same inherent variability. That is, we want to have a statistical test that will ask whether the true V for shrews is the same as the true V for elephants. The trouble is that there is no statistical test for comparing V's from different samples except when these samples are very large, much larger than is usually the case.

The purpose of this note is to point out a very simple property of variance that produces a simple way to measure relative variability and at the same time to make an exact statistical test.

The use of the coefficient of variation is based on an underlying assumption that two variables, X and Y, are really identical except that Y is K times as large as X. That is, the distribution of Y values can be exactly reproduced from the distribution of X values by multiplying every X by K. Now it is a well-known result in statistics that

if $\quad Y = KX \qquad (1)$

then $\quad \bar{Y} = K\bar{X} \qquad (2)$

and $\quad s_Y^2 = K^2 s_X^2 \quad$ or $\quad s_Y = K s_X \qquad (3)$

where $\bar{X}$, s_X^2, and s_X are the mean, variance and standard deviation of X. When these relations hold, it is easy to see why the coefficient of variation

$$V = \frac{100 \times \text{mean}}{\text{standard deviation}}$$

measures intrinsic variation since, from 2 and 3 above,

$$V_Y = \frac{100\bar{Y}}{s_Y} = \frac{100K\bar{X}}{Ks_X} = \frac{100\bar{X}}{s_X} = V_X .$$

Since it is the relationships 1, 2 and 3 above that make the coefficient of variation work, we can take advantage of them to construct a different method of estimating relative variability. This is done by *taking logarithms* (to any base) *of the measurements.* If $Y = KX$, then

$$\log Y = \log K + \log X .$$

Now since $\log K$ is a constant, and therefore has no variance

$$s^2_{\log Y} = s^2_{\log X} \qquad \text{and} \qquad s_{\log Y} = s_{\log X} .$$

We then see that *the variance (or standard deviation) of the logarithms of measurements gives a measure of intrinsic variability which is invariant under a multiplicative change.* One useful concomitant of this fact is that it does not matter in what units the variable is measured. The variance of the logarithms is the same whether the measurement is expressed in feet, inches, millimeters or cubits.

The advantage of the variance or standard deviation of the logarithms over the coefficient of variation is that all the usual statistical tests can be performed. For example, to test the question whether X and Y really have the same intrinsic variation we simply look up the ratio $s^2_{\log X}/s^2_{\log Y}$ in the table of the F distribution, with $N_X - 1$ and $N_Y - 1$ degrees of freedom. To test whether the intrinsic variability of X is equal to some *a priori* value σ^2 we look up the ratio $(N_X - 1)s^2_{\log X}/\sigma^2$ in the table of the chi-square distribution, with $N_X - 1$ degrees of freedom.

This property of logarithms has been pointed out at least once before, by Wright (1952), but the only record known to me of its having been used in systematic zoology is Bader and Lehman (1965). In his article, Wright points out that the variance of logarithms can be approximated by the logarithms of $(1 + V^2)$. For coefficients of variation less than 30% or so, an even closer approximation is that the variance of logarithms (to the base e) is equal to the squared coefficient of variation. This latter approximation is particularly handy since the simple coefficient of variation can be used as a descriptive measure of relative variability, and its square can be used in the F test described above.

REFERENCES

Bader, R. S., and W. H. Lehman. 1965. Phenotypic and genotypic variation in odontometric traits of the house mouse. Amer. Midland Nat. 74 (1):28–38.

Simpson, G. G., A. Roe and R. C. Lewontin. 1960. Quantitative Zoology. Harcourt, Brace and Co., Inc., New York.

Wright, S. 1952. The genetics of quantitative variability. *In* E. C. R. Reeve and C. Waddington [eds.], Quantitative inheritance, p. 5–41. H.M.S.O., London.

RICHARD C. LEWONTIN is Professor of Zoology at the University of Chicago, Chicago, Illinois 60637.