

UNIVERSITY OF  
ILLINOIS LIBRARY  
AT URBANA-CHAMPAIGN  
NATURAL HIST. SURVEY





NH  
NATURAL HISTORY SURVEY

AUG 6 1985

LIBRARY

# FIELDIANA Zoology

Published by Field Museum of Natural History

---

New Series, No. 15

## THE EFFECTIVENESS OF METHODS OF SHAPE ANALYSIS

CLIFF A. LEMEN

March 31, 1983

Publication 1343



## THE EFFECTIVENESS OF METHODS OF SHAPE ANALYSIS





# FIELDIANA

## Zoology

Published by Field Museum of Natural History

---

New Series, No. 15

### THE EFFECTIVENESS OF METHODS OF SHAPE ANALYSIS

CLIFF A. LEMEN

*Research Associate  
Department of Zoology  
Field Museum of Natural History*

*School of Life Sciences  
University of Nebraska  
Lincoln, Nebraska 68588*

Accepted for publication September 24, 1981

March 31, 1983

Publication 1343

*Library of Congress Catalog Card No. : 82-83505*

ISSN 0015-0754

PRINTED IN THE UNITED STATES OF AMERICA

## CONTENTS

|                                      |    |
|--------------------------------------|----|
| Introduction .....                   | 1  |
| Methods of Data Transformation ..... | 2  |
| Methods of Shape Analysis .....      | 3  |
| Sizeout .....                        | 3  |
| Correlation .....                    | 5  |
| Standardization .....                | 7  |
| Log Transformation .....             | 9  |
| Ratios .....                         | 9  |
| Regression-Residual Analysis .....   | 10 |
| Log-Sizeout .....                    | 11 |
| Congruence of Techniques .....       | 13 |
| Summary .....                        | 16 |
| Acknowledgments .....                | 16 |
| Literature Cited .....               | 16 |

## LIST OF ILLUSTRATIONS

|                                                                                                                                                                           |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1. Mechanism of the sizeout analysis for a two-variable case .....                                                                                                        | 4  |
| 2. The morphological measurements of two OTUs, wolf and dog, plotted against one another to form scattergram .....                                                        | 7  |
| 3. Log-transformed data with parallel iso-OTU vectors .....                                                                                                               | 11 |
| 4. The shape coefficients of the different methods plotted against one another in scattergram to obtain visual representation on the congruence of these techniques ..... | 14 |
| 5. A phenogram of the methods of shape analysis made on the inter-method correlations of Table 6 .....                                                                    | 15 |

## LIST OF TABLES

|                                                                                                                                        |    |
|----------------------------------------------------------------------------------------------------------------------------------------|----|
| 1. The loadings of the morphological characters on the first principal component obtained by a PCA on the canid data set .....         | 4  |
| 2. Two shape coefficient matrices: distances between OTUs in sizeout space and inter-OTU correlation coefficients .....                | 6  |
| 3. Correlation coefficients among canids, using both standardized and log-transformed data .....                                       | 8  |
| 4. Two shape coefficient matrices: distances between OTUs in ratio space and inter-OTU distances, based on the regression method ..... | 12 |
| 5. Shape coefficients based on the distances between OTUs in log sizeout space .....                                                   | 12 |
| 6. The congruence of the methods of shape analysis discussed in this report obtained by using inter-method coefficients .....          | 13 |



## INTRODUCTION

A common goal in multivariate morphological studies is to compare the shapes of the organisms under study. On an intuitive level the distinction between size and shape is obvious. Most people could agree with the dictionary definitions of shape as the relative position of all points composing the outline or external surface of an object and size as the space an object occupies. But these two concepts can be difficult to separate in multivariate analyses of morphological data (Sneath & Sokal, 1973). The data sets from these analyses normally consist of linear measurements on a series of morphological characters such as skull length, tooth row, etc. These measurements are affected by both the shape and the size of an organism. The goal of shape analysis is to separate these two parts of the measurements so that shape comparisons can be made between organisms of different sizes.

This paper tests the effectiveness of several techniques that have been developed for shape analysis, investigates how they work, and shows why some methods fail, using an experimental approach similar to those followed by Moss (1968, 1971), Crovello (1969), Manischewitz (1973), Minkoff (1965), and Rohlf (1972). I am defining effectiveness as the ability to classify objects of the same shape as similar, regardless of the size of the objects (Moss, 1968; Mosimann, 1970). This is in contrast to other kinds of shape analyses which have the goal of documenting and quantifying consistent changes of shape with increasing size (Gould, 1966; Sweet, 1980; for a discussion of this sort of analysis in conjunction with allometric growth, or documenting the morphological distinctiveness of predetermined groups of organisms, see Albrecht, 1980).

The first problem in testing different methods of shape analysis is that the relationships of shape among real objects are unknown and can only be inferred (Corruccini, 1973); therefore, no criteria exist to evaluate methods of analyzing shape. To produce objects of known shape relationships, I first measured 11 morphological characters on eight species of canids. Then I made artificial operational taxonomic units (OTUs) with known shape relationships by scalar multiplication of the data from two canids, kit fox and wolf. For example, all 11 measurements of the wolf were multiplied by a constant of 0.53 to produce an iso-OTU about the size of a kit fox. On paper at least, I have produced two canids of different size but of the same shape. Similar scalar multiplications were performed to produce a mid-sized wolf, a wolf-size kit fox, and a mid-sized kit fox. Of course, isometric enlargement of OTUs is rare or nonexistent in real biological data, but my procedure has the strength that the shape relationships within the iso-wolves and the iso-kit foxes are known.

Several methods of shape analysis were applied to the canid data set, such as sizeout, correlation, ratio distance, regression analysis, and log-sizeout. My

research indicates that the best results are obtained from ratio distance, log-sizeout, and correlation of log-transformed data. The following section presents the description of each method and analysis of its strengths and weaknesses.

### METHODS OF DATA TRANSFORMATION

In general, neither the means nor the variances of characters are equal. Clearly, total head and body length measurements will have a larger mean, and probably a higher standard deviation, than a measurement taken on the molar tooth row. This produces a problem: characters with large means and/or variances contribute more to the determination of shape coefficients between OTUs than do characters with small means and/or variances (Sneath & Sokal, 1973). *A priori* there is no reason to weight these characters differently. A 10% change in molar tooth row is not necessarily less important than a 10% change in head and body length. However, because head and body length has a greater absolute measurement, it will tend to affect the shape coefficients more. Some methods must be found to equalize the effect of variables before a shape analysis is done. Two methods of data transformation are discussed here. One method is standardization (Sneath & Sokal, 1973). The standardized score of a character can be given as:  $\text{Score} = (x - \bar{x})/s_x$ , where  $\bar{x}$  is the mean of that character for the organisms included in the study and  $s_x$  its standard deviation. Thus, the standardized score is equivalent to the z-value used in statistics. After standardization, the means of all characters are 0.0 and the standard deviations 1.0. Variables then contribute equally to the analysis. Because of the subtraction of the mean and division by the standard deviation, the standardized score of a measurement is, in part, dependent on the other objects included in the study. The same organism could receive very different scores for a character in two analyses which contained different sets of organisms.

Some of the problems produced by standardization have already received attention. Hudson et al. (1966) and Sneath & Sokal (1973) pointed out that measurements with small variances may be heavily affected by simple measurement error. The magnification of these errors to equal status with the other characters through standardization would be a mistake. Also, Rohlf & Sokal (1965), Rohlf (1962), and Underwood (1969) found that standardization normally reduces the average correlation coefficient between OTUs. Finally, Sneath & Sokal (1973) note that standardization reduces the atypicality of aberrant OTUs. Manischewitz (1973) tested standardization with several other methods of transformation and found it to be the most reliable method. However, as I will discuss below, other problems also affect the use of standardization in correlation analysis. These problems are so severe that the standardization of data in correlation analysis is impossible.

There are other methods available for transforming data; one of the most promising is log transformation. Logarithmic transformation has long been associated with normalizing the curve shape of right skewed data; however, as pointed out by Moriarty (1977) and Lewontin (1966), the effects of logarithmic transformation go beyond normalization. Perhaps most important is the property that, in log-transformed data, the standard deviation of a variable is proportional to the coefficient of variation (CV) of the untransformed variable. Thus, after transformation, each variable will tend to contribute to the analysis in proportion to its CV. Because the mean of a variable may, in part, determine

its relative contribution to the analysis, the mean of the transformed data can be subtracted from the log-transformed data. Other properties of log-transformed data are discussed in the section on log-sizeout shape analysis.

## METHODS OF SHAPE ANALYSIS

### SIZEOUT

In morphometric data there is normally a high intercharacter correlation with size; large animals tend to have large tails, skulls, feet, etc. Thus, when viewing a character correlation matrix from morphometric data, one usually notices a great many high positive values, especially between size-related characters. In a principal components analysis (PCA), the first principal component is defined as that vector in hyperspace which explains the maximum possible variation of the data. Thus, if most or all characters are highly size dependent, the first principal component will be highly correlated with size (Jolicoeur & Mosimann, 1960; Jolicoeur, 1963). This is a general phenomenon and has been observed repeatedly with morphometric data (Sneath & Sokal, 1973). Of course, it is not a rule, because size may be less important in some data matrices (i.e., the first principal component may not be highly correlated with size).

Once it is determined that the first principal component is a size factor, the effect of this axis can be removed mathematically from a data set (Rohlf et al., 1971). Figure 1 may help the reader to visualize this process. Two characters are measured on a series of canids, and a principal components analysis is performed on the data. As can be seen in the figure, the first principal component is highly size related. If this axis is eliminated, only the second principal component is left, and distances between OTUs on this second vector represent "sizeout" distances. Multivariate analyses use more than two characters, but the method remains the same: the first principal component is removed and distances are calculated between OTUs based on all the remaining components. These sizeout distances can be interpreted as indicators of shape similarity; i.e., the smaller the sizeout distance between OTUs the greater the shape similarity.

Sizeout analysis depends on the first principal component being the size factor, but exactly what does this mean? The first principal component is a composite of all the original variables, and those characters which are size related are generally more important in determining the first principal component. An example of these loadings is given in Table 1. Notice that, although all loadings are high, indicating that the first principal component is highly size related, all characters are not equally high. The smallest loading is for cranial width; this means that, for this group of animals, cranial width is least related to the linear "size factor" (the first principal component). An important point to remember is that the loadings of these characters are determined by the shapes of the animals included in the study. With another group of animals of different shapes, cranial width might be more highly correlated with the first principal component.

The loadings for skull length and cranial width are 0.983 and 0.913, respectively. A graph can be made by plotting these two numbers against one another to define the sizeout vector for these two character axes (fig. 1). As already discussed, it is the collapsing of this first size factor that results in sizeout. Also plotted in Figure 1 are the OTUs on which the PCA was performed.

In Figure 1 the iso-OTUs are connected to obtain an iso-wolf and an iso-kit

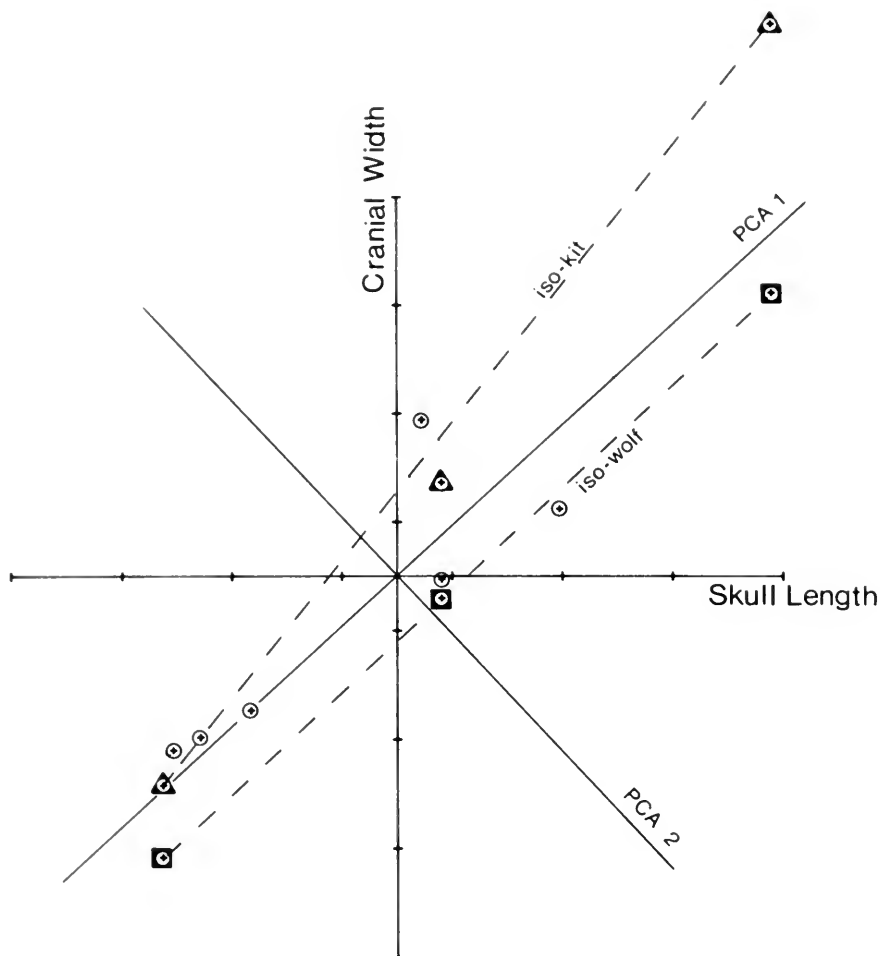


FIG. 1. Mechanism of the sizeout analysis for a two-variable case. Notice that the iso-wolf vector (marked iso-wolf) is parallel to the first principal component (PCA 1), but the iso-kit fox vector (iso-kit) is not. This leads to the incorrect shape relationships found among the iso-kit foxes.

TABLE 1. The loadings of the morphological characters on the first principal component obtained by a PCA on the canid data set.

|                          |       |
|--------------------------|-------|
| Skull length             | 0.983 |
| Tooth row length         | 0.929 |
| Molar tooth row length   | 0.986 |
| Width between tooth rows | 0.970 |
| Zygomatic width          | 0.982 |
| Nasal length             | 0.979 |
| Cranial width            | 0.913 |
| Dentary length           | 0.987 |
| Coronoid height          | 0.987 |
| Width between incisors   | 0.953 |
| Dentary thickness        | 0.937 |



fox line. Because the data have been standardized, all the kit foxes and wolves no longer lie on a straight line. For the purpose of this example, I have connected the largest and smallest iso-OTUs. The wolf line is more similar in slope to the sizeout vector than is the kit fox vector; because of the mathematics of sizeout analysis, the more similar the slope of an iso-OTU vector is to the sizeout vector, the more perfectly the analyses will remove the effect of size from data. This can be seen graphically if one imagines the collapse of the sizeout vector to the origin, leaving only the second principal component. All three wolves, because of the nearly parallel nature of the iso-wolf vector, collapse to nearly one point on the second vector; however, this is not true for the iso-kit foxes. This is only a two-dimensional example, but when all 11 characters are considered, the sizeout procedure is still less effective for the kit foxes than for the wolves. In fact, the distance between the largest and smallest iso-kit foxes is in the upper 46% of all distances in the sizeout distance matrix (table 2). This magnitude of error indicates that sizeout can produce significant changes in results. Iso-enlargements do not exist in real studies, but the problem remains. Sizeout cannot remove the effect of size from all groups equally, and the ability of sizeout to function properly will tend to decrease as animals of more diverse shapes are included in the analysis. In this example, the shape relationships of the iso-wolves were correctly calculated, whereas incorrect answers were obtained for the iso-kit foxes. In another study, which included more canids, the sizeout analysis also failed on wolves. Evidently, the first principal component was shifted and no longer parallel to the iso-wolf vector in the second analysis.

#### CORRELATION

Shape analysis using correlation (Michener & Sokal, 1957) involves calculating the inter-OTU product-moment correlation coefficient for all possible pairs of OTUs in a study. This coefficient is calculated by the formula:

$$r_{1,2} = \frac{\sum_{i=1}^n (X_{i,1} - \bar{X}_1)(X_{i,2} - \bar{X}_2)}{\sum_{i=1}^n (X_{i,1} - \bar{X}_1)^2 \sum_{i=1}^n (X_{i,2} - \bar{X}_2)^2}$$

where  $X_{i,1}$  is the measurement of the  $i$ th character on the first OTU,  $\bar{X}_1$  is the mean of all measurements on OTU<sub>1</sub>, and  $r_{1,2}$  is the product-moment correlation coefficient between OTU<sub>1</sub> and OTU<sub>2</sub>.

A scattergram (fig. 2) can give a graphical representation of this correlation for a pair of OTUs (dog and wolf). A high positive correlation between two OTUs may indicate great similarity in shape, whereas low values mean little similarity (Rohlf & Sokal, 1965).

One of the problems already pointed out by Rohlf & Sokal (1965) with the use of correlation analysis is that OTUs may be highly correlated ( $r = 1.0$ ) and not be of similar shape. Their example illustrates this point: the measurements 1-2-3 on object 1 would correlate perfectly with the measurements 1-2-3 on object 2 or 2-4-6 on object 3, and this is consistent with our ideas on shape relationship; however, object 4 with the measurements 101-102-103 would also be perfectly correlated with OTU 1, even though these two OTUs are not the same shape. Although this problem has been noted, its real impact in an analysis is difficult to determine.

TABLE 2. Two shape coefficient matrices (because such matrices are symmetrical, only one-half of each is presented). In the lower portion of the matrix (below asterisks) are the distances between OTUs in sizeout space; in the upper portion of the matrix are the inter-OTU correlation coefficients, using standardized data.

|             | Wolf | Dog  | Coyote | Hunting dog | Red fox | Arctic fox | Gray fox | Kit fox | Small wolf | Mid kit | Large kit | Mid wolf |
|-------------|------|------|--------|-------------|---------|------------|----------|---------|------------|---------|-----------|----------|
| Wolf        | **** | 0.38 | 0.08   | -0.02       | -0.51   | -0.35      | -0.84    | -0.48   | 0.96       | -0.92   | -0.95     | 0.99     |
| Dog         | 0.27 | **** | 0.70   | -0.62       | -0.08   | -0.43      | -0.32    | -0.18   | 0.31       | -0.28   | -0.28     | 0.34     |
| Coyote      | 0.30 | 0.12 | ****   | -0.71       | 0.37    | -0.40      | 0.10     | -0.18   | -0.03      | -0.03   | 0.01      | 0.02     |
| Hunting dog | 0.51 | 0.55 | 0.53   | ****        | -0.46   | 0.17       | -0.05    | -0.26   | 0.05       | -0.23   | -0.20     | 0.02     |
| Red fox     | 0.37 | 0.23 | 0.16   | 0.52        | ****    | 0.39       | 0.62     | -0.08   | -0.66      | 0.50    | 0.58      | -0.59    |
| Arctic fox  | 0.30 | 0.22 | 0.20   | 0.42        | 0.16    | ****       | 0.33     | -0.03   | -0.31      | 0.18    | 0.21      | -0.33    |
| Gray fox    | 0.35 | 0.21 | 0.16   | 0.44        | 0.13    | 0.13       | ****     | 0.15    | -0.89      | 0.73    | 0.80      | -0.88    |
| Kit fox     | 0.31 | 0.19 | 0.17   | 0.45        | 0.19    | 0.14       | 0.13     | ****    | -0.30      | 0.68    | 0.55      | -0.38    |
| Small wolf  | 0.08 | 0.28 | 0.32   | 0.50        | 0.40    | 0.33       | 0.37     | 0.31    | ****       | -0.89   | -0.95     | 0.99     |
| Mid kit     | 0.49 | 0.32 | 0.28   | 0.53        | 0.21    | 0.24       | 0.18     | 0.20    | 0.50       | ****    | 0.99      | -0.92    |
| Large kit   | 0.72 | 0.53 | 0.48   | 0.68        | 0.39    | 0.45       | 0.38     | 0.43    | 0.73       | 0.23    | ****      | -0.96    |
| Mid wolf    | 0.04 | 0.27 | 0.31   | 0.50        | 0.39    | 0.31       | 0.36     | 0.31    | 0.04       | 0.50    | 0.73      | ****     |

The scientific names of the species used in this study are as follows: wolf, *Canis lupus*; dog, *C. familiaris*; coyote, *C. latrans*; African hunting dog, *Lycyon pictus*; arctic fox, *Alopex lagopus*; gray fox, *Urocyon cinereoargenteus*; kit fox, *Vulpes macrotis*; red fox, *V. fulva*; small wolf, mid wolf, mid kit fox, and large kit fox, all iso-OTUs of *C. lupus* and *V. macrotis*.

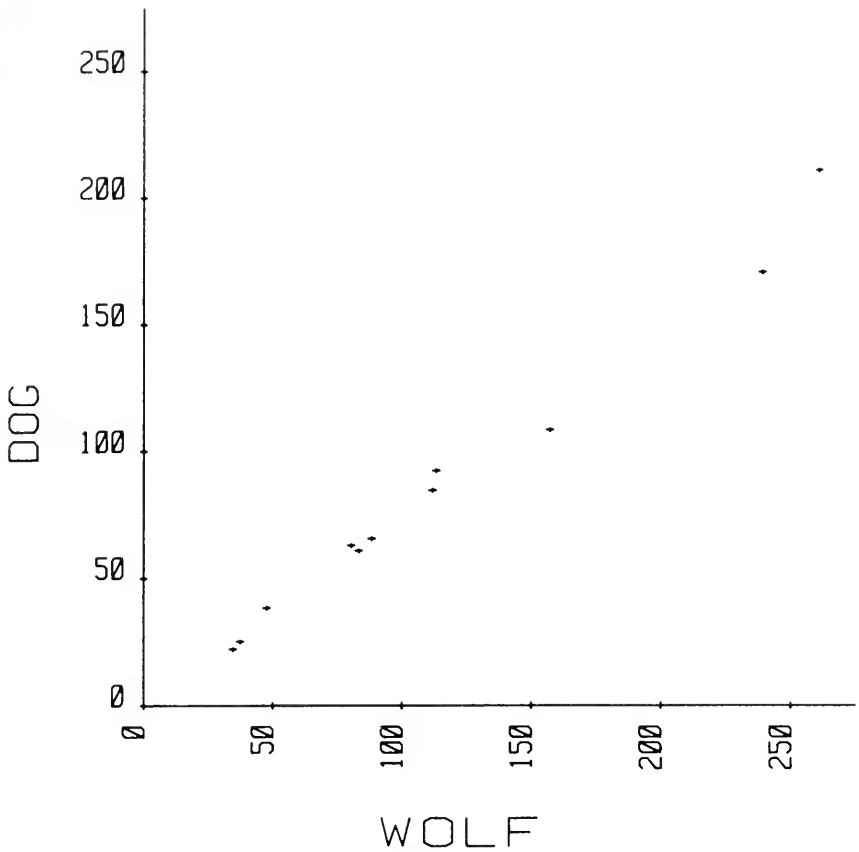


FIG. 2. The morphological measurements of two OTUs, wolf and dog, are plotted against one another to form this scattergram. The data have not been standardized; therefore, the correlation coefficient is high ( $r = 0.993$ ). Standardization reduces the  $r$ -value to 0.375.

The importance of a character in determining  $r$  is strongly affected by the mean value of the characters; for this reason, some data transformation is needed. The two methods discussed here are standardization and log transformation.

#### *Standardization*

Tables 2 and 3 give the results of correlation analysis of two separate standardized data sets, both of which include the iso-wolves and iso-kit foxes. It is clear from these tables that the results of these correlation analyses, which should be based on shape alone, give inconsistent results with known relationships of the iso-OTUs. The differences in results indicate two things: (1) size is evidently important in determining product-moment correlation coefficient because, in every case within an analysis, iso-OTUs which are more similar in size have larger  $r$ -values than do iso-OTUs of very different sizes

TABLE 3. Below the asterisks are the correlation coefficients on standardized data between the canids, but in this case, a larger group of canids was included in the standardization process (although these OTUs are not shown here). In the upper portion of the matrix are the inter-OTU correlation coefficients, using log-transformed data.

|             | Wolf  | Dog   | Coyote | Hunting dog | Red fox | Arctic fox | Gray fox | Kit fox | Small wolf | Mid kit | Large kit | Mid wolf |
|-------------|-------|-------|--------|-------------|---------|------------|----------|---------|------------|---------|-----------|----------|
| Wolf        | ****  | 0.68  | 0.33   | 0.11        | -0.75   | -0.68      | -0.94    | -0.88   | 1.00       | -0.88   | -0.88     | 1.00     |
| Dog         | 0.06  | ****  | 0.72   | -0.41       | -0.47   | -0.63      | -0.66    | -0.52   | 0.67       | -0.52   | -0.53     | 0.67     |
| Coyote      | -0.15 | 0.76  | ****   | -0.63       | 0.05    | -0.48      | -0.27    | -0.28   | 0.34       | -0.28   | -0.28     | 0.33     |
| Hunting dog | -0.12 | -0.54 | -0.49  | ****        | -0.44   | 0.13       | -0.09    | -0.34   | 0.11       | -0.34   | -0.34     | 0.11     |
| Red fox     | -0.31 | 0.33  | 0.66   | -0.24       | ****    | 0.60       | 0.77     | 0.60    | -0.75      | 0.60    | 0.60      | -0.75    |
| Arctic fox  | -0.55 | -0.16 | -0.02  | 0.20        | 0.61    | ****       | 0.62     | 0.42    | -0.68      | 0.42    | 0.42      | -0.68    |
| Gray fox    | -0.71 | 0.16  | 0.51   | 0.09        | 0.76    | 0.61       | ****     | 0.75    | -0.94      | 0.75    | 0.75      | -0.94    |
| Kit fox     | -0.71 | -0.03 | 0.09   | -0.11       | 0.24    | 0.30       | 0.45     | ****    | -0.88      | 1.00    | 1.00      | -0.88    |
| Small wolf  | 0.72  | -0.27 | -0.51  | -0.14       | -0.73   | -0.53      | -0.89    | -0.51   | ****       | -0.88   | -0.88     | 1.00     |
| Mid kit     | -0.73 | 0.23  | 0.45   | 0.03        | 0.67    | 0.47       | 0.81     | 0.77    | -0.93      | ****    | 1.00      | -0.88    |
| Large kit   | -0.70 | 0.27  | 0.49   | 0.06        | 0.71    | 0.47       | 0.84     | 0.68    | -0.96      | 0.99    | ****      | -0.88    |
| Mid wolf    | 0.89  | -0.17 | -0.41  | -0.15       | -0.62   | -0.57      | -0.88    | -0.62   | 0.97       | -0.92   | -0.93     | ****     |

(Moss, 1968); and (2) the results of correlation analysis are not consistent from run to run.

There are a variety of reasons for these failures (also see Minkoff, 1965, for discussion of correlation analysis and allometric growth). One of the most important reasons has to do with standardization process. The iso-kit foxes illustrate how this problem arises. The small kit fox is the smallest canid in the study; thus, its skull length and dentary thickness are both relatively small. When the large kit fox data set was created, its skull length, by definition, became equal to that of the largest canid, the wolf. However, the dentary thickness of a jaw of a wolf is proportionally much greater than that of a kit fox due to allometric growth. Therefore, even though the skull of the large iso-kit fox is as large as that of the real wolf, its dentary is much thinner. When the data are standardized, the small kit fox receives very small scores for both skull length and dentary thickness ( $-1.1078$  and  $-1.132$ , respectively). In contrast, the large kit fox receives a large score for skull length ( $1.678$ ), but only a slightly higher than average score ( $0.588$ ) for its dentary thickness (as compared with  $2.269$  for the wolf). When looking at only these two traits, the small kit fox has standardized scores which indicate that its skull length and jaw thickness are equally small (the ratio of jaw thickness to skull length is  $1.05$ ). This is in sharp contrast to the large kit fox whose scores indicate a large skull length but much smaller jaw thickness than would have been predicted from the scores of the small kit fox (the ratio for the large kit fox is  $0.348$ ). This difference and others reduce the correlation between the small and large kit fox from the expected value of  $1.0$  to  $0.553$  (table 1).

The result just cited is from a data set where most of the animals were fox-like. In a second data set where more wolf-like OTUs were added, the normal mid- to large-sized animals had a much thicker dentary than before; the mean dentary thickness shifted from  $19.3$  mm in the first run to  $22.2$  mm in the second analysis. The result is that the large kit fox looks all the more peculiar after standardization in the second run. This time the scores of the small kit fox were skull length,  $-1.315$ ; and dentary thickness,  $-1.371$  (the ratio is  $1.04$ ). The scores of the large kit fox were skull length,  $1.460$ ; and dentary thickness,  $0.203$  (a ratio of  $0.139$ ). Clearly, according to the standardized scores, these two iso-OTUs, at least for this trait, appear to have different shapes.

#### *Log Transformation*

The results of the correlation analysis on log-transformed data are shown in Table 3. These results are consistent with the known shape relationships of the iso-OTUs.

#### RATIOS

Distance in standardized ratio space has a unique property as compared with that in the previous methods in that ratios are shape characters (Simpson et al., 1960; Corruccini, 1975). The axes in hyperspace are now shape axes instead of simple raw data axes, and any change along an axis now represents an actual change in shape. Because ratios formed with size-related characters as denominators are shape measures, there is no need for a method to remove the effect of size further. It is true that a ratio may be highly correlated with linear

size, and this is often related to allometric growth patterns. For instance, large canids generally have relatively narrow crania as compared with small canids. Therefore, a negative correlation exists between the ratio of cranial width/skull length and some measure of size such as skull length or head and body length. This relationship should not be confused with the size correlation discussed under sizeout analysis. The correlation between the cranial width/skull length ratio and size represents the allometric change of shape with size, whereas the positive linear correlation between cranial width and skull length represents the degree of maintenance of shape with increasing size.

As with sizeout analysis, distances computed in ratio space measure the similarity in shape of OTUs; small distances indicate high shape similarities. Standardization, which had such a detrimental effect on correlation, does not adversely affect distance in ratio space. The results of ratio analysis on the iso-OTUs are entirely consistent with our understanding of their true shape relationships (table 4). The only effect standardization has is to assess differences in shape relative to the total difference present in the data matrix. Thus, if two OTUs vary greatly on the ratio axis on which other OTUs do not, these two OTUs would be more distant than if the other OTUs that were included in the analysis varied greatly on this axis as well. This may not be a problem as long as one remembers that distance in ratio space after standardization is relative. In many cases relative shape distances are sufficient; however, if some absolute measure is needed, then standardization cannot be used.

It should be noted here that Atchley et al. (1976) have recently strongly questioned the use of ratios. Their article was equally vigorously rebutted by Hills (1978), Dodson (1978), and Albrecht (1978); but see Atchley (1978) and Atchley & Anderson (1978). At this point it is certainly safe to say that ratios are controversial. My own view from reading all the papers concerned is that reasonable, thoughtful use of ratios may still be a powerful tool.

#### REGRESSION-RESIDUAL ANALYSIS

Several methods are combined under this topic. All involve using the data to generate a vector that is size related and then removing the effect of that vector on the data set. Methods that can be used here are regression, reduced major axis, and partial correlation. This technique is similar to sizeout, except that now one variable is specified as the size factor. The general idea is to regress this variable against all other variables. The residuals from this analysis indicate the relative size of each variable. This method suffers from the same problem as sizeout. The slope and intercept of the regression lines are affected by the OTUs included in the study. As in the sizeout example, the iso-wolves and iso-foxes are not affected equally by this analysis (table 4). Once again the iso-wolves are more effectively characterized than the iso-kit foxes.

This kind of analysis, often in conjunction with log transformation, can be useful in removing the effects of allometric growth from data. In such cases, the effect of consistent changes in shape can be removed. However, care should be used to know exactly what is being removed from an analysis. At the extreme, a researcher has used this approach (partial correlation) to remove the effect of size from chromosome number and color pattern.

## LOG-SIZEOUT

The log-sizeout method is similar to the sizeout and regression-residual methods in that the effect of a size vector is removed from the data matrix. The difference in the log-sizeout method is that the size vector is not determined by the data matrix. This is possible because, after a data matrix is log transformed, all the iso-OTU lines have equal slopes in hyperspace. A simple bivariate plot can illustrate this. In Figure 1, skull length and cranial width are plotted. Notice once again the two iso-OTU lines are of different slopes. Because of the difference in slope, no single vector can remove the effect of size from both of these groups. However, when the data are log transformed, both iso-OTU lines now have a slope of 1, with different intercepts (fig. 3). Log-sizeout removes the effect of size by removing the effect of a vector with slope 1 in all dimensions

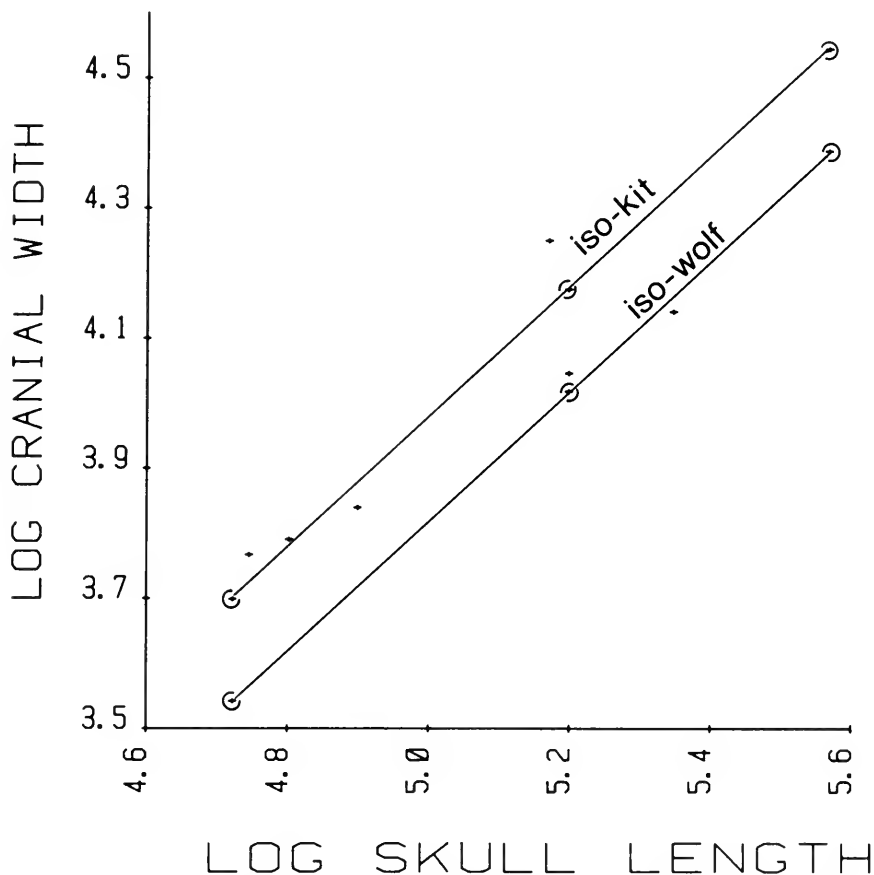


FIG. 3. Once the data set has been log transformed, all iso-OTU vectors are parallel and have slope 1.00. Compare this situation with standardized data plotted in Figure 1.

TABLE 4. Two shape coefficient matrices: below are the distances between OTUs in ratio space, and above are inter-OTU distances, using the regression method.

|             | Wolf | Dog   | Coyote | Hunting dog | Red fox | Arctic fox | Gray fox | Kit fox | Small wolf | Mid kit | Large kit | Mid wolf |
|-------------|------|-------|--------|-------------|---------|------------|----------|---------|------------|---------|-----------|----------|
| Wolf        | **** | 34.70 | 36.80  | 40.90       | 35.10   | 35.60      | 35.60    | 34.70   | 11.70      | 45.00   | 57.80     | 6.30     |
| Dog         | 1.30 | ****  | 9.20   | 45.60       | 20.50   | 16.10      | 15.50    | 13.90   | 25.10      | 20.60   | 32.10     | 29.30    |
| Coyote      | 1.23 | 0.57  | ****   | 43.20       | 14.40   | 12.00      | 10.40    | 11.45   | 27.90      | 15.30   | 26.30     | 31.80    |
| Hunting dog | 1.70 | 1.97  | 1.91   | ****        | 39.20   | 35.90      | 37.40    | 38.80   | 39.10      | 42.30   | 49.60     | 39.50    |
| Red fox     | 1.68 | 1.28  | 0.88   | 1.91        | ****    | 10.10      | 10.10    | 16.00   | 29.30      | 17.20   | 26.00     | 31.50    |
| Arctic fox  | 1.71 | 1.20  | 1.23   | 1.79        | 1.00    | ****       | 5.20     | 10.40   | 27.90      | 12.40   | 23.40     | 31.20    |
| Gray fox    | 1.90 | 1.32  | 0.99   | 1.84        | 0.62    | 0.97       | ****     | 8.50    | 27.50      | 11.10   | 22.80     | 31.00    |
| Kit fox     | 1.83 | 1.12  | 1.14   | 2.07        | 1.16    | 0.73       | 1.00     | ****    | 24.70      | 11.40   | 24.90     | 31.00    |
| Small wolf  | 0.00 | 1.03  | 1.23   | 1.70        | 1.68    | 1.70       | 1.90     | 1.83    | ****       | 35.80   | 49.10     | 5.30     |
| Mid kit     | 1.83 | 1.12  | 1.14   | 2.07        | 1.16    | 0.73       | 1.00     | 0.00    | 1.83       | ****    | 13.50     | 39.90    |
| Small kit   | 1.83 | 1.12  | 1.14   | 2.07        | 1.16    | 0.73       | 1.00     | 0.00    | 1.83       | 0.00    | ****      | 53.00    |
| Mid wolf    | 0.00 | 1.03  | 1.23   | 1.70        | 1.68    | 1.71       | 1.90     | 1.83    | 0.00       | 1.83    | 1.83      | ****     |

TABLE 5. Shape coefficients based on the distances between OTUs in log sizeout space.

|             | Wolf | Dog  | Coyote | Hunting dog | Red fox | Arctic fox | Gray fox | Kit fox | Small wolf | Mid kit | Large kit | Mid wolf |
|-------------|------|------|--------|-------------|---------|------------|----------|---------|------------|---------|-----------|----------|
| Wolf        | **** |      |        |             |         |            |          |         |            |         |           |          |
| Dog         | 0.11 | **** |        |             |         |            |          |         |            |         |           |          |
| Coyote      | 0.14 | 0.06 | ****   |             |         |            |          |         |            |         |           |          |
| Hunting dog | 0.22 | 0.23 | 0.23   | ****        |         |            |          |         |            |         |           |          |
| Red fox     | 0.23 | 0.17 | 0.17   | 0.25        | ****    |            |          |         |            |         |           |          |
| Arctic fox  | 0.22 | 0.16 | 0.13   | 0.19        | 0.09    | ****       |          |         |            |         |           |          |
| Gray Fox    | 0.28 | 0.21 | 0.17   | 0.24        | 0.09    | 0.11       | ****     |         |            |         |           |          |
| Kit fox     | 0.25 | 0.17 | 0.14   | 0.24        | 0.09    | 0.10       | 0.09     | ****    |            |         |           |          |
| Small wolf  | 0.00 | 0.11 | 0.14   | 0.22        | 0.23    | 0.21       | 0.28     | 0.25    | ****       |         |           |          |
| Mid kit     | 0.25 | 0.17 | 0.14   | 0.24        | 0.09    | 0.10       | 0.09     | 0.00    | 0.25       | ****    |           |          |
| Large kit   | 0.25 | 0.17 | 0.14   | 0.24        | 0.09    | 0.10       | 0.09     | 0.00    | 0.25       | 0.00    | ****      |          |
| Mid wolf    | 0.00 | 0.11 | 0.14   | 0.22        | 0.23    | 0.22       | 0.28     | 0.25    | 0.00       | 0.25    | 0.25      | ****     |



related to size and zero slope for variables that are not size related. The effect of this vector can be removed from the distance matrix by the equation:

$$D_{\text{log-sizeout } 1,2} = \sqrt{D_{\text{log } 1,2}^2 - (N\bar{X}_{\text{log } 1} - N\bar{X}_{\text{log } 2})^2}$$

where  $N$  is equal to the number of size-related variables,  $\bar{X}_{\text{log } 1}$  is the mean of the log-transformed size variables, and  $D_{\text{log } 1,2}$  is the euclidean distance between two OTUs in log-transformed space. The quantity  $N\bar{X}_{\text{log } 1}$  is equal to the location of an object, on the size vector. This location is one definition for size. The equation can be read as the distance from object 1 to object 2 in log-transformed hyperspace minus their difference in size, which leaves only the shape differences to affect distance.

Analysis of the canid data indicates that all iso-OTUs are correctly classified (table 5).

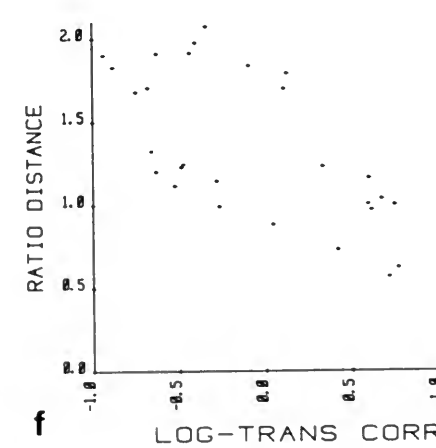
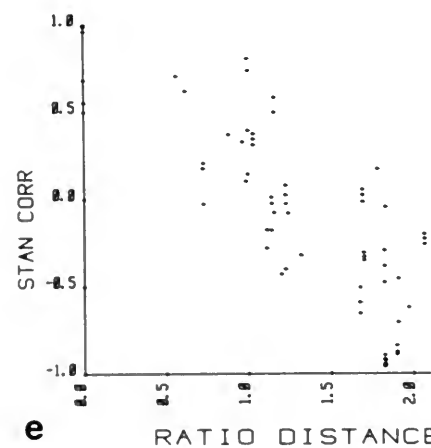
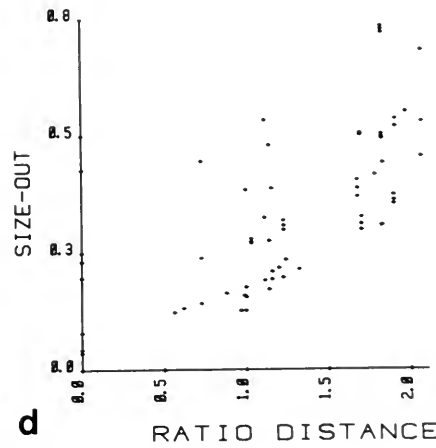
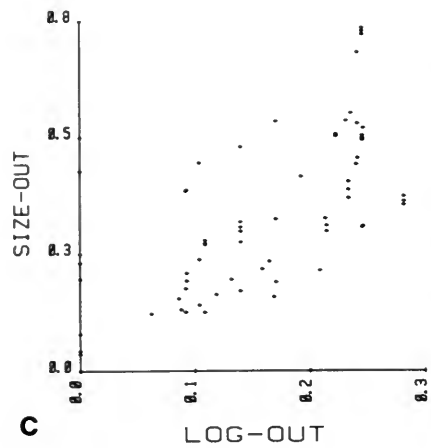
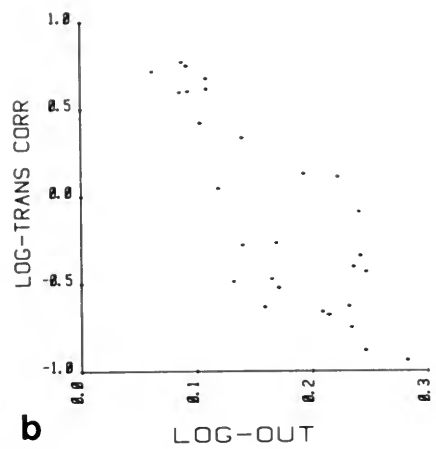
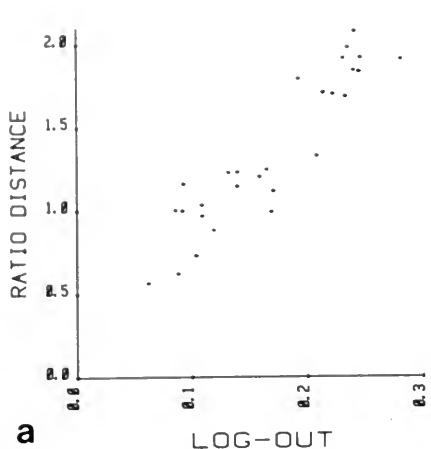
### CONGRUENCE OF TECHNIQUES

Another way of comparing methods of shape analysis is to see how similarly the methods assess shape relations among a group of OTUs. One way to do this is to correlate the similarity (or dissimilarity) matrices of all the techniques. Several of these scattergrams are shown in Figure 4a-h, and Table 6 contains all the "cophenetic" correlation coefficients between methods. Not surprisingly, in view of the evidence already presented, the correlation between methods is not always particularly high. As an example, two methods mentioned as possibilities by Sneath & Sokal (1973), sizeout and correlation, are correlated with an  $r$ -value of only 0.623 (fig. 4a). To get an overview of the relationships of the methods, I constructed a phenogram based on the correlations (fig. 5). The most striking point of this phenogram is that techniques using similar methods give similar results.

The sizeout and regression methods are similar mathematically in their way of removing a size vector. Consistent with that fact is that the shape relationships generated by these two methods are most similar to one another. Likewise with the correlation of standardized data and log-transformed data, the mathematics are most similar and the results are more similar to one another than to the other methods. The log-sizeout method is harder to classify mathematically. On one hand it can be visualized as the removal of a size vector of slope 1 in all size-related dimensions, much the same as sizeout. On the other hand it resembles the ratio technique, as it entails the subtraction of logarithms (a process closely related to division of non-log numbers). The phenogram

TABLE 6. The congruence of the methods of shape analysis discussed in this report can be seen, using these inter-method correlation coefficients.

|             | Log-sizeout | Ratio  | Log corr | Stan corr | Sizeout | Regress |
|-------------|-------------|--------|----------|-----------|---------|---------|
| Log-sizeout | 1.0         |        |          |           |         |         |
| Ratio       | 0.959       | 1.0    |          |           |         |         |
| Log corr    | -0.895      | -0.801 | 1.0      |           |         |         |
| Stan corr   | -0.887      | -0.822 | 0.924    | 1.0       |         |         |
| Sizeout     | 0.667       | 0.677  | -0.531   | -0.623    | 1.0     |         |
| Regress     | 0.738       | 0.685  | -0.563   | -0.655    | 0.915   | 1.0     |



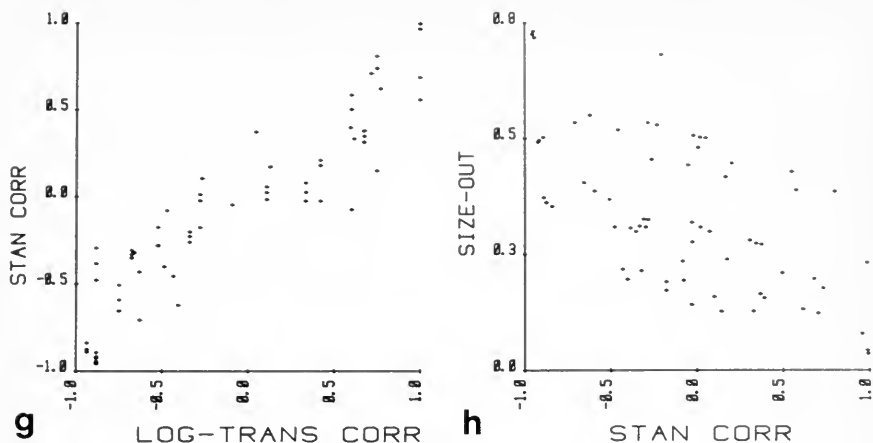


FIG. 4. The shape coefficients of the different methods are plotted against one another in scattergrams to obtain a visual representation on the congruence of these techniques. **a**, Comparison of distance of OTUs in ratio space and log-sizeout space; this yielded the highest correlation,  $r = 0.959$ . **b**, Comparison of log-sizeout with the correlation coefficients, using log-transformed data. **c**, Comparison of the log-sizeout method with regular sizeout; notice the wide scatter ( $r = 0.667$ ) between these two methods. Along the vertical axis are several circled points; these are the iso-OTU comparisons where shapes of the OTUs are known to be identical. Although log-sizeout correctly classifies these relationships, sizeout does a poor job. **d**, Comparison of distances in ratio space with distances in sizeout space. **e**, Comparison of distances in ratio space and correlation coefficients on standardized data. As mentioned in the text, the correlation method failed here, but comparison with graph **c** shows that correlation was more effective than sizeout in classifying iso-OTUs. **f**, Comparison of the correlation coefficients on log-transformed data with distances in ratio space. **g**, Comparison of the correlation coefficients on log-transformed data with the correlation coefficients on standardized data; notice the tight relationship ( $r = 0.924$ ), which indicates the similarity of the two methods. **h**, Comparison of the correlation coefficients on standardized data with distance in sizeout space; notice the surprisingly low correlation of these two techniques ( $r = -0.623$ ).

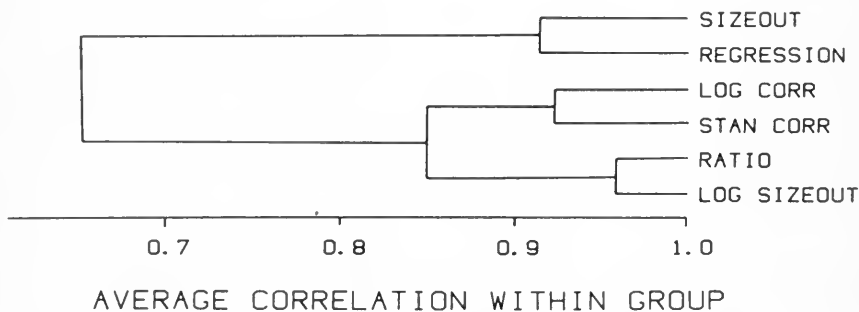


FIG. 5. A phenogram of the methods of shape analysis studied in this report was made based on the inter-method correlations of Table 6. The phenogram shows the methods fall into three groups, with the groups of correlation of log-transformed data (LOG CORR) and the correlation of standardized data (STAN CORR) and ratio distances (RATIO) and distances in log-sizeout space being most similar. Sizeout distances (SIZEOUT) and the regression method (REGRESSION) are more distantly related.

indicates that the results of log-sizeout agree most closely with those of ratio technique and are only distantly related to results of sizeout or regression.

### SUMMARY

Six methods of shape analysis were tested, and half of them—sizeout, regression analysis, and correlation of standardized data—failed to classify iso-OTUs correctly. Three others—correlation of log-transformed data, log-sizeout, and ratio distance—all passed my test equally well. Even among the successful tests, however, there are considerable differences in the estimated shape relationships. Based on the data presented here, it is not possible to distinguish if one of the three methods is more successful than the others, or if differences in results represent different views of shape, internally consistent and equally valid.

Having said this, I will state that my own slight preference is for the log-sizeout methods. Correlation, even with log transformation, suffers from some conceptual problems mentioned above. Ratios, which on the surface appear the most satisfactory of shape indices, have been attacked as producing spurious correlations (Atchley et al., 1976).

I believe that studies such as this one are necessary for critical judgment of methods of shape analysis. Using iso-OTUs has proved partially successful in giving insights into how these methods work, and sometimes fail to work. More studies, perhaps with more sophisticated artificial OTU shape relationships, are needed to continue this work.

### ACKNOWLEDGMENTS

I wish to thank Dr. Patricia W. Freeman and the Division of Mammals for the use of their specimens and facilities and the Advanced Technology Laboratories of Field Museum for the use of their computer equipment. I would also like to thank Sarah Derr Bruner for typing several drafts of this manuscript.

### LITERATURE CITED

- ALBRECHT, G. 1978. Some comments on the use of ratios. *Syst. Zool.*, **27**: 67-71.
- . 1980. Multivariate analysis and the study of form, with special reference to canonical variate analysis. *Am. Zool.*, **20**: 679-693.
- ATCHLEY, W. R. 1978. Ratios regression, intercepts and the scaling of data. *Syst. Zool.*, **27**: 78-83.
- ATCHLEY, W. R., AND D. ANDERSON. 1978. Ratios and the statistical analysis of biological data. *Syst. Zool.*, **27**: 71-78.
- ATCHLEY, W. R., C. T. GASKIN, AND D. ANDERSON. 1976. Statistical properties of ratios. I. Empirical results. *Syst. Zool.*, **25**: 137-148.
- CORRUCCINI, R. S. 1973. Size and shape in similarity coefficients based on metric characters. *Am. J. Phys. Anthropol.*, **38**: 743-754.
- . 1975. Multivariate analysis in biological anthropology: Some considerations. *J. Human Evol.*, **4**: 1-19.
- CROVELLO, T. J. 1969. Effects of change of characters and of number of characters in numerical taxonomy. *Am. Midl. Nat.*, **81**: 68-86.
- DODSON, P. 1978. On the use of ratios in growth studies. *Syst. Zool.*, **27**: 62-67.
- GOULD, S. J. 1966. Allometry and size in ontogeny and phylogeny. *Biol. Rev.*, **41**: 587-640.
- HILLS, M. 1978. On ratios—A response to Atchley, Gaskin and Anderson. *Syst. Zool.*, **27**: 61-62.

- HUDSON, G. E., R. A. PARKER, J. VANDEN BERGE, AND P. J. LANZILLOTTI. 1966. A numerical analysis of the modifications of the appendicular muscles in various genera of gallinaceous birds. *Am. Midl. Nat.*, **76**: 1-73.
- JOLICOEUR, P. 1963. The multivariate generalization of the allometry equation. *Biometrics*, **19**: 497-499.
- JOLICOEUR, P., AND MOSIMANN, J. E. 1960. Size and shape variation in the painted turtle, a principal component analysis. *Growth*, **24**: 339-354.
- LEWONTIN, R. 1966. On the measurement of relative variability. *Syst. Zool.*, **15**: 141-142.
- MANISCHEWITZ, J. R. 1973. Prediction and alternative procedures in numerical taxonomy. *Syst. Zool.*, **22**: 176-184.
- MICHENER, C. D., AND R. R. SOKAL. 1957. A quantitative approach to a problem in classification. *Evolution*, **11**: 130-162.
- MINKOFF, E. C. 1965. The effects of classification on slight alterations in numerical technique. *Syst. Zool.*, **14**: 196-213.
- MORIARTY, D. J. 1977. On the use of variance of logarithms. *Syst. Zool.*, **26**: 92-93.
- MOSIMANN, J. E. 1970. Size allometry: Size and shape variables with characterizations of the log normal and generalized gamma distributions. *J. Am. Stat. Assoc.*, **65**: 930-942.
- MOSS, W. W. 1968. Experiments with various techniques of numerical taxonomy. *Syst. Zool.*, **17**: 31-47.
- . 1971. Taxonomic repeatability: An experimental approach. *Syst. Zool.*, **20**: 309-330.
- ROHLF, F. J. 1962. A numerical taxonomic study of the genus *Aedes* (Diptera: Culicidae) with emphasis on the congruence of larval and adult classifications. Ph.D. thesis, University of Kansas, 98 pp.
- . 1972. An empirical comparison of three ordination techniques on numerical taxonomy. *Syst. Zool.*, **21**: 271-280.
- ROHLF, F. J., AND R. R. SOKAL. 1965. Coefficients of correlation and distance in numerical taxonomy. *Univ. Kansas Sci. Bull.*, **45**: 3-27.
- ROHLF, F. J., J. KISHPAUGH, AND D. KIRK. 1971. NT-SYS, numerical taxonomy system of multivariate statistical programs. Tech. Rep. State University of New York at Stony Brook, New York.
- SIMPSON, G. G., A. ROW, AND R. C. LEWONTIN. 1960. *Quantitative Zoology*. Harcourt, Brace and World, Inc., New York, 440 pp.
- SNEATH, P. H. A., AND R. R. SOKAL. 1973. *Numerical taxonomy, the principles and practice of numerical classification*. W. H. Freeman and Company, San Francisco, 573 pp.
- SWEET, S. S. 1980. Allometric inference in morphology. *Am. Zool.*, **20**: 643-652.
- UNDERWOOD, R. 1969. The classification of constrained data. *Syst. Zool.*, **18**: 312-317.







Field Museum of Natural History  
Roosevelt Road at Lake Shore Drive  
Chicago, Illinois 60605-2496  
Telephone: (312) 922-9410

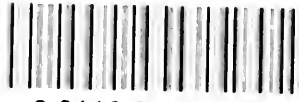








UNIVERSITY OF ILLINOIS URBANA



3 0112 018406972