

finding the correlation between V and $U - V$ where $U = \log X$, $V = \log Y$. This is given by $\rho_{V, U-V} = (\rho\sigma_U - \sigma_V) / \{\sigma_U^2 + \sigma_V^2 - 2\rho\sigma_U\sigma_V\}$ where ρ = correlation of U , V and σ_U^2, σ_V^2 refer to their variances. When $\rho = 0$ this reduces to $-\sigma_V / (\sigma_U^2 + \sigma_V^2)$. Why did Atchley et al. take such trouble (and so much computer time) establishing results about X and Y empirically when it can be done analytically in terms of the logarithms? Why is the correlation between X and Y more meaningful than that between $\log X$ and $\log Y$?

Atchley et al. made great play out of the fact that $\rho_{V, X/Y}$ is large and negative if X and Y are independent (the same is true in terms of logs). Yet biologists use ratios for measurements with high positive correlations, not zero correlations, this being the whole point of using ratios. If ρ is positive and high and if $\sigma_U = \sigma_V$ (usually true on a log scale) then $\rho_{V, U-V}$ will be low, for then

$$\rho_{V, U-V} = (\rho - 1) / \{2 - 2\rho\} \\ = \{(1 - \rho) / 2\}.$$

Thus for $\rho = 0.8$, $\rho_{V, U-V} \approx -0.32$ whereas for $\rho = 0$, $\rho_{V, U-V} \approx -0.71$. Note that for $\rho = -0.8$, $\rho_{V, U-V} \approx -0.95$!

3. A function of a set of linear measurements X_1, \dots, X_p can only be regarded as measuring shape if it is invariant when all of the X 's are changed in the same proportion. This leads one to ratios or to some extension of the same idea based on logarithms (e.g., Penrose's [1954] size and shape measures based on the logarithms of the variables). It is a common fallacy to assume that this condition of invariance implies that shape must be

statistically independent of size—this is not so. The condition is a mathematical one referring to a single set of measurements.

4. If a principal component analysis is based on variables U_1, \dots, U_p (which are the logarithms of X_1, \dots, X_p), then the result will certainly differ from a principal component analysis based on $U_1 - U_2, U_2 - U_3, \dots, U_{p-1} - U_p$, or any other linear transformation of the U 's. The difference is quite predictable and it can make more sense to use a linear transformation rather than the U 's themselves. For example, with two variables it is often useful to transform to $U_1 + U_2$ and $U_1 - U_2$, i.e., to use $\log(X_1 X_2)$ and $\log(X_1/X_2)$. Since the correlation between these will be low (provided U_1 and U_2 have high positive correlation) it will often be possible to analyze them separately. Penrose's method of size and shape (on a log scale) effectively uses an orthogonal linear transformation in which $(U_1 + \dots + U_p) / p$ is the first transformed variable. Shape differences are then based on the other transformed variables.

REFERENCES

- ATCHLEY, W. R., C. T. GASKINS, AND D. ANDERSON. 1976. Statistical properties of ratios I. Empirical results. *Syst. Zool.* 25:137-148.
PENROSE, L. S. 1954. Distance size and shape. *Ann. Eugenics* 18:337-343.

MICHAEL HILLS

*Biometrics Section
British Museum (Natural History)
Cromwell Road
London SW7 5BD, England*

On the Use of Ratios in Growth Studies

Ratios have a long history of use among biologists, and the practice is unlikely to diminish markedly in the foreseeable future. There has been, however, recent recognition of certain theoretical (Atchley, et al., 1976) and practical (Blackith and Reymont, 1971) liabilities that may de-

tract from the general usefulness of ratios. The purpose of this contribution is to draw attention to several beneficial properties of ratios discovered in bivariate and multivariate studies of growth series of living and fossil reptiles (Dodson 1975a, b, c; 1976). A common feature

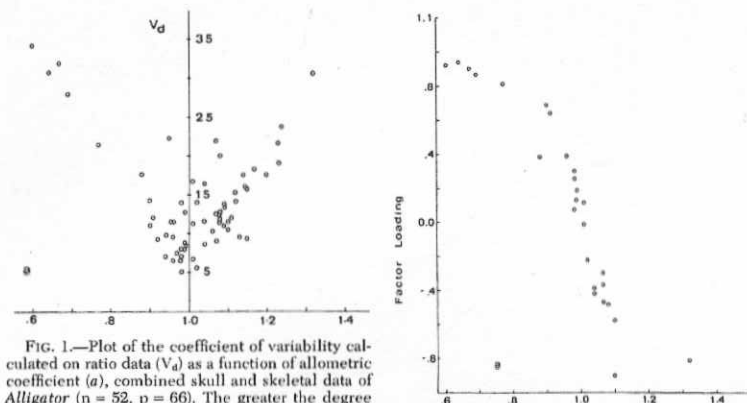


FIG. 1.—Plot of the coefficient of variability calculated on ratio data (V_d) as a function of allometric coefficient (a), combined skull and skeletal data of *Alligator* ($n = 52$, $p = 66$). The greater the degree of allometry, the greater the coefficient so calculated.

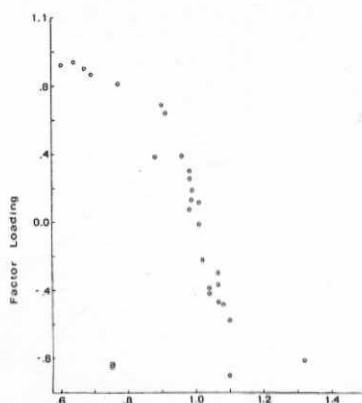


FIG. 2.—Plot of factor loadings of the first principal component as a function of the allometric coefficient (a); ratio data of *Alligator* skull ($n = 52$; $p = 27$).

of these studies is the great size span of the specimens, ranging from a factor of three in the *Sceloporus* study (Dodson, 1975b) to a factor of 18 in the *Alligator* study (Dodson, 1975a). In each of the examples discussed below, ratios were used only after equivalent procedures had been executed with the raw data. In this manner the effects of ratios could be carefully judged.

The first case involves calculation of the coefficient of variation (V) of a growth series. In his classic biometric study of the Permian amphibian, *Diplocaulus*, Olson (1951) discovered that calculation of V for a growth series gave meaningless results; the inflated values of V clearly did not reflect typical within-population variability. When he converted his data to ratios, values of V remained unacceptably high. Allometry is the key to this puzzle: ratios reduced specimens to unit length but do not eliminate changes of shape. If no change of shape (isometry) obtains during growth, size is eliminated by ratios and V does approximate the required measure of variability. But the greater the degree of allometry, the more V is inflated by the effect of size alone.

A plot of V as a function of the allometric coefficient (a) yields a V-shaped plot with minimum values (the sought-after measure of variability) centered about the isometric coefficient of 1.0 (Fig. 1). The tight envelope of points on the ascending limbs of the plot represent characters with similar levels of variability; only those points which lie above the envelope are of genuinely higher variability. This provides in effect a criterion of subtraction for the allometric effect. Because it facilitates work with growth series rather than just static adult samples, the coefficient calculated with ratio data has been named the coefficient of dynamic variability, V_d (Dodson, 1975a, b).

A frequently found feature in principal components analyses of growth series is that the first principal component accounts for a very high proportion of the total variance, leaving a small and sometimes statistically insignificant residual to interpret. When the first component has positive loadings of approximately equal value, it is usually interpreted as the size component (Blackith and Reymont, 1971).

TABLE 1. FIRST PRINCIPAL COMPONENT AND ALLOMETRIC COEFFICIENTS OF *Alligator* SKULL (n = 52, p = 27).

Variable	Factor loading (Raw data)	Factor loading (Ratio data)	Allometric coefficient
1	.99	-.47	1.07
2	.99	-.15	1.01
3	.99	.30	.98
4	.99	.13	.99
5	.99	.39	.96
6	.99	—	—
7	.99	-.91	1.10
8	.99	-.25	1.02
9	.98	.87	.69
10	.98	.93	.60
11	.99	-.37	1.07
12	.99	.11	1.01
13	.99	.64	.91
14	.99	.91	.67
15	.97	.38	.88
16	.99	.69	.90
17	.99	-.81	1.32
18	.99	.08	.98
19	.96	.82	.77
20	.99	-.58	1.10
21	.99	.19	.99
22	.98	-.30	1.07
23	.99	-.39	1.04
24	.99	-.42	1.04
25	.99	-.48	1.08
26	.98	.26	.98
27	.96	.94	.64

The size component in the study of *Alligator* skulls (number of specimens [n] 52; number of variables [p] 27; Dodson, 1975a) accounts for 97.9% of the total variance (the next three components account for 0.8%, 0.6% and 0.3% respectively). When ratios are used instead of raw data, the value of the first component plummets to 32.0% (while the next three components increase to 19.6%, 11.6% and 8.9% respectively). The first component ceases to be a size component in the usual sense because negative signs crop out and the factor loadings assume unequal value. However, size is reconstituted in an interesting way, for now the first component contrasts positively allometric with negatively allometric variables (Table 1). That it does so in an orderly fashion is shown by a plot of factor loadings as a function of allometric coefficients (Fig. 2).

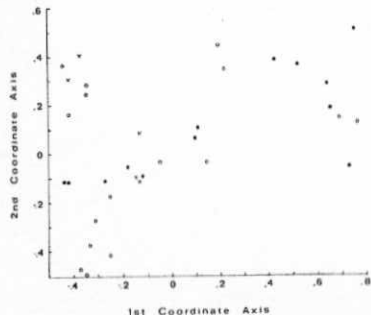


FIG. 3.—Principal coordinates analysis (p = 24) (raw data) of growth series of skulls of *Alligator* (n = 17) (O); *Crocodilus* (n = 13) (●); and *Gavialis* (n = 5) (x).

Another example of the utility of ratios in a multivariate study stems from a principal coordinates analysis (p = 24) with growth series of skulls from three species of crocodylians: *Alligator mississippiensis* (n = 17); *Crocodilus acutus* (n = 13); and *Gavialis gangeticus* (n = 5). Although these crocodylians differ so broadly in their morphology that they are immediately distinguishable one from another, the principal coordinates analysis of the raw data failed to bring about any significant separation among the three (Fig. 3). However, when ratio data are substituted, an excellent separation occurs (Fig. 4). The reason for this is that the variance due to size within the *Alligator* series is so great that juveniles manifest a greater metric resemblance to juvenile *Crocodilus* than they do to adults of their own kind. Boyce (1964) experienced similar difficulty in aligning skulls of juvenile pongids with their respective adults. The use of ratios reduces the within-group variance to a sufficiently low level that between-group variance predominates, resulting in the required separation.

DISCUSSION

Ratios are useful in studies of growth series and in other biometrical applica-

tions. In view of certain undesirable properties, however, the continued use of ratios has been questioned. Atchley et al. (1976) presented an important critique of ratios that must be considered carefully. Their study, which they described as "empirical," was in fact based on analysis of computer-generated pseudo-random vectors from a parametric multivariate normal distribution, the properties of which simply are not relevant to biologically meaningful tests of ratios. They concluded, after executing many hours of computer runs, that ratios shift the distribution of normal variables towards a right-skewed leptokurtic configuration; that ratios do not remove the effect of size from the data set; that spurious correlation among variables may be induced where none existed before; that ratios inflate the overall dependency of the correlation matrix and inflate the eigenvalue of the first principal component; and that in multivariate studies interpretations based on ratio data differ from those based on raw data.

The comments that follow derive from experience based on real biological data sets, in which variance due to size was a problem; that is, results of procedures executed on the raw data were unsatisfactory, requiring further manipulation of data to obtain comprehensible results. While other procedures could have been chosen, ratios were carefully tested and proved to be useful.

That ratios change the shape of distributions is significant, but how important this is to the practicing biologist is uncertain. Tests for normality are rarely carried out as a routine procedure; indeed the relevance of the normal distribution to the empirical world with which the biologist deals on a daily basis is doubtful. Reyment (1971) demonstrated that distributions which pass tests of bivariate normality may fail tests of multivariate normality, and that a transformation that has the effect of eliminating kurtosis may simultaneously increase skewness. Reyment (1971: 367) concluded: "Although the desire to make a set of data multivar-

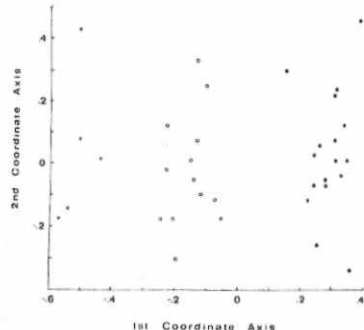


FIG. 4.—Principal coordinates analysis (p = 24) (ratios) of growth series of skulls of *Alligator* (n = 17) (O); *Crocodilus* (n = 13) (●); and *Gavialis* (n = 5) (x).

iate normal by means of some transformation or other may make good sense from the purely statistical point of view, there is some doubt as to the biologic rationale of such a step, as there can be no hiding the fact that this may tend to obscure the relationships the investigator wishes to probe." Fortunately, many statistical procedures (including generalized distance) are robust to departures from normality. It is the practical matter of robustness (as of yet inadequately explored), rather than theoretical niceties of distribution, with which the biologist is properly concerned.

That ratios do not remove the effect of size from the data set is so widely appreciated (e.g., Huxley, 1932; Gould, 1966, 1971; Corruccini, 1975, etc.) as to deserve no further discussion. The issue of spurious correlation is a more challenging point. However, the situation of zero correlation between numerator and denominator variables to which the authors alluded is biologically irrelevant for continuous characters in growth studies; because all characters increase in absolute size, they are all positively intercorrelated (in the study of *Alligator*, the lowest correlation coefficient in the 27 × 27

TABLE 2. SAMPLE CORRELATION MATRICES OF *Alligator* SKULL.¹

Log raw data					
VAR 1	1.000				
2	.997	1.000			
3	.996	.999	1.000		
4	.994	.995	.996	1.000	
5	.988	.998	.993	.998	1.000
	VAR 1	2	3	4	5
Log ratios					
VAR 1	1.000				
2	.592	1.000			
3	.411	.797	1.000		
4	.315	.398	.568	1.000	
5	.050	.154	.508	.842	1.000
	VAR 1	2	3	4	5

¹Note the sharp reduction of levels of correlation when ratios are used. (Basal skull length [var. 6] is the standard variable.)

matrix for cranial raw data was .890). When the effect of size is great, ratios can decrease the general level of correlation (Table 2); indeed, negative correlations may appear, indicating that while all structures increase in *absolute* size, some structures become *relatively* larger while others become *relatively* smaller.

Thus, using ratio data for *Alligator*, for the relationship snout length ($a = 1.10$) versus orbit length ($a = .69$) $r = -.804$; and for width of the upper temporal opening ($a = 1.32$) versus width of the foramen magnum ($a = .64$), $r = -.629$. The unconditional assertion by Atchley et al. (1976) concerning inflation of correlation by ratios, in spite of indisputable evidence to the contrary (Corruccini's [1975] statement on the subject was labeled "completely false"), indicates that the authors fail to appreciate intuitively what the taking of ratios means: literally the reduction of all specimens to the same unit length. Size is not eliminated, because allometry remains. The desirable effect achieved by ratios is the emphasis on the growth of structures relative to one another, of which the general lowering of correlations and the appearance of negative correlations is an expression.

Atchley et al. (1976) correctly established that in a multivariate analysis the use of ratios may lead to a substantially different interpretation from one based

on raw data. This insight is not novel; the factor analyst has at his disposal a wide range of options as to types of rotation and numbers of axes to abstract, each of which yields a unique result. The issue is not which solution is true and which others are false, but rather which provide useful biological insight (Gould, 1967). In the principal components analysis discussed above (Table 1; Fig. 2), the first component of the ratio analysis clearly provided useful, independently verified information. Once again, usefulness is the sole criterion.

CONCLUSIONS

Ratios doubtlessly have been used unwisely in metric studies for many years. During the past decade our knowledge of the significance of ratios, as well as their liabilities, has improved considerably. Atchley et al. (1976) have made an important contribution on the theoretical properties of ratios, but precisely because of their lack of empiricism, they failed to demonstrate the applicability of their findings to the real biological world. They unfortunately chose to base their analyses on pseudo-random vectors from a multivariate normal, random distribution, and to assume non-correlation between numerator and denominator variables in ratios, situations most unlikely to be encountered in nature. I demonstrate from studies of real data sets that ratios have desirable properties in both bivariate and multivariate studies of continuously distributed growth data. In each of the studies described, ratios were used advisedly, only after equivalent operations with raw data had been executed; and similar caution is to be recommended to all investigators. The desirable properties of ratios reported herein occurred in growth series in which the largest members were between 3 and 18 times greater than the smallest. As the beneficial effects of ratios were more evident at the upper end of the spectrum than at the lower, there probably exists a limit (yet to be determined—probably in the range of 1.5 to 2.0) below which the deleterious

effects outlined by Atchley et al. outweigh the benefits presented herein. I believe, however, that careful biometricians will decide that, to paraphrase Mark Twain, "rumors of demise of ratios are grossly exaggerated."

ACKNOWLEDGMENTS

I thank David Pilbeam for his help and encouragement in overcoming the initial "failure" of principal coordinates analysis to ordinate correctly the *Alligator-Crocodylus-Gavialis* series. I am also very grateful to Wayne W. Moss for his thorough reading and constructive criticism of an earlier draft. D. Wexlar drafted the figures.

REFERENCES

- ATCHLEY, W. R., C. T. GASKINS, AND D. ANDERSON. 1976. Statistical properties of ratios. I. Empirical results. *Syst. Zool.* 25:137-148.
- BLACKITH, R. E., AND R. A. REYMENT. 1971. Multivariate morphometrics. Academic Press, New York.
- BOYCE, A. J. 1964. The value of some methods of numerical taxonomy with reference to hominoid classification. In Heywood, V. H. and J. McNeill (eds.), *Phylogenetic and phenetic classification*. Syst. Assoc. London Publ. 6:47-65.
- CORRUCCINI, R. S. 1975. Multivariate analysis in biological anthropology: some considerations. *J. Hum. Evol.* 4:1-19.
- DODSON, P. 1975a. Functional and ecological sig-

- nificance of relative growth in *Alligator*. *J. Zool., London* 175:315-355.
- DODSON, P. 1975b. Relative growth in two sympatric species of *Sceloporus*. *Amer. Midl. Nat.* 94:421-450.
- DODSON, P. 1975c. Taxonomic implications of relative growth in lambeosaurine hadrosaurs. *Syst. Zool.* 24:37-54.
- DODSON, P. 1976. Quantitative aspects of relative growth and sexual dimorphism in *Protoceratops*. *J. Paleontol.* 50:929-940.
- GOULD, S. J. 1966. Allometry and size in ontogeny and phylogeny. *Biol. Rev.* 41:587-640.
- GOULD, S. J. 1967. Evolutionary patterns in pelycosaurian reptiles: a factor-analytic study. *Evolution* 21:385-401.
- GOULD, S. J. 1971. Geometric similarity in allometric growth: a contribution to the problem of scaling in the evolution of size. *Amer. Nat.* 105:113-136.
- HUXLEY, J. S. 1932. *Problems of relative growth*. Dial Press, New York.
- OLSON, E. C. 1951. *Diplocaulus*—a study in growth and variation. *Fieldiana Geol.* 11:57-154.
- REYMENT, R. A. 1971. Multivariate normality in morphometric analysis. *Math. Geol.* 3:357-368.

PETER DODSON

Laboratories of Anatomy
School of Veterinary Medicine
University of Pennsylvania
Philadelphia, Pennsylvania 19104

Syst. Zool., 27(1), 1978, pp. 67-71

Some Comments on the Use of Ratios

Atchley et al. (1976) have recently called attention to some of the statistical consequences of using ratio data. Their empirical results should provide ample warning for biologists who might consider converting raw data to ratios. However, while Atchley et al. presented examples which are of heuristic value because of their relative simplicity, several of these same examples are potentially misleading to readers looking for correlations between the empirical findings and common usages of ratios in biology.

In discussing the utility of ratios for the control of size differences, Atchley et al. proposed the situation in which one variable (X_1) is converted to a ratio ($Y = X_1/$

X_2), where the denominator variable ($X_2 = Z$) is uncorrelated with X_1 ($\rho_{12} = 0$) but equal in terms of coefficients of variation ($\delta_1 = \delta_2$). They demonstrated that under such conditions the ratio Y is correlated with the variable Z ($r_{YZ} = -0.71$); apparently, in this situation, the use of ratios does not eliminate size from the raw variable X_1 (if X_2 is considered to be a vector which measures size). However, the original assumption of zero correlation ($\rho_{12} = 0$) indicates that X_1 , by itself, displays exclusively non-size variation; accordingly, there would be little point in scaling the data to correct for size. Of course, nothing prevents the careless researcher from overlooking the independence of the two variables, utilizing a ratio to cor-