BIOMETRICAL METHODS IN THE STUDY OF INVERTEBRATE FOSSILS

JOHN IMBRIE

BULLETIN
OF THE
AMERICAN MUSEUM OF NATURAL HISTORY
VOLUME 108 : ARTICLE 2 NEW YORK : 1956

BIOMETRICAL METHODS IN THE STUDY OF INVERTEBRATE FOSSILS

BULLETIN OF THE AMERICAN MUSEUM OF NATURAL HISTORY
Volume 108, article 2, pages 211–252, figures 1–10, tables 1–7

Issued January 30, 1956

Price: 75 cents a copy

BIOMETRICAL METHODS IN THE STUDY OF INVERTEBRATE FOSSILS

JOHN IMBRIE

Research Associate, Department of Geology and Paleontology
The American Museum of Natural History
Associate Professor of Geology
Columbia University

BULLETIN
OF THE
AMERICAN MUSEUM OF NATURAL HISTORY
VOLUME 108 : ARTICLE 2 NEW YORK : 1956

CONTENTS

INTRODUCTION

STUDENTS OF INVERTEBRATE FOSSILS hold widely divergent opinions concerning the applicability of biometrical methods. One extreme of modern practice is illustrated in the monograph by Deleers and Pastiels (1952) on collections of *Lingula* from the Belgian lower Carboniferous, a study combining exceedingly complete quantitative data with elaborate statistical analysis. A completely different approach, exemplified by Wang (1949) in an excellent modern treatment of an Ordovician brachiopod fauna, employs a minimum of quantitative data and no formal statistics.

Advocates of more widespread use of quantitative methods in invertebrate paleontology have indicated the advantages of these techniques over procedures that do not make explicit use of statistics. These arguments can, in general, be reduced to three:

1. Descriptions of taxonomic characters that can be simply and logically related to measurements (or counts) are always more accurate if observations are made and presented quantitatively. Thus, "mean length, 10.3 mm." is far more revealing than "length average for the genus," or "medium size."

2. With the abandonment of the typological concept, it is no longer desirable to limit the description of a species to observations on the holotype. In fact, a logical consequence of the acceptance of the interbreeding population as the basic taxonomic unit is the expansion of species descriptions to include observations on representative groups of individuals. (This procedure, of course, in no way diminishes the importance of the holotype as a nomenclatorial reference point.) The most direct and perhaps the simplest way of accomplishing this is to present descriptive data in tabular form. Such tables are often impractical to publish, however, and usually obscure rather than emphasize essential characteristics of the sample. Clearly, descriptive devices are needed that will summarize compactly the essential characteristics of a series of observations. Just such devices, called statistics, have been designed. They have been repeatedly tested empirically; their theoretical basis has been thoroughly scrutinized; and they are very widely understood. Statistics are therefore recommended as the most efficient tools to use in dealing with quantitative observations on collections of invertebrate fossils.

Now it is often possible to record some of the important characteristics of collections without the formal computation of statistics. A photograph or measurement of a "typical specimen," for example, conveys in a general way the same information as a series of arithmetic means. Similarly, measurement of the smallest and largest observed specimens (observed range) in a sample of known size is a reasonable way of recording information on variability. Useful as these substitutes are for the formal computation of statistics, they have definite limitations in invertebrate paleontology. The problem of measuring biological variability in typical collections of invertebrate fossils is a case in point. Such samples usually include a mixture of indistinguishable age classes so that the observed range of many characters is not related to inherent biological variability. Variability is in this case best measured by a statistic (the coefficient of relative dispersion) for which there is no effective qualitative substitute.

3. Although an essential part of taxonomy involves the study of groups of actual specimens, the science is fundamentally concerned with the drawing of inferences about living populations. Viewed in this way, every species description is an act of faith based on the assumption that from the characteristics of the specimens actually at hand it is possible to draw useful inferences concerning the original population. Now the science of statistics is largely concerned with the analysis of sample-population relationships. So many useful guides to understanding have been worked out and tested that it is surely improvident for a taxonomist not to take advantage of this generalized experience as he attacks his special problems. This is not to say that any amount of statistical work on a sample can ever tell the taxonomist exactly what the characteristics of the original population were. But by application of statistical principles he can do the next best thing: predetermine the chance he is willing to take of making an error (say, 1 in 100), and then describe an interval within which he is sure

(99% sure) that the population value lies. Closely related to this process of establishing confidence limits for estimates of population characteristics is the problem of statistical discrimination: Given two collections that differ in some character, what is the probability that this difference arose by chance in the sampling of two identical populations? A satisfactory answer to this question is a logical prerequisite to sound taxonomic discrimination. In many cases, of course, the decision is quite clear without formal statistical calculation. In others, the taxonomist is grateful for an unbiased guide.

These arguments in favor of the biometric approach to taxonomic problems are certainly cogent, and it is worth while to inquire why such techniques have not in fact been more widely adopted in invertebrate paleontology. Is it because the paleontologist lacks sufficient special training to apply the methods of statistics? This question can be answered firmly in the negative, because a number of students, in addition to arguing for a more widespread use of quantitative methods, have provided clear explanations of procedures with examples drawn from invertebrate paleontology. Papers by Burma (1948, 1949), Miller (1949), Olson and Miller (1951), and Kermack (1954) may be cited as typical examples of this group. In addition to these publications relating directly to invertebrate fossils there is a large and growing literature on the application of biometrics to taxonomic problems in general. Many of these works, e.g., Simpson and Roe (1939), Cazier and Bacon (1949), and Mayr *et alii* (1953, chap. 7), are designed for the biologist lacking special mathematical training. Slightly more demanding in terms of mathematical requirements, but by no means formidable, are a number of clearly written basic statistical texts, e.g., Snedecor (1946), Dixon and Massey (1951), and Wilks (1951). Hence the sparsity of biometrical analyses on invertebrate fossils must be due to something other than a lack of special training.

Two serious objections have been raised against the use of biometrical techniques: first, that these techniques cannot be validly applied to most invertebrate fossils; and, second, that the techniques are too time-consuming to justify their general application. These objections, and others, are considered at some length in the present paper. The conclusion is reached that there are techniques, at present not widely applied, which violate neither statistical nor paleontological principles, and which are valuable tools in attacking several kinds of taxonomic problems. Further, the statistics necessary to characterize the average paleontologic sample can be computed in no more time than it takes to prepare a thin-section, provided that suitable computational aids are employed. It seems therefore that at the species level much can be gained and nothing lost by the inclusion of suitable statistical data. The full scientific fruits of biometrical investigation will not be realized until the description of a large number of samples includes statistical characterization. This basic task should not, indeed cannot, be deferred until every genus is known and every species qualitatively diagnosed.

THE PLACE OF BIOMETRICS IN TAXONOMY

BIOMETRICS AT THE SPECIFIC AND INFRASPECIFIC LEVELS

BIOMETRICS IS HERE CONSIDERED to be the statistical treatment of quantitative morphological data. Thus defined, biometrics has a limited application in the field of taxonomy —a limitation due, first, to the fact that many important taxonomic characters cannot be efficiently represented numerically, and, second, to the fact that many taxonomic problems which do involve quantitative data can be most effectively dealt with by nonstatistical means. At the generic level of classification, for example, there is rarely a need for statistical analysis. The genus is usually diagnosed on a limited set of relatively constant characters, so that in practice any complete adult specimen can be identified generically. In groups so defined there is normally little need or place for statistical aids in description and discrimination.

When populations, subspecies, and closely related species are dealt with, however, taxonomic problems of an essentially different sort are encountered. At these levels discrimination is usually based on group tendencies rather than on the possession of key characters that distinguish all members of one group from all members of another.[1] In precisely these circumstances biometrical procedures can be most effectively and usefully applied. Techniques discussed and illustrated in this paper are therefore designed primarily for specific and infraspecific levels.

In one sense the existence of morphologically overlapping populations is merely a taxonomic problem subtle enough to require the aid of statistical methods. Viewed broadly, however, intergradation on specific and infraspecific levels offers a challenging opportunity to the paleontologist equipped with biometrical tools—an opportunity to record with reasonable precision some of the results of dynamic evolutionary processes.

SPECIES CONCEPT IN INVERTEBRATE PALEONTOLOGY

Because the present paper is concerned primarily with the application of biometric techniques to the lower levels of classification, it is desirable to review certain aspects of the concept of fossil species.

Mayr *et alii* (1953, p. 25) define species as "groups of actually (or potentially) interbreeding natural populations which are reproductively isolated from other such groups," and probably most paleontologists today accept this view of living species. In applying this definition to fossils, the taxonomist encounters two principal difficulties. He is, first, faced with the problem of subdividing continuously evolving lineages and, second, required to find evidence on the breeding habits of organisms long dead.

The first difficulty mentioned above is more theoretical than real, for in spite of the extended attention this matter has received, the fact remains that continuous fossil records of evolving lineages are rare. Even when they do occur, subdivision need not be entirely arbitrary. The limits of stratigraphic units, for example, may define useful and, in a sense, natural taxonomic boundaries (Simpson, 1943, p. 176). From the point of view of the biologist, a more satisfactory subdivision can perhaps be achieved by the designation of one particular population as the midpoint of the morphologic range of the species. Controversy then centers about the criteria used to define the morphological range of the species. Although some students advocate qualitative, and others quantitative, evaluation, the object should be to determine with reasonable probability a morphological range that would include just those populations potentially capable of interbreeding with the population chosen as midpoint.

We are thus led to consider the second and basic difficulty involved in applying the neontologic definition of species to the fossil record, namely, that there is never any direct evidence concerning the breeding habits of fossilized organisms. Lacking such evidence, the paleontologist must proceed very much as a neontologist would in analyzing allopatric populations: he must make the best possible inference from the morphologic, geographic, and stratigraphic data at hand.

In doing this, the paleontologist must make certain simplifying assumptions. For example, it must be assumed that in general

[1] See Simpson (1945, pp. 20–22) for an enlightening discussion of the dimensions of classification.

there is a positive correlation between reproductive isolation and morphological divergence. Barring some unheralded advance in analytical techniques, this will inevitably mean that some valid species are lumped and some ecophenotypic variants split.

The paleontologist, lacking contrary evidence, must often make another assumption: that specimens found together at the same locality and horizon did in fact live together. This assumption considerably simplifies taxonomic judgment, for it is axiomatic that truly sympatric populations showing no evidence of hybridization are valid separate species.

Fossil species that are demonstrably congeneric and sympatric are of considerable importance in applied taxonomy, for they provide the student with some insight into the degree of morphological change associated with specific divergence. The experience gained in this way provides a yardstick for evaluating morphological differences between geographically or stratigraphically separated collections.

Some students hold that it is possible and desirable to distinguish subspecies that develop simultaneously in different localities from those that evolve during a portion of geologic time. The former have been called geographic, and the latter chronologic, subspecies. In fact some paleontologists (e.g., Burma, 1948, p. 741) advocate that the term subspecies be restricted to geographic subspecies. It has been noted by several students, however, that there is no basic difference between these supposedly contrasting categories, as both represent secular accumulation of genetic differences. Moreover there is a practical objection to distinguishing geographic from chronologic subspecies, for it is rarely possible to demonstrate simultaneity in the stratigraphic record.

Study of living populations has shown that if sufficiently rigorous methods are employed, significant morphologic and genetic differences between two populations can be demonstrated (see Mayr *et al.*, 1953, p. 31). Hence species and subspecies must be considered as collective categories, in the sense that they are composed of local populations no two of which are identical. The only logical alternative to this concept would be to grant each local population a formal name. The application of such a scheme to modern organisms would yield a totally impractical taxonomy. In invertebrate paleontology it would lead to chaos. Yet several modern students of fossils have proposed just this, under the guise of employing the subspecies or variety to designate the smallest statistically demonstrable unit.

SOME TAXONOMIC PROCEDURES SUBJECT TO BIOMETRIC ANALYSIS

Characterization

A sample is said to be characterized when morphological attributes judged to be taxonomically important are recorded for study or publication. Because systematic knowledge of fossil organisms is based primarily on characterizations of paleontological samples, it is obviously important that published descriptions of unit characters include compact, informative, and objective data on the entire suite of available specimens. Whenever characters selected for study can be expressed quantitatively, these objectives are most easily achieved by statistical techniques. It is demonstrated in a later section of this paper that adequate statistical characterization of any pair of variates requires the calculation of only seven quantities: N, $\bar{x}$, $\bar{y}$, s_x, s_y, r, and OR_x. Characterizations of this sort are, of course, to be viewed as complementary to qualitative descriptions. Statistical data alone are incapable of recording the morphological subtleties of the simplest biological form. Conversely, words and pictures alone cannot record the essential group features of a sample.

Statistical Discrimination

Because no two collections of fossils are ever exactly identical, most taxonomic problems at one stage involve a balancing of similarities and differences. In order that sound taxonomic judgments may be made, therefore, it is necessary to bear in mind that observed differences between any two samples can always be explained in one of two ways. First, the samples may have been derived from the same or identical populations, in

which case the differences are the chance result of taking samples of limited size from large populations. Second, the samples may represent populations that were in fact not identical. Differences of the first sort are termed "statistically non-significant"; differences of the second type are referred to as "statistically significant"; and the process of arriving at a judgment of statistical significance or non-significance is here designated statistical discrimination. Thus the aim of statistical discrimination is to answer the question: What are the chances that the observed differences between two fossil collections are due solely to random sampling error?

When closely related populations are dealt with, it is usually impossible to arrive at an exact answer to questions of statistical significance. The best that can be done is to calculate the probability that the observed difference is due to chance alone. In a judgment of this probability, three factors are involved: the size of the samples, the variability of the samples, and the degree of difference between the samples. Whenever samples are large, relatively invariable, or separated by a considerable morphological gap, statistical discrimination can be satisfactorily accomplished by visual inspection alone. In such instances it would, of course, be a needless waste of time to apply elaborate mathematical procedures. When differences are more subtle, and collections smaller or more variable, the taxonomist welcomes the objective guides provided by formal statistical analysis.

As some of the important techniques of biometrical discrimination are outlined in later portions of this paper, it would serve no purpose to elaborate them here. In essence, these methods employ quantitative measures of sample size, variability, and degree of difference in order that a numerical statement of the probability of significant difference can be achieved. Provided that samples have been statistically characterized, none of the discrimination techniques described below takes more than a few minutes to apply. Thus if there is any doubt about the statistical validity of separating two samples, there can be little practical objection to mathematical analysis.

Taxonomic Discrimination

Having determined that two samples are statistically distinct, the taxonomist must decide whether or not recognition of separate formal categories is justified. The process of arriving at judgments of this sort is here referred to as taxonomic discrimination.

Some paleontologists hold that whenever statistical differences between two samples can be demonstrated, these samples should be considered separate species or subspecies. This procedure is said to offer the most objective means of defining the limits of a taxonomic unit. However, if systematically employed, such a practice would result in a separate named species for almost every collection. Furthermore, this procedure violates the modern theoretical concept of species and subspecies as collective categories.

Taxonomic discrimination is properly based on a combination of morphologic, geographic, and stratigraphic evidence. Evaluation of this sort is thus fundamentally nonstatistical and ideally should never rest on biometrical data alone. However, to the extent to which taxonomic judgments are based on morphological data, the systematist can frequently benefit from statistical aids designed to measure the degree of morphological overlap between related populations.

Three methods of overlap analysis can be employed: univariate analysis, which considers overlap in one character at a time; bivariate analysis, which considers the amount of overlap in two related characters simultaneously; and multivariate analysis, which considers overlap in a multidimensional continuum. Because differences and similarities between real populations always involve a large number of characters, multivariate analysis is theoretically preferable. Unfortunately, such methods are too laborious and abstract for wide application. In practice, therefore, biometrical aids in both taxonomic and statistical discrimination are best confined to univariate and bivariate analyses. This limitation is not so serious as it may seem, provided that statistical methods are applied at a late stage in an investigation as a means of testing critical hypotheses. For example, if careful scrutiny has indicated

that two collections are most readily distinguished on the basis of two or three unit characters, a univariate or a bivariate biometrical study of these characters may yield valuable data.

Several convenient methods of describing univariate morphological overlap are in use. These include the population range diagram (Simpson and Roe, 1939, p. 318; Hubbs and Perlmutter, 1942; Cazier and Bacon, 1949, p. 384; Hubbs and Hubbs, 1953) and a method for estimating percentage overlap between pairs of populations (Mayr *et al.*, 1953, pp. 145–147).

Because univariate techniques have a somewhat limited application in invertebrate paleontology, attempts have been made to analyze morphological overlap in bivariate distributions. Burma (1948, p. 748) uses regression lines and the standard error of estimate as a means of estimating population ranges. Klauber (1943, p. 56) outlines the use of a measure of overlap between two bivariate samples at selected growth stages. Kotaka (1953) and Pastiels (1953, pls. 10, 11) employ ellipses of equal probability as a means of describing range of variation. None of these methods, however, has been tested widely enough to justify further attention here.

SURVEY OF SOME BASIC STATISTICAL CONCEPTS

THE PRESENT paper is an attempt to provide an elementary presentation of biometrical procedures especially relevant to invertebrate paleontology.[1] Both univariate and bivariate analyses are considered. Univariate techniques are treated very briefly, partly because many lucid explanations are available, and partly because these techniques are less widely applicable than bivariate methods.

POPULATION AND SAMPLE IN PALEONTOLOGY

Statisticians and biologists commonly use the word "population" in distinctly different but obliquely related ways, a circumstance that has been responsible for many erroneous conclusions. In order to avoid confusion, therefore, it is necessary here to define several terms.

The group of specimens actually studied during the course of an investigation is referred to here as the *statistical sample*, or simply the *sample*.

The *total interbreeding population* is defined as the entire group of organisms with which the individuals in the sample were potentially capable of interbreeding. Thus the total interbreeding population consists of an unknown but real group of organisms which existed at different times and places. In essence, this is the neontological concept of species with its inherent and arbitrary chronological limitation removed.

At any instant of geologic time, the total interbreeding population is composed of a number of *local interbreeding populations* separated geographically and exhibiting slight but measurable morphologic and genetic differences. The entire group of such local populations living at any one time corresponds to the neontological concept of species.

The *total fossil population* can be defined as all the members of the total interbreeding population that are preserved today as fossils.

The *local fossil population* includes all fossils present in the geographically and stratigraphically restricted body of rock from which the actual collection is taken. It is this population that is sampled in paleontological studies. Therefore, statistical inferences based on a sample apply directly only to the local fossil population which the sample represents.

Just as a collection of specimens is considered a sample of a local fossil population, so the local fossil population may itself be considered a sample of one or more local interbreeding populations. Although the first is made by human and the second by natural agency, both samples are imperfect reflections of the populations from which they are derived. In drawing inferences about living populations, therefore, a paleontologist must consider the quality of the available sample.

The ideal sample is said to be random, i.e., a sample taken in such a way that every individual in the population stands an equal chance of being chosen. If this condition is not met, the sample is biased. Now the size distribution of most paleontological samples is strongly biased. This condition is due not only to the geological factors that intervene between living and fossil populations, but also to the technical and psychological difficulties of securing a truly random sample of the local fossil population.

Owing to the prevalence of biased size distributions in paleontology, univariate analysis must be applied with great caution. Bivariate analysis, on the other hand, is relatively little affected by bias of this sort, and is therefore recommended for most taxonomic problems involving invertebrate fossils.

UNIVARIATE ANALYSIS

CALCULATION OF SOME IMPORTANT STATISTICS

An indispensable step in most biometrical investigations is the calculation of certain quantities that summarize important features of a sample. Such a quantity, calculated directly from observations on the sample, is called a *statistic*.

The *arithmetic mean*, or simply *mean*, is a

[1] In preparing this paper, the present writer has had the advantage of discussing many of the statistical problems involved with Prof. Howard Levene of Columbia University, but bears the sole responsibility for any errors.

statistic very widely used to express the "average" or "central" value of a series of numbers. The mean is defined as the sum of the individual observations divided by the number of observations. Expressed as a formula,

$$\bar{x}=\frac{\sum(x)}{N} \qquad (1)$$

where

$\bar{x}$ = the mean of x,
$\sum$ = the sum of all quantities indicated in parentheses,
x = the value of any variate, and
N = the number of individuals in the sample.

It is often necessary to have a quantitative measure of the variability of a sample. One of the most useful of such measures is a statistic called the *standard deviation*, defined as the positive square root of the sum of the squares of the deviations from the mean, divided by one less than the number of observations. Thus,

$$s=\sqrt{\frac{\sum(d^2)}{N-1}} \qquad (2)$$

where

s = the standard deviation, and
$d=x-\bar{x}$, the difference between any observation and the mean.

The use of these formulas will be made clear by reference to table 1.

The standard deviation, as is the mean, is expressed in whatever units of measurement were used in the recording of the original data. In table 1, for example, the standard deviation is 2.24 mm. In other words, the standard deviation is a measure of absolute variation. But the amount of variation displayed in any organism is usually proportional to the absolute size of the organism. Hence, if the objective of an investigation is to compare the variability of two organisms of different size, a meaningful comparison cannot be made on the basis of the standard deviation alone. Such comparisons can, however, be made by expressing the standard deviation as a percentage of the mean. The measure of relative variation thus derived is called the *coefficient of variation* (V) and is defined as

$$V=\frac{100s}{\bar{x}}. \qquad (3)$$

TABLE 1

CALCULATION OF THE MEAN AND STANDARD DEVIATION FOR A HYPOTHETICAL SAMPLE (DATA IN MILLIMETERS)

x	d	d^2
9	+3	9
7	+1	1
6	0	0
5	−1	1
3	−3	9

$\sum(x)=30$

$N=5$

$$\bar{x}=\frac{\sum(x)}{N}=\frac{30}{5}=6 \text{ mm.}$$

$$s=\sqrt{\frac{\sum(d^2)}{N-1}}=\sqrt{\frac{20}{4}}=\sqrt{5}=2.24 \text{ mm.}$$

Because V is a ratio, it is a pure number without dimension, and can be used to compare the variability of organisms (or organs) of different size. For the data in table 1,

$$V=\frac{100\ (2.24 \text{ mm.})}{6 \text{ mm.}}=37.$$

Another statistic useful in biometrical characterization is the *observed range* (*OR*). This is simply a record of the smallest and largest values of the variate observed in the sample. For the data in table 1, *OR* equals 3 mm.–9 mm. Note that a larger sample would be very likely to have a larger *OR*, but essentially the same standard deviation.

ESTIMATION OF PARAMETERS

In most paleontological work it is impossible to observe the entire local fossil population. This population exists, however, and has real (but unknown) values of the mean, standard deviation, and other "statistics." Population values of this kind are designated *parameters*. Thus for every sample statistic there is a corresponding population parameter. Usually a Greek letter is used for the parameter, and a Roman letter for the statistic (see table 2).

Now the value of a statistic computed for any sample will almost never be exactly equal to the corresponding population param-

TABLE 2

SYMBOLS FOR SOME STATISTICS AND PARAMETERS

	Statistic	Parameter
Univariate analysis		
Mean	$\bar{x}$	μ (mu)
Standard deviation	s	σ (sigma)
Bivariate analysis		
Growth ratio	a	α (alpha)
Initial growth index	b	β (beta)
Correlation coefficient	r	ρ (rho)

eter. One of the central problems of biometrical analysis, therefore, is to devise methods for judging the reliability of any statistic as an estimate of a parameter. This can be accomplished by the computation of a quantity known as the *standard error*.

To illustrate the use of standard errors as measures of reliability, consider the data in table 1. The sample mean ($\bar{x}=6$ mm.) is viewed as an estimate of the population mean (μ). The standard error of the mean ($\sigma_{\bar{x}}$) is then calculated according to the formula

$$\sigma_{\bar{x}}=\frac{s}{\sqrt{N}}. \qquad (4)$$

Thus

$$\sigma_{\bar{x}}=\frac{2.24\ mm.}{\sqrt{5}}=\frac{2.24 mm.}{2.24}=1.00 mm.$$

This value is a measure of the reliability of $\bar{x}$ as an estimate of the population mean. The smaller the value of the standard error the more reliable the estimate. Note that a larger N or a smaller s would yield a smaller standard error and correspondingly greater reliability.

The general nature of the relationship formalized in equation 4 is intuitively obvious. The real advantage of the standard error lies in its precise definition of the relative importance of the factors involved.

In order to make an exact interpretation of any calculated standard error of the mean (i.e., to derive a confidence interval) it is necessary to assume that the values of the variate x in the original population are distributed normally, i.e., that they form a frequency distribution that can be closely approximated by the normal curve. Experience has shown that this assumption is valid for the vast majority of biological variates in statistical populations that are homogeneous with respect to sex, age, stratigraphic position, and geographic location. Moreover, deviations from normality are not serious if the available sample is large.

With regard to the data of table 1, a confidence interval for the mean of the population can be obtained as follows:

1. Choose the risk one is willing to take of making an erroneous inference about the population. This risk is expressed in terms of probability and symbolized as P. Thus a P of 0.05 indicates that one wishes to determine a range of values such that the chances are no greater than five in 100 (5%) of excluding the real population mean. In this example, P is chosen as 0.05.

2. Determine the number of *degrees of freedom* (*d. f.*).[1] In the calculation of the confidence interval for the mean,

$$d.f.=N-1.$$

In this example

$$d.f.=5-1=4.$$

3. Find the value t corresponding to the number of degrees of freedom (*d. f.*) and the probability level P. Values of t are tabled in all statistical texts (e.g., Snedecor, 1946, p. 65). In this example t is 2.776. The confidence interval is then given by

$$\mu=\bar{x}\pm t\sigma_{\bar{x}}.$$

Thus

$$\mu=6\text{ mm.}\pm 2.78\,(1.00\text{ mm.})=3.22\text{ mm.}-8.78\text{ mm.}$$

One is sure that 95 per cent of the time this calculated interval will include the value of the population mean. This is called the 95 per cent confidence interval. Expressed another way, one may conclude that about 95 per cent of all other samples of the same size will possess means that fall in the designated interval.

Other values of P may of course be selected. For $P=0.01$, $t=4.604$, and the corresponding 99 per cent confidence interval is

[1] The concept of degrees of freedom is difficult to define. As an approximation, the number of degrees of freedom can be described as the number of independently determinable quantities.

1.40–10.60. In both of these calculated intervals the range is quite large, as would be expected with such a small and highly variable sample.

Other formulas may be used to compute standard errors for all commonly used statistics. The calculation of confidence intervals for parameters other than the mean, however, involves concepts not discussed above.

Statistical Discrimination

The first step in approaching many taxonomic problems involves an evaluation of the observed difference between two samples. Given two samples that have different values of some statistic, the investigator may explain this difference by one of two hypotheses:

1. The populations represented by the samples are identical (with respect to this character), and the observed difference is due to chance sampling error.

2. The populations represented by the samples actually differ, and the observed sample difference is a reflection of this original difference.

In many problems the proper choice between these alternatives is intuitively evident after careful consideration of the magnitude of the difference, the number of specimens, and the variability of the samples. Although frequently not explicit in published taxonomic work, this judgment is without exception implicit in any evaluation of the difference between two samples. The process of statistical discrimination is therefore not confined to studies containing formal biometrical data.

There are many problems, however, in which subtle morphological differences are encountered in small or highly variable samples. Under these circumstances formal statistical tests should be applied, not because they achieve positive judgments, but because they permit the investigator to select the chance he is willing to take of making an erroneous decision.

Table 3 illustrates a standard statistical approach to problems of discrimination. Consider first the difference between samples 1 and 2. Statistical treatment begins with the formulation of the following hypothesis: that the populations from which the samples have been drawn have equal means. From the very nature of the problem one can never state with certainty that this hypothesis (the null hypothesis) is either true or false. But one can do the next best thing by determining the chance of rejecting the null hypothesis when it is actually true, i.e., the chance of saying that the populations are different when they are really the same.

The probability of making this error is called the *level of significance*. In taxonomic work a probability of 1 per cent or less ($P \leqq 0.01$) is commonly taken as a reliable indication that the populations actually differ. If the probability is more than 5 per cent

TABLE 3

Tests for the Significance of the Difference Between the Means of Three Hypothetical Samples

Sample	$\bar{x}$	s	N
1	20.6	2.5	10
2	24.3	2.8	12
3	18.1	2.6	8

$$t=\frac{(\bar{x}_1-\bar{x}_2)\sqrt{\frac{N_1N_2}{N_1+N_2}}}{\sqrt{\frac{(N_1-1)s_1^2+(N_2-1)s_2^2}{N_1+N_2-2}}}$$

Sample 1 vs. sample 2:

$$t=\frac{-3.7\sqrt{120/22}}{\sqrt{\frac{9(6.25)+11(7.84)}{20}}}=-3.24$$

$t_{.01}(20\ d.f.)=2.845$

$P<0.01$

Sample 1 vs. sample 3:

$$t=\frac{2.5\sqrt{80/18}}{\sqrt{\frac{9(6.25)+7(6.76)}{16}}}=2.07$$

$t_{.05}(16\ d.f.)=2.120$

$P>0.05$

($P \geq 0.05$), a real difference is usually not considered to have been established. If the probability lies between 1 per cent and 5 per cent ($0.01 < P \leqq 0.05$), the difference is normally judged to be probably significant.

The hypothesis of equal means for samples 1 and 2 may be tested by calculating the statistic t according to the formula

$$t = \frac{(\bar{x}_1 - \bar{x}_2)\sqrt{N_1 N_2/(N_1+N_2)}}{\sqrt{\dfrac{(N_1-1)s_1^2+(N_2-1)s_2^2}{N_1+N_2-2}}}. \qquad (5)$$

This formula, given by Simpson and Roe (1939, p. 211), relates the differences between the means ($\bar{x}_1 - \bar{x}_2$), the sample sizes (N_1 and N_2), and the standard deviations (s_1 and s_2) in such a way that the probability of a significant difference can be quantitatively evaluated.[1]

Note that the value of t will be large if the absolute difference between the means is large, if the samples are large, or if the standard deviations are small—as would be expected from an intuitive analysis. In this example, the absolute value of t is 3.24. To evaluate this figure it is necessary to compute the number of degrees of freedom (*d. f.*) according to the expression

$$d.f. = N_1 + N_2 - 2.$$

Entering a table of t (e.g., Snedecor, 1946, p. 65) one finds that for 20 degrees of freedom the probability of obtaining an absolute value of t larger than 2.845 is 0.01. Because the observed value of t (3.24) is greater than 2.845, the difference is judged to be significant at the 1 per cent level. In other words, one is sure 99 per cent of the time that the observed difference in length reflects a real difference in the populations. In a comparison of sample 1 with sample 3, t is calculated as 2.07. Because this value is less than 2.120, the 5 per cent level of t for 16 *d. f.*, the observed difference in sample means is judged not to be significant. In other words, as there are more than five chances in 100 that the observed difference is due to chance, one is not justified in concluding that the populations are distinct. Note that this does not prove that the populations are identical; it indicates only that on the basis of the available evidence a difference cannot be conclusively demonstrated. Given larger samples of the same populations it might in fact be possible to document a significant difference.

[1] This formula has the advantage that several related operations are summarized in one expression. Most statistical texts treat the constituent operations as units, e.g., Snedecor (1946, p. 80) and Dixon and Massey (1951, p. 103).

BIVARIATE ANALYSIS

Relative Growth

Because the paleontologist deals with the remains of organisms that have long since ceased to grow, he may easily fall into the habit of thinking of fossils in static terms. Such an attitude is inevitably fostered, rather than diminished, by the use of univariate statistical devices in the description of samples.

Bivariate analysis, on the other hand, involves the dynamic concept of relative growth. The focal point of interest here is the pattern of growth, i.e., the route by which the adult stage is attained. This method has two advantages. First, it provides a means of characterizing a sample in a way that has general taxonomic value regardless of the ability of the investigator to identify the growth stages represented in the sample. Second, by shedding light on morphogeny, bivariate analysis provides a better understanding of the underlying genetic mechanism. What is inherited, after all, is a growth pattern rather than a static adult character.

Although the study of relative growth is as old as the field of taxonomy itself, the publication of Huxley's classic "Problems of relative growth" (1932) placed the subject on a firm quantitative basis and pointed the way to a deeper understanding of many phenomena connected with organic growth. The present summary is taken largely from this work and from published studies based thereon.[2]

Consider first a scatter diagram constructed by the recording of pairs of measurements on a single ideal organism at various

[2] For general reviews of this subject, the reader is referred to Huxley (1932), Simpson and Roe (1939), Reeve (1940), Kermack and Haldane (1950), Zuckerman (1950), Olson and Miller (1951), and Kermack (1954).

stages in its growth (fig. 1A). The characters measured (x and y) might represent length and width of a brachiopod valve, height and diameter of an echinoid calyx, postorbital and preorbital length of a skull, or any other pair of variates of which the relationship during growth represents a significant morphological characteristic of the organism. The points recorded in this ideal case fall on a smoothly curving line, which thus represents a pattern or path of relative growth. Such a line is here called a *line of relative growth*, or simply a *growth line.*

The striking feature of the majority of these growth lines is this: they are very closely approximated by graphs of the equation

$$y=bx^a \tag{6}$$

where y and x stand for the variates, and a and b are mathematical parameters that take on particular values for different curves. A growth pattern of this type is termed *simple allometric.*[1] Studies on a wide variety of animals and plants have repeatedly confirmed this relationship, not only for the growth of a single organism through time, but also for the static distribution of values characterizing a population at one point in time.

Before complications that arise in studying real populations are discussed, it is advantageous to note certain mathematical features of the ideal (though commonly approximated) situations illustrated in figures 1A and 1B. The fact that the growth line in figure 1A is curved indicates that the two measured dimensions are increasing at different rates and that the ratio y/x changes as the organism grows. But because the relationship between the variates is described by the allometric equation $y=bx^a$, an exponential equation, the logarithms of the original data plot as a straight line (fig. 1B). In other words, the specific growth rates of x and y,[2] though unequal, maintain a ratio that is constant and equal to a. In special cases, where the specific growth rates are equal ($a=1$), the growth is called *isometric*, and a plot of the original data yields a straight line.

The mathematical parameters a and b are thus of critical importance in studies of relative growth. The exponent a, called the *growth ratio*, is a pure number, as it expresses the ratio between two specific growth rates. Graphically, it is the slope of the growth line on double logarithmic paper. The coefficient b, called the *initial growth index*, is the absolute value of y when x equals 1. For mathematical treatments of allometric data, it is necessary to transform the exponential equation into its linear logarithmic equivalent, thus:

$$Y=aX+B \tag{7}$$

where

$$X=\log x,\ Y=\log y,\ \text{and}\ B=\log b.$$

In certain instances the simple allometric growth relationships described above do not obtain. Deviations from the usual pattern can be easily detected by plotting observations on double logarithmic paper and observing deviations from linearity.

The calculation of growth constants for real samples is always complicated by the fact that the plotted points do not lie exactly along a smooth mathematical curve. In almost every case, however, the points tend to lie close to a simple allometric growth line. When such a tendency is evident there is no serious objection to the assumption that the population as a whole tended to follow a simple allometric pattern and that the observed dispersion or variation from the ideal condition is due to inherent biological variability, distortion, or errors of measurement. Thus the growth line of a bivariate population is analogous to the mean of a normal univariate population in that both represent a morphological norm around which observed values tend to cluster. One represents the average value actually attained by a given growth stage; the other, the average path by which the adult form was attained. Figure 1C represents a typical cluster of points around an allometric growth line. The amount of scatter increases with growth, owing to the tendency for the absolute amount of variation to be proportional to absolute size. In such a sample, where a wide

[1] For a general discussion of terminology, see Reeve and Huxley (1945, p. 123).

[2] The specific growth rate is the rate at which the logarithm of a dimension is changing. This measure gives equal weight to equal percentage increases.

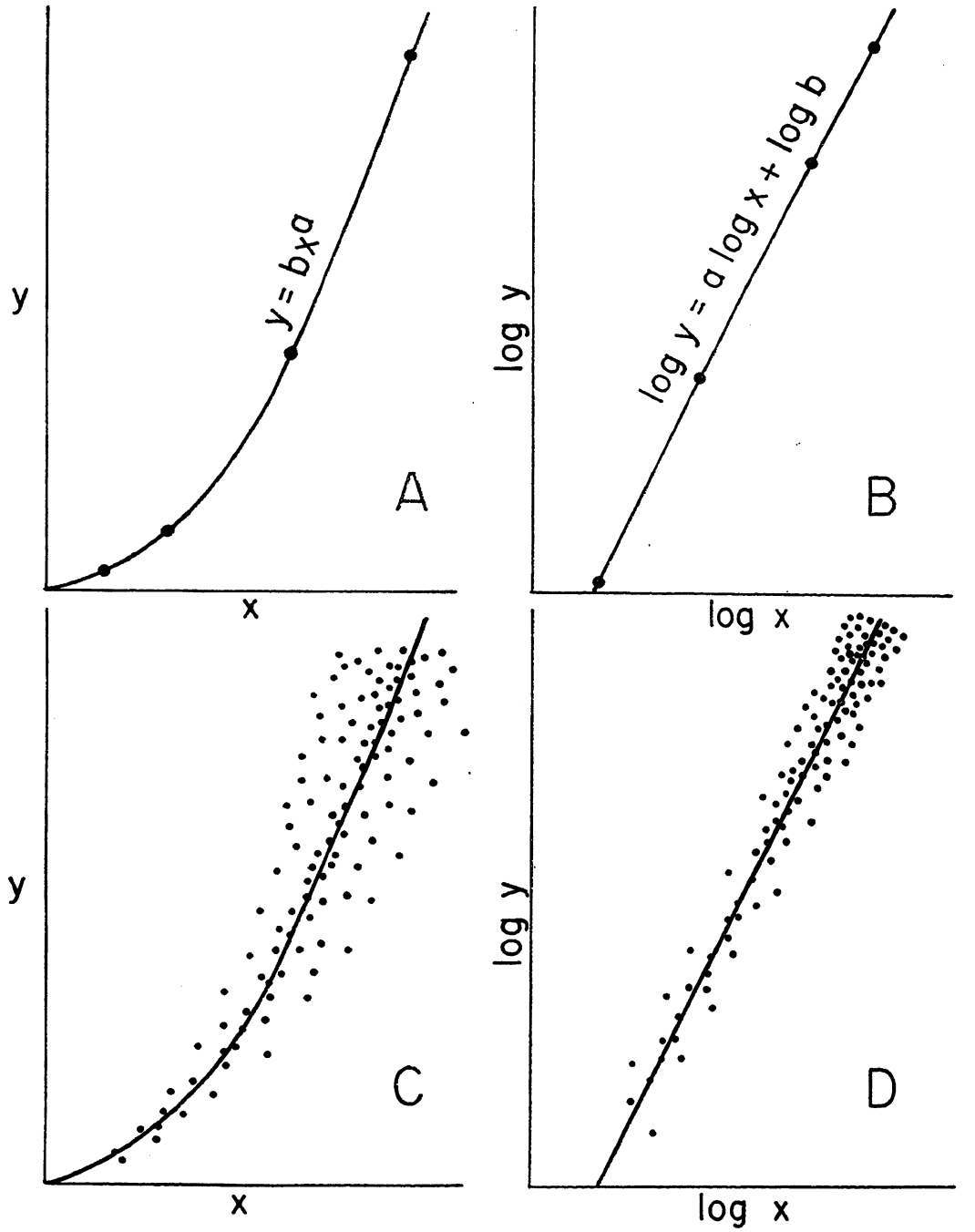

FIG. 1. Hypothetical scatter diagrams showing relative growth in two linear measurements, x and y. A. Ideal case of simple allometric growth in which the four plotted points represent growth stages in a single individual. B. Ideal case of simple allometric growth in which data of case A are plotted as log x and log y. C. Simple allometric growth as shown by a sample including individuals at various stages of growth. D. Data of case C plotted as log x and log y.

range in size is represented, there is little difficulty in recognizing the allometric nature of the growth pattern. If the logarithms of the data are plotted, as in figure 1D, the trend is straight and the amount of scatter essentially constant.

Finally, consider the effect of collecting a sample that embraces only a relatively small part of the total size range, a situation by no means uncommon in invertebrate paleontology. If the curvature of the true growth line is strong, or if the amount of scatter is small, it might still be possible to demonstrate an allometric (i.e., curved) pattern of growth. Usually, however, the curvature is so slight or the amount of scatter so large (or both) that the underlying allometric relationship cannot be convincingly demonstrated. Under these circumstances, the observed facts can be expressed by means of a linear equation of the form

$$y=ax+b. \tag{8}$$

In mathematical terms, a is the slope of the line and b the y-intercept. As b will usually not equal zero, the ratio x/y will normally vary with growth.

Estimation of Parameters of Relative Growth

Regression and Reduced Major Axis

In every series of observations made for the purpose of recording patterns of relative growth there is some inherent scatter or variability. In order that allometric studies may proceed, therefore, it is necessary that one choose some standard method for constructing the growth line that best approximates, or "fits," the observed trend. Two statistics are required to define this line: the growth ratio and the initial growth index. These quantities (a and b) can then be regarded as estimates of the corresponding growth parameters (α and β) of the population from which the sample was drawn.

The question naturally arises as to the proper criterion for judging goodness of fit. The simplest procedure is to plot the observations and locate by visual estimation a line that passes through the "middle" of the distribution. If there is little variation in a sample, and if the purposes of the investigation do not justify the application of more objective but slightly more time-consuming methods, this procedure may give satisfactory results. Usually, however, an objective algebraic solution to the problem is desired. At least four basically different methods for obtaining a line of best fit can be employed, as follows:

1. Major Axis: A line that minimizes the sum of the squares of the perpendicular distances from each point to the desired line is called the major axis. In figure 2 this perpendicular distance is indicated by the distance BF. On an intuitive basis, this procedure seems to be entirely reasonable. Critical analysis of the properties of this line, however, has shown that its slope changes with the unit of measurement. It is thus unsuitable for taxonomic problems (Kermack and Haldane, 1950, p. 30).

2. Regression of y on x: The regression of y on x is defined as the line that minimizes the sum of the squares of deviations from that line, the deviations being measured perpendicular to the x-axis. (In mathematical terms, this means that x is taken as the independent, and y as the dependent, variable.) In figure 2 this vertical distance corresponds to the distance AE. Regression analysis, though widely employed in the calculation of allometry, has one serious weakness: the assumption that all the dispersion involved is due to deviations in one variate. From the biological point of view this assumption is never warranted, for biological variability and observational errors are always involved in both variates (Kermack and Haldane, 1950, p. 30; Kermack, 1954, p. 488).

3. Regression of x on y: The regression of x on y is defined as the line that minimizes the sum of the squares of the deviations from that line, the deviations being measured perpendicular to the y-axis. This horizontal distance is indicated in figure 2 as the distance DK. As in the case previously described, use of this line entails unwarranted assumptions of independence.

4. Reduced Major Axis: This line minimizes the sum of the areas of the triangles formed by lines drawn from each point to the desired line and parallel with the x and y axes. In figure 2 this area is equivalent to the

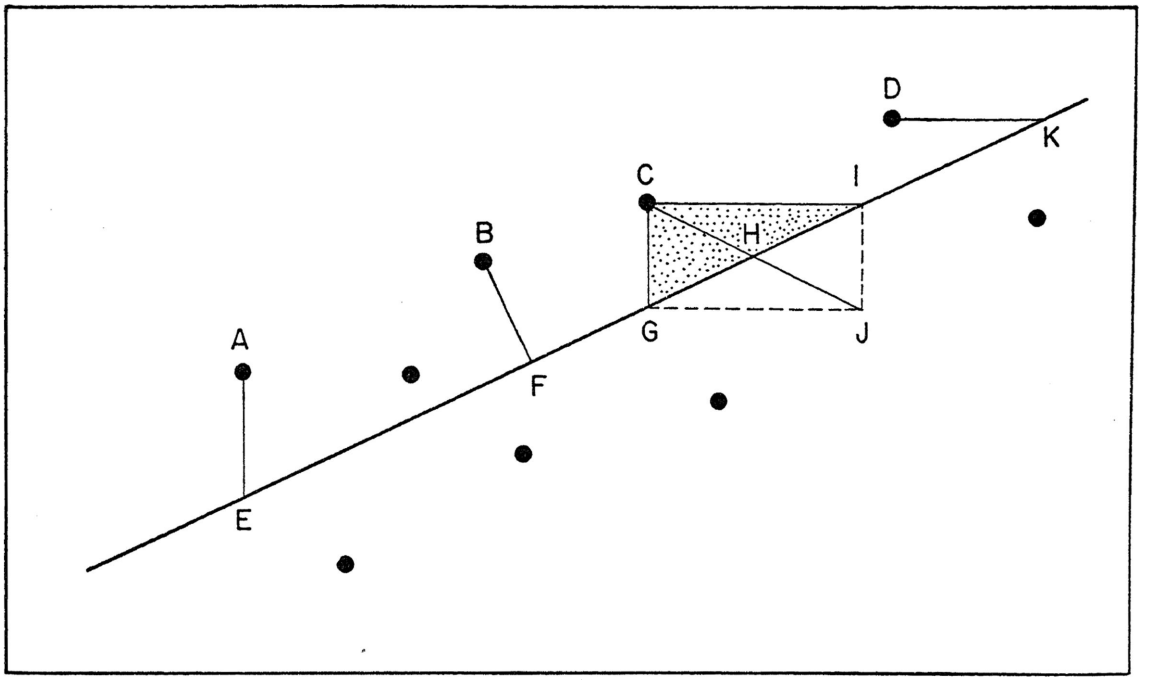

FIG. 2. Diagram to show various methods of fitting a line to a scatter of points. A regression line y on x minimizes the sum of the squares of the deviations measured as AE. A regression line x on y minimizes the corresponding sum of deviations measured as DK. A major axis minimizes the sum of the squares of the deviations measured as BF. A reduced major axis minimizes the sum of the areas of triangles GCI.

triangle GCI. Geometrically this is the same as minimizing the sum of the products of the legs of these triangles (e.g., $CG \times CI$). From the strictly biological point of view it is difficult to find an *a priori* justification for the use of this line in computing allometry. However, from both theoretical and empirical biometrical studies, the reduced major axis emerges as the best available statistical tool. (See especially Teissier, 1948; Kermack and Haldane, 1950; Kruskal, 1953; and Kermack, 1954.) To summarize some of the arguments for this line: (1) it makes no assumptions of independence; (2) it is invariant under change of scale; (3) it is simple to compute; and (4) results obtained from its use are intuitively more reasonable than corresponding results obtained from regression analysis, as is shown below.

On figure 3 growth lines characterizing two samples of the brachiopod *Pholidostrophia gracilis* have been constructed by the use of reduced major axes and regression lines. The growth patterns recorded by the reduced major axes deviate only slightly from colinearity, and this deviation is statistically not significant at the 1 per cent level. Growth lines based on regression analysis, however, are strikingly different for the same pair of samples. From an inspection of the points represented by these lines (fig. 10), it is clear that there is no demonstrable difference between the growth patterns—a judgment that is supported by laboratory examination of the collections. Generalizing from this example, one is justified in concluding that the reduced major axis is to be preferred over regression lines in dealing biometrically with problems of relative growth.[1]

Having concluded that the reduced major

[1] As the amount of scatter or dispersion (measured by r) decreases, all the lines discussed above tend to coincide. If the scatter is small enough (say $r = 0.95$), the results obtained from regression analysis will be essentially the same as those from the reduced major axis.

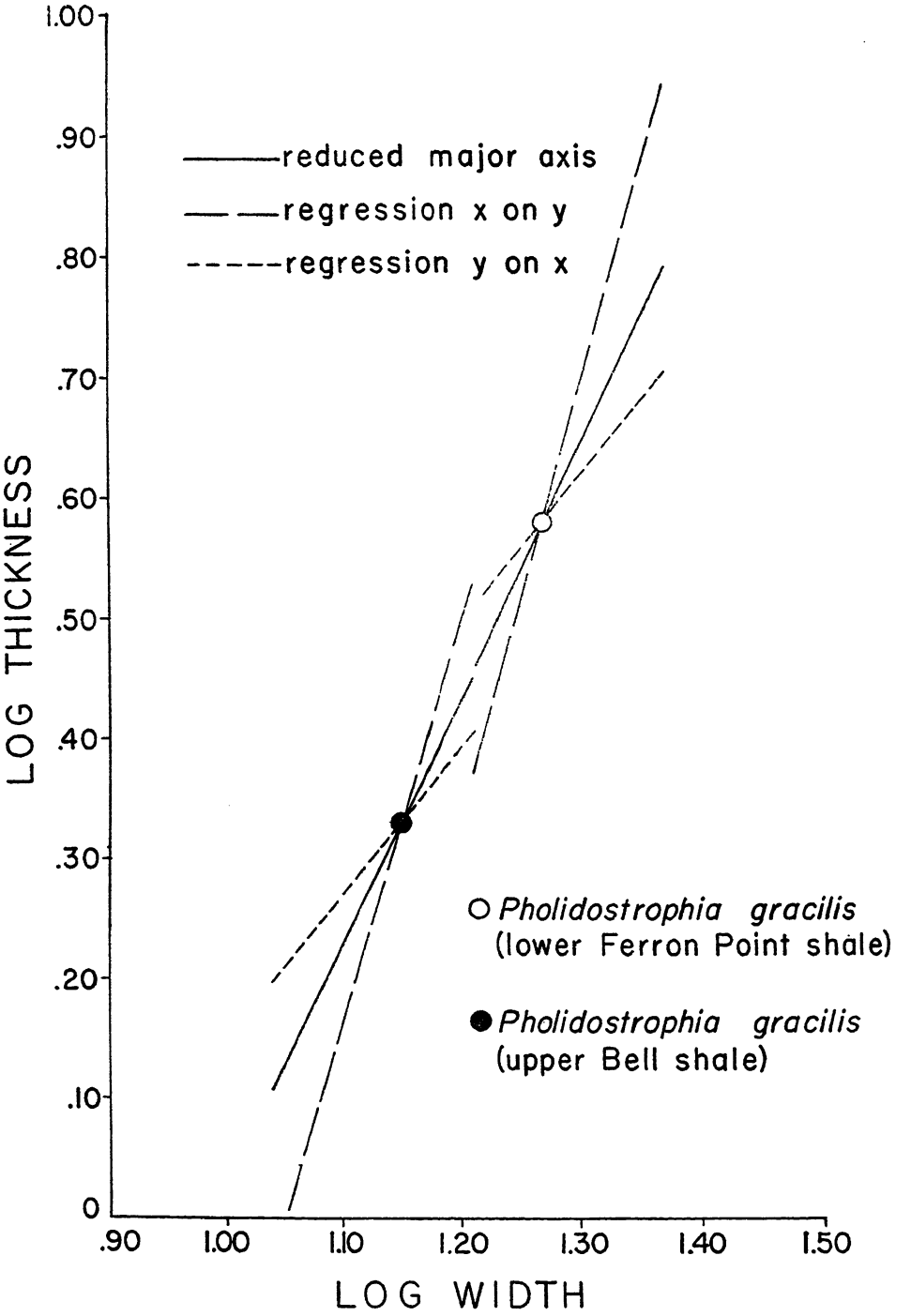

FIG. 3. Samples of *Pholidostrophia gracilis* characterized by regression lines and the reduced major axis for log-width and log-thickness. Length of some lines less than observed range. Plotted points correspond to joint means of samples. Original data plotted on figure 10. Ferron Point sample is *P. g. gracilis;* Bell shale sample *P. g. nanus.*

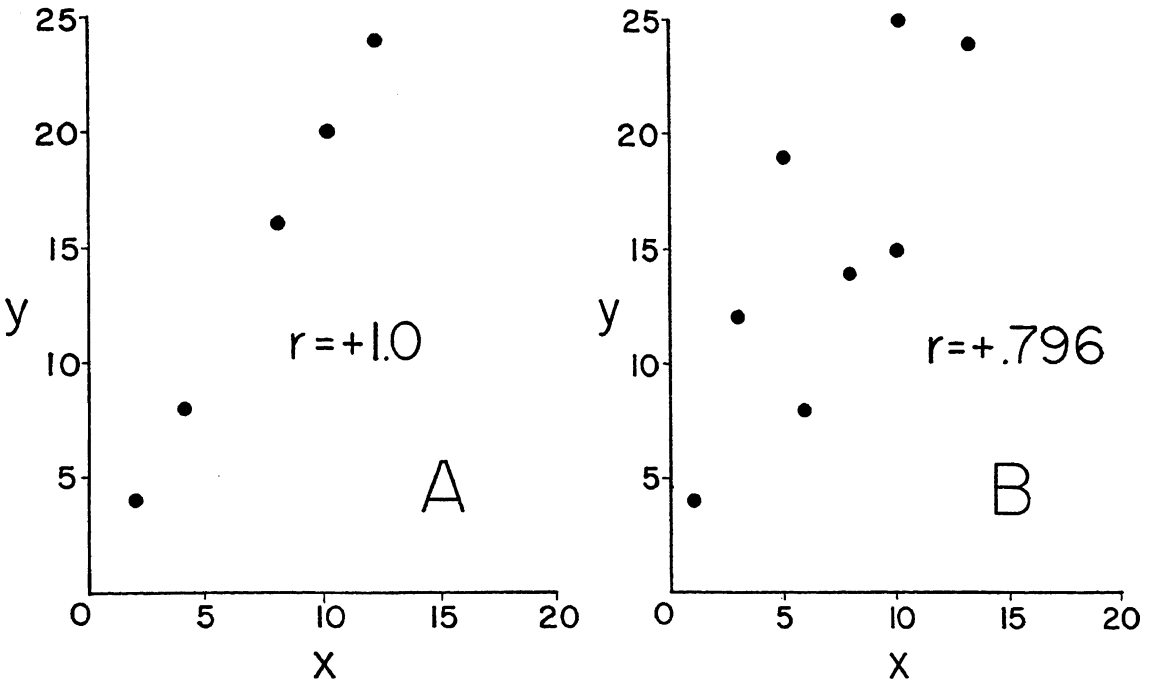

FIG. 4. Scatter diagrams illustrating the correlation coefficient r. A. Perfect linear correlation. B. Significant but imperfect linear correlation. Plotted points correspond to data in table 4.

axis represents the best approach to the problem of calculating growth coefficients, one now must consider the formulas for the appropriate statistics and standard errors. First, however, some understanding of another statistic, the correlation coefficient, is necessary. For readers unfamiliar with this statistic a brief description is included below.

COEFFICIENT OF CORRELATION

Consider the following pairs of observations on two variates, x and y:

x	y
2	4
4	8
8	16
10	20
12	24

Because a given change in one variate is accompanied by an exactly proportional change in the other, the two variates are perfectly correlated. This relationship is shown graphically in figure 4A by the fact that the plotted points fall exactly on a straight line which trends at an angle to both axes. Figure 4B illustrates another set of observations in which the points tend to cluster along a straight line at an angle to both axes. Here the correlation (statistical, but not necessarily functional) is less marked than in the former case. The *coefficient of correlation*, symbolized by the letter r, is a measure of the strength of linear correlation. The coefficient varies between one and zero, with a value of one indicating perfect linear correlation, zero an absence of any correlation, and intermediate values corresponding to intermediate levels.[1] The coefficient takes on positive values if an increase in one variate is associated with an increase in the other, and negative values if an increase in one is associated with a decrease in the other.

The formula for the calculation of r from raw data may be given as

$$r=\frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}. \qquad (9)$$

[1] A word of caution is in order with regard to the interpretation of r. While it is true that a correlation giving a value of $r=0.8$ is better than one giving a value of $r=0.4$, it is not true that the former is twice as good as the latter. Any statistical text will provide a more detailed explanation of the meaning and use of this coefficient.

TABLE 4

CALCULATION OF THE CORRELATION COEFFICIENT BY THE LONG METHOD FOR A HYPOTHETICAL SET OF DATA

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
1	4	−6	−11	+66	36	121
3	12	−4	− 3	+12	16	9
5	18	−2	+ 3	− 6	4	9
6	8	−1	− 7	+ 7	1	49
8	14	+1	− 1	− 1	1	1
10	15	+3	0	0	9	0
10	25	+3	+10	+30	9	100
13	24	+6	+ 9	+54	36	81

$\sum(x) = 56$ $\quad \sum(x-\bar{x})(y-\bar{y}) = 162$

$\sum(y) = 120$ $\quad \sum(x-\bar{x})^2 = 112$

$\bar{x} = 7$ $\quad \sum(y-\bar{y})^2 = 370$

$\bar{y} = 15$ $\quad N = 8$

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} = \frac{162}{\sqrt{(112)(370)}} = +0.796$$

The calculation of r for the data illustrated in figure 4B is shown in table 4. Shorter methods for obtaining r are described below.

CALCULATION OF STATISTICS

The basic formulas for the computation of the growth ratio and the initial growth index are given by Kermack and Haldane (1950). For a straight line of the form $y = ax + b$,

$$a = \frac{s_y}{s_x} \tag{10}$$

$$\sigma_a = a\sqrt{\frac{1-r^2}{N}} \tag{11}$$

$$b = \bar{y} - \bar{x}a \tag{12}$$

where

a = growth ratio
b = initial growth index
σ_a = standard error of a
$\bar{x}$ = mean of x
$\bar{y}$ = mean of y
s_x = standard deviation of x
s_y = standard deviation of y
r = correlation coefficient.

Kermack and Haldane also give the formula for the standard error of b, but this quantity is rendered unnecessary by certain techniques described below. These formulas should be used only if a plot of the original data on arithmetic graph paper indicates that a straight line will fairly represent the trend. This will be true (1) in the relatively rare instances in which growth proceeds isometrically, or (2) in the common situations in which the variability is too great and the range of observations too small for an allometric relationship to be demonstrated.

If the original data cannot be represented with sufficient accuracy by a straight line, the trend can almost always be approximated by a line of the form $y = bx^a$. The appropriate computations may be carried out in two ways:

1. By transformation of the original data into logarithmic form. This operation can be symbolized by the notation

$$X = \log x$$
$$Y = \log y.$$

If relative growth does in fact follow the pattern $y = bx^a$, a plot of X and Y will produce a straight trend. This trend can in turn be described by the linear equation $Y = aX + B$ where B is defined as $\log b$. Then

$$a = \frac{s_Y}{s_X} \tag{13}$$

$$\sigma_a = a\sqrt{\frac{1-r'^2}{N}} \tag{14}$$

$$B = \overline{Y} - \overline{X}a \tag{15}$$

where

s_Y = standard deviation of Y
s_X = standard deviation of X
$\overline{X}$ = mean of X
$\overline{Y}$ = mean of Y
r' = correlation coefficient of X and Y.[1]

If it is desired to write the equation in terms of the original data (in the form $y=bx^a$), it is only necessary to compute $b=\text{antilog } B$.

2. By approximation formulas. Kermack and Haldane (1950) give formulas for computing constants of the logarithmic equation $Y=aX+B$ and the logarithmic correlation coefficient r' directly from the (arithmetic) statistics $\bar{x}$, $\bar{y}$, s_x, s_y, and r. These formulas usually give only approximate results, because they are derived on the assumption that the logarithms are normally distributed. Kermack and Haldane indicate that moderate deviations from log-normality do not seriously affect the accuracy of results. In the experience of the present writer, however, results obtained by the approximation formulas are unsatisfactory for the small, skewed samples often encountered in paleontology.

Statistical Discrimination

In a previous section of this paper the problem of statistical discrimination in univariate samples is discussed. When bivariate data are dealt with, there are analogous problems of statistical discrimination. Here, a pattern of relative growth is computed for each of two samples, and the question is: Do the sample growth patterns differ significantly? As before, there are many cases in which a graphic plot of the data indicates at a glance that the lines of relative growth are so far from coincidence that a significant difference may be regarded as certain without formal statistical evaluation. In other circumstances, however, particularly when samples are small and display considerable variability, an objective evaluation of probabilities is desirable.

If two samples characterized by growth lines of the form $y=ax+b$ are given, the problem of statistical discrimination may present itself in different ways. Four representative situations are illustrated in figure 5.

[1] This definition is employed here to avoid confusion with the usual meaning of r. Kermack and Haldane, however, reverse this symbolism and use r for the correlation coefficient of the logarithms.

Figure 5A illustrates the general case in which growth patterns differ in both slope and position, as indicated by differences in a and b, respectively. Figure 5B illustrates a case in which two growth lines have the same slope ($a_1=a_2$), but differ in position ($b_1\neq b_2$). In figure 5C the slopes differ but the difference in position, as measured by the initial growth index, is zero. Figure 5D represents a situation in which two samples follow an identical growth pattern and thus differ fundamentally only in size. Note, however, that because the initial growth indices do not equal zero ($b_1=b_2\neq 0$), the two samples will exhibit different y/x ratios.

Statistical discrimination in bivariate analysis begins with the formulation of the null hypothesis (represented in fig. 5D) that two samples were drawn from populations having identical growth patterns. The procedure recommended here for testing the null hypothesis is a shortened and simplified version of a method developed by Kermack (1954). Four steps are involved.

Step 1: Characterize each sample by computing the basic bivariate statistics (see formulas 1, 2, and 9).

Sample 1	Sample 2
N_1	N_2
$\bar{x}_1$	$\bar{x}_2$
$\bar{y}_1$	$\bar{y}_2$
s_{x_1}	s_{x_2}
s_{y_1}	s_{y_2}
r_1	r_2

Step 2: Compute for each sample the statistics necessary to define a line of relative growth of the form $y=ax+b$ (see formulas 10, 11, and 12).[2]

Sample 1	Sample 2
a_1	a_2
σ_{a_1}	σ_{a_2}
b_1	b_2

Step 3: Test the hypothesis that the growth lines characterizing the populations from which the two samples were drawn

[2] If a line of the form $y=bx^a$ is required, the logarithms of the original data should be used and a line of the form $Y=aX+B$ computed. A procedure exactly analogous to the one described in subsequent paragraphs can then be followed.

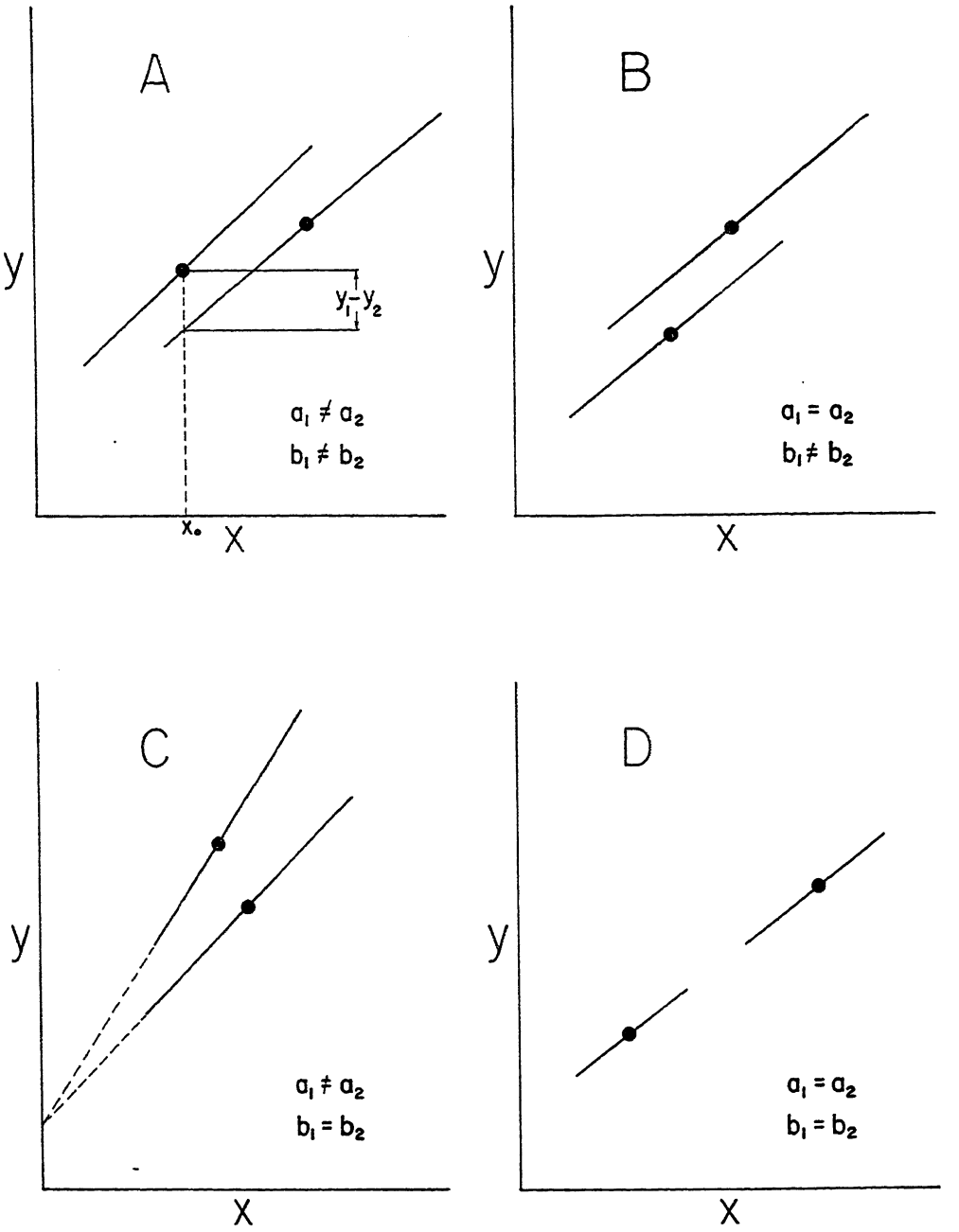

FIG. 5. Diagrams illustrating various ways in which two samples can differ in patterns of relative growth. Lines represent isometric growth of form $y=ax+b$. Length of solid lines corresponds to observed sample ranges. Points represent joint means of samples. A. Two lines with slight difference in slope and considerable difference in position. Vertical distance between lines, y_1-y_2, shown along the ordinate x_0, when $x_0=$ mean of one sample. B. Two lines differing in position but not slope. C. Two lines differing in slope but having identical values of initial growth index b. D. Two colinear line segments.

have the same slope. This can be done by calculating the statistic z, where

$$z=\frac{a_1-a_2}{\sqrt{\sigma_{a_1}^2+\sigma_{a_2}^2}}. \quad (16)$$

If the observed value of z is less than 1.96, the probability (P) that the observed difference (a_1-a_2) arose by chance is greater than 0.05. If the observed value of z is greater than 1.96, the probability that so great a difference was observed purely by chance is 0.05 or less. If z is greater than 2.58, the corresponding probability is 0.01. Other probability levels can be obtained from tables of z in any standard statistical text.[1]

In most taxonomic work, if P is less than 0.05 ($z>1.96$) the hypothesis of equal slopes is rejected, and the observed difference is considered statistically significant. Under these circumstances there is no need to apply step 4, unless additional information on the nature of the difference is sought. If, however, P is greater than 0.05 ($z<1.96$), the hypothesis of identical slopes is normally accepted, and a test is made for the significance of positional differences as indicated below.

STEP 4: Test the hypothesis that the growth lines characterizing the populations from which the two samples were drawn are identical over the size ranges represented in the samples. Clearly, this hypothesis is to be tested only if step 3 has demonstrated no significant difference in slope. At least three statistical tests of this hypothesis are available:

a. A test based on the standard error of the initial growth indices. The rather complicated expression for this standard error is given by Kermack and Haldane (1950, p. 40). This method in effect tests the significance of the difference between the computed growth lines where they cross the y-axis. Graphically it is evident, and mathematically it has been proved (Kermack, 1954, p. 410), that whenever the sample means are large multiples of the unit of measurement, as is generally true, a great deal of inaccuracy is inevitable in the estimate of the initial growth indices. This method is therefore less efficient than procedures discussed below.

b. A test based on the assumption that the slopes of the original populations (α_1 and α_2) are actually identical (Kermack, 1954, p. 409). The procedure is to form a combined estimate of the true slope ($\dot{a}$) based on data from both samples and then to test the significance of the distance between two theoretical parallel lines having slope $\dot{a}$. This method has the advantage that the distance between the two lines is rendered constant. Essentially the only disadvantage is that the computations are laborious.

c. A test based on the assumption that the slopes of the original populations (α_1 and α_2) actually differ and that the observed difference (a_1-a_2) is an estimate of the true difference. The test described in step 3 has shown that there is no significant difference in slopes[2]; hence the amount of error introduced by this assumption would normally be small. Because the mechanics of this test are quite simple, it is here recommended as the best generally applicable technique for testing positional differences. The procedure is as follows:

1. Choose some value of x, say x_0, for which it is desired to test the observed difference between the growth lines. Different values of x_0 may be selected according to the nature of the problem. For example, if the investigator is seeking to show that a significant difference in growth pattern exists, he will normally choose a value x_0 such that the vertical distance between the growth lines (y_1-y_2) is at a minimum for the range of x values represented in the samples. Conversely if the investigator is seeking to demonstrate that there is no significant difference in growth pattern, he will normally choose a value of x_0 such that the vertical distance is at a maximum.

2. Test the hypothesis that, for the value $x=x_0$, the true growth lines of the populations coincide—in other words, that y_1-y_2

[1] Strictly speaking, assignment of probability levels to different values of z is valid only if reasonably large samples are used, say $N_1+N_2=35$ or more. Borderline cases involving small samples must be interpreted with caution.

[2] Note that this test does not prove that the populations are identical with respect to slope, only that one's data are insufficient to prove that a real difference exists.

$=0$ when $x_1=x_2=x_0$. For this purpose, compute the statistic z, where

$$z=\frac{x_0(a_1-a_2)+(b_1-b_2)}{\sqrt{\sigma_{a_1}^2(x_0-\bar{x}_1)^2+\sigma_{a_2}^2(x_0-\bar{x}_2)^2}}. \quad (17)^1$$

As above, if z is greater than 1.96, the difference is taken to be significant at the 5 per cent level. If z is less than 1.96, the observed difference will usually not be accepted as significant.

If the observed differences in slope are slight, and if $\bar{x}_2 \gg \bar{x}_1$, it may be possible to set x_0 equal to $\bar{x}_1$ or $\bar{x}_2$ without seriously affecting the validity of the test. This permits a saving in computation, as is evident from formula 17. For example, if x_0 is designated as $\bar{x}_1$ (as in fig. 5A), then

$$z=\frac{\bar{x}_1(a_1-a_2)+(b_1-b_2)}{\sigma_{a_2}(\bar{x}_1-\bar{x}_2)}.$$

Concept of Variability in Bivariate Samples

Biometrical characterization of any sample should include a measure of the amount of variation displayed. Two aspects of variation must be distinguished: *absolute variation*, recorded in terms of the units of measurement; and *relative variation*, the absolute variation expressed as a percentage of average size.

It will be helpful to review briefly the concepts of relative and absolute variation as applied to univariate data (fig. 6A). The best objective measure of absolute variation in this case is s_x, the standard deviation. The standard deviation is converted into a measure of relative variation (V) when it is stated as a percentage of the mean.

The coefficient V is useful in paleontology because organisms tend to exhibit an absolute variation that is proportional to absolute size. By computing V for two samples of different average size, therefore, one can compare inherent morphological variability, provided that each sample is homogeneous with respect to growth stage. Usually, however, individuals in a collection of fossil invertebrates vary widely in age, and objective criteria are lacking to distinguish growth stages. Under these circumstances V is unrelated to inherent biological variability, because its value is determined primarily by geological factors extrinsic to the organisms.

The difficulty just mentioned can be avoided if bivariate data are assembled and variation is defined in terms of the dispersion of values about the line of relative growth. This definition of variability is illustrated in figure 1. In figures 1A and 1B, the amount of dispersion about the growth line, and by definition the amount of variation, is zero. The points represented in figures 1C and 1D, on the other hand, are scattered about the growth line. In such a case, the amount of scatter is taken as an index of the amount of morphological variation. Statistical formalization of this concept takes two basic forms, as follows:

1. Dispersion Around a Regression Line: Figure 6B illustrates the concepts involved in the computation of the amount of dispersion around a line of relative growth when a regression line (in this case the regression of x on y) is used to represent the data. As y is taken as the independent variate, all the dispersion is attributed to deviations in the x direction. Thus the appropriate measure of absolute dispersion is defined by the standard deviation of the horizontal distances (d_x) computed from each point to the regres-

[1] This expression is derived as follows. The numerator is equal to y_1-y_2, the vertical distance between the lines when $x_1=x_2=x_0$. From equation 8,

$$y_1=a_1x_0+b_1$$

and

$$y_2=a_2x_0+b_2.$$

Hence

$$y_1-y_2=x_0(a_1-a_2)+(b_1-b_2).$$

The denominator is the standard error of the distance y_1-y_2, which may be symbolized

$$\sigma_{y_1-y_2}.$$

The formula for this standard error is given by Kermack (1954, p. 410) as

$$\sigma_{y_1-y_2}^2=\frac{s_{y_1}^2}{N_1}(1-r_1^2)\frac{(x_0-\bar{x}_1)^2}{s_{x_1}^2}+\frac{s_{y_2}^2}{N_2}(1-r_2^2)\frac{(x_0-\bar{x}_2)^2}{s_{x_2}^2}.$$

From (10) and (11),

$$\sigma_a=a\sqrt{\frac{1-r^2}{N}}=\frac{s_y}{s_x}\sqrt{\frac{1-r^2}{N}}.$$

Hence

$$\sigma_{y_1-y_2}=\sqrt{\sigma_{a_1}^2(x_0-\bar{x}_1)^2+\sigma_{a_2}^2(x_0-\bar{x}_2)^2}.$$

Note that if $x_0=\bar{x}_1$,

$$\sigma_{y_1-y_2}=\sigma_{a_2}(\bar{x}_1-\bar{x}_2).$$

sion line. This standard deviation, called the standard error of estimate ($s_{x \cdot y}$ in this case; $s_{y \cdot x}$ if the deviations are measured in the y direction), is the logical measure of absolute variation about a regression line. To minimize the work of computation, the formula

$$s_{x \cdot y} = s_x \sqrt{1 - r^2} \tag{18}$$

should be used.[1]

Having obtained a measure of absolute variation around the regression line, one can proceed to define a measure of relative variation analogous to the coefficient of variation. This is necessary because the amount of dispersion around a line of relative growth normally increases with increasing size (fig. 1C). In fact it is generally approximately true that the amount of dispersion is proportional to size, as noted by Burma (1948, p. 748) and Klauber (1943, p. 54). This relationship is demonstrated by the tendency for scatter diagrams constructed for the logarithms of linear measurements to exhibit constant dispersion (see fig. 1D).

It will prove useful in the analysis of more complex situations to develop a dynamic interpretation of the situation in figure 6B. For this purpose the biological equivalent of the X-Y plane is considered to be a bivariate morphological field in which any particular observed combination of characters defines a point. The cluster of points in figure 6B thus represents a group of individuals each of which has moved from the origin to its observed position. The average path of growth through this field is taken (in this case) as the regression line x on y, and the average position achieved by the group is the joint mean of x and y, the point $(\bar{x}, \bar{y})$. In order to achieve a measure of relative dispersion it is natural to compare the absolute amount of variation in the x-direction ($s_{x \cdot y}$) with the average distance traveled in the x-direction by the entire group ($\bar{x}$). This ratio, expressed in per cent, is here called the *coefficient of relative dispersion from regression*, $D_{x \cdot y}$. Thus

$$D_{x \cdot y} = \frac{100 s_{x \cdot y}}{\bar{x}} . \tag{19}$$

[1] This is accurate enough for large samples. For small samples, the result should be multiplied by the factor

$$\sqrt{(N-1)/(N-2)}.$$

For the regression of y on x,

$$D_{y \cdot x} = \frac{100 s_{y \cdot x}}{\bar{y}} . \tag{20}$$

Both $D_{x \cdot y}$ and $D_{y \cdot x}$ have a standard error (σ_D) given by

$$\sigma_D = \frac{D}{\sqrt{2N}} . \tag{21}$$

Klauber (1943, p. 31), describing variation in rattlesnakes, uses the symbol "V" for the coefficient of relative dispersion. Although it is true that both V and D measure variation, the bases for defining variability are quite different, and it seems wise to provide distinct terms and symbols.

In the interpretation of $D_{y \cdot x}$ and $D_{x \cdot y}$ it is necessary to bear in mind that use of these coefficients involves the assumption that all the variation is in the y or x directions, respectively. As this is never true, it follows that these statistics systematically overestimate the amount of variation actually present in any character. Provided that samples are treated uniformly, however, and that the underlying assumptions are understood, the coefficient of relative dispersion from regression is a valuable addition to biometry.

2. Dispersion Around Reduced Major Axis: By the use of the reduced major axis it is possible to derive measures of absolute and relative variation that do not involve assumptions of independence. Consider the scatter of points around the reduced major axis in figure 6C. Each point deviates from the line by a certain horizontal distance (d_x) and a certain vertical distance (d_y). One can therefore measure the total dispersion in the bivariate morphological field by computing the vector sum of d_x and d_y. This vector sum is equal to the diagonal distance d. The total dispersion about the reduced major axis can then be expressed as the standard deviation (s_d) of these diagonal distances. From Teissier (1948, p. 30) we have

$$s_d = \sqrt{2(1-r)(s_x^2 + s_y^2)}. \tag{22}$$[2]

[2] Teissier actually gives the variance of half of the diagonal distance as

$$(1-r)(s_x^2 + s_y^2)/2,$$

whence the standard deviation of the whole diagonal is derived.

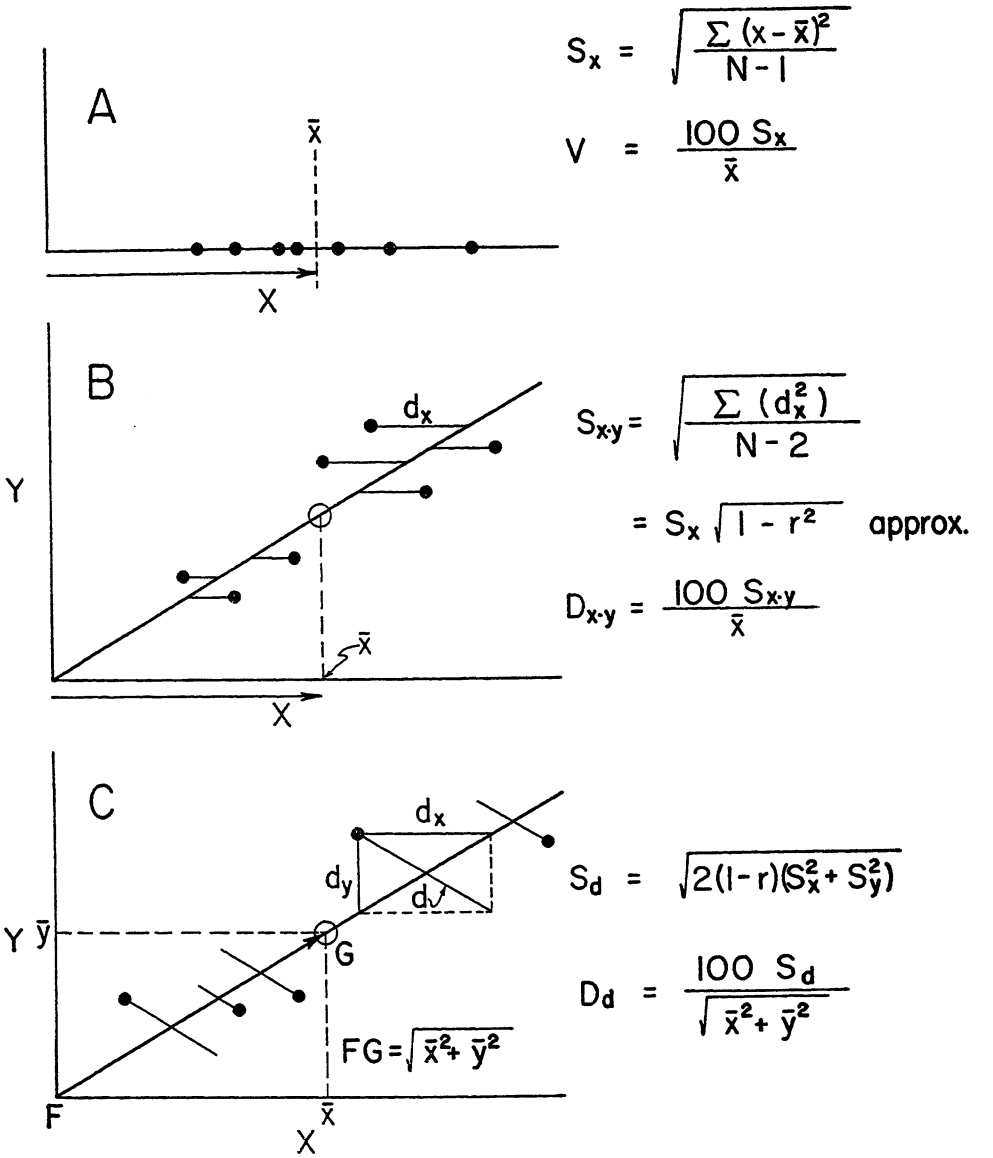

FIG. 6. Diagrams to illustrate concepts of absolute and relative variation in univariate and bivariate analysis. Symbols used in formulas explained in text. A. Univariate case. Observed values of variate x plotted on abscissa. No ordinate scale. Absolute dispersion measured by standard deviation (s_x). Length of arrow corresponds to average distance traveled in univariate morphological field. Relative dispersion measured by V. B. Bivariate case. Absolute dispersion measured by standard deviation ($s_{x.y}$) of distances (d_x) measured from regression of x on y. Length of arrow corresponds to average distance traveled horizontally in bivariate morphological field. Relative dispersion measured by $D_{x.y}$. C. Bivariate case. Absolute dispersion measured by standard deviation (s_d) of distances (d) measured from reduced major axis. Length of arrow corresponds to average distance traveled in bivariate morphological field. Relative dispersion measured by D_d.

This statistic can be used as an objective measure of absolute variation. Moreover, because a deviation in either y or x affects s_d, this statistic is a measure of shape variability, whereas V, $D_{x \cdot y}$, and $D_{y \cdot x}$ measure variability in only one dimension.

To arrive at a meaningful measure of relative dispersion one has only to compare the absolute variation s_d with the average distance traveled by the sample in the bivariate morphological field. This distance can be defined as the length of a line (FG) extending from the origin to the joint mean $(\bar{x}, \bar{y})$, or

$$\sqrt{\bar{x}^2+\bar{y}^2}$$

The suggested measure of relative dispersion D_d is thus given by

$$D_d = \frac{100 s_d}{\sqrt{\bar{x}^2+\bar{y}^2}} = 100 \sqrt{\frac{2(1-r)(s_x^2+s_y^2)}{\bar{x}^2+\bar{y}^2}} . \quad (23)$$

The quantity D_d can be called the *coefficient of relative dispersion about the reduced major axis*. Statistically, it is the ratio, expressed in per cent, between the standard deviation of the vector sums of the deviations from the reduced major axis in the x and y directions and the distance from the origin to the joint mean of the sample. Biologically, it expresses the amount of shape variation as a proportion of the average shape attained by the sample.

SUMMARY AND CONCLUSIONS

THE PRESENT PAPER outlines some of the statistical methods and concepts that can be applied to problems involving invertebrate fossils. In the application of these methods it must be remembered that statistical conclusions are never better than the data on which they are based. Provided that undistorted specimens are measured, however, and that the characters used represent the best distinguishing characteristics, the resulting statistical conclusions will be biologically valid.

Perhaps the chief disadvantage of statistical methods is that they are time-consuming. Before dismissing biometrical techniques on this basis, however, the investigator should weigh the advantages of biometrical procedures against the cost in terms of time. With suitable computational short cuts (described below) and a calculating machine, a bivariate sample of 50 specimens can easily be characterized statistically in 30 minutes. The average time consumed in testing similarities and differences between two samples is about 15 minutes. The corresponding figures when calculation is performed without a machine are approximately 60 and 20 minutes, respectively.

The choice of a technique for any particular problem will depend primarily on the nature of the paleontological samples at hand. Univariate statistical techniques can be meaningfully applied if it can be determined that the entire sample (or reasonably large subsample) consists of individuals at the same growth stage. The sample is then adequately characterized by sample size (N), mean ($\bar{x}$), and standard deviation (s). A more elaborate description would include the coefficient of variation (V), the standard error of the mean ($\sigma_{\bar{x}}$), and the observed range (OR). With these statistical data it is possible not only to test the hypothesis that the means of two populations are the same, but also to make a reasonable estimate of the range of variation in a population (see Simpson, 1941).

Bivariate statistical techniques can be applied when objective criteria for distinguishing growth stages are lacking. The following step-by-step procedure is here recommended:

1. PRELIMINARY EVALUATION

Identify pairs of logically related quantitative characters likely to be taxonomically significant.

2. GRAPHIC ANALYSIS

Construct a scatter diagram of the original data. The advantages in doing this are twofold. First, differences between growth patterns that are obviously significant can be detected at once. Second, growth patterns can be identified as isometric or allometric. If growth is allometric, the original data must be transformed into logarithms and a line of the form $Y=aX+B$ fitted. Fortunately, most paleontological samples can be treated satisfactorily by the isometric form of the growth equation, $y=ax+b$.

3. STATISTICAL CHARACTERIZATION

For each pair of variates compute the following statistics[1]:

N = number of pairs of measurements
$\bar{x}$ = mean of x (formula 1)
$\bar{y}$ = mean of y (formula 1)
s_x = standard deviation of x (formula 2)
s_y = standard deviation of y (formula 2)
r = correlation coefficient (formula 9)
OR_x = observed range of x

Publication of these basic bivariate statistics insures that the sample has been adequately characterized. From these quantities all the other measures discussed in this paper can be computed by simple operations. For special purposes it may be desirable to record some of the following:

a = slope of the growth line (formula 10)
σ_a = standard error of the slope (formula 11)
b = initial growth index (formula 12)
$D_{x \cdot y}$ = coefficient of relative dispersion from regression x on y (formula 19); alternatively, $D_{y \cdot x}$ (formula 20)
$\sigma_{D_{x \cdot y}}$ = standard error of $D_{x \cdot y}$ (formula 21)
D_d = coefficient of relative dispersion from the reduced major axis (formula 23); standard error unknown.

[1] If statistics are computed from the logarithms of the original data, the symbol x is replaced by X; y by Y; r by r'; and b by B.

4. Statistical Discrimination

Whenever there is reasonable doubt that morphological differences between two samples (except for differences in absolute size) are statistically significant, apply the following two-step procedure.

Step 1: Test for significant difference in the slope of the growth lines by computing the absolute value of z, where

$$z=\frac{a_1-a_2}{\sqrt{\sigma_{a_1}{}^2+\sigma_{a_2}{}^2}}.$$

Then, if z is greater than 1.96, the observed difference is significant on the 5 per cent level. If z is less than 1.96, the difference is usually judged to be not significant. If no statistically significant difference is shown, proceed to step 2.

Step 2: Test for difference in the position of the growth lines. By inspection, choose some value of x (x_0) for which it is desired to test the observed distance between the two growth lines (see p. 237). If possible, let x_0 equal the mean of one sample (say $\bar{x}_1$) and compute z from

$$z=\frac{\bar{x}_1(a_1-a_2)+(b_1-b_2)}{\sigma_{a_2}(\bar{x}_1-\bar{x}_2)}.$$

Evaluate z as discussed above.

If it is desired to test for positional difference at some point x_0 not equal to $\bar{x}_1$ or $\bar{x}_2$, compute z as

$$z=\frac{x_0(a_1-a_2)+(b_1-b_2)}{\sqrt{\sigma_{a_1}{}^2(x_0-\bar{x}_1)^2+\sigma_{a_2}{}^2(x_0-\bar{x}_2)^2}}$$

and evaluate as above.

5. Taxonomic Discrimination

Evaluate taxonomically the morphological differences shown to be statistically significant.

ILLUSTRATIVE EXAMPLES

EXAMPLE 1. CHARACTERIZATION OF A SAMPLE OF *Pholidostrophia gracilis gracilis* WITH REGARD TO LENGTH AND WIDTH

TABLE 5 contains a series of measurements of length and width on the Devonian brachiopod *Pholidostrophia gracilis gracilis* Imbrie. A convenient and rapid method for computing the basic bivariate statistics is described below.

1. Enter the data on a correlation form of the type illustrated in figure 7.[1] Consider x and y as the width and length, respectively. Vertical columns in the main body of the table represent 25 class intervals along the x axis, and the 25 horizontal rows represent class intervals along the y axis. Three different sets of class intervals are printed on the form for convenience, with blank spaces left for the entry of any other set as required. The printed scale running from 1 to 25 is chosen as suitable for the data in this example. In some problems it may be necessary to group the data in order to fit the form. Such grouping will, of course, introduce small inaccuracies in the results. After the scale is selected, it is necessary to note the midpoints of the lowest intervals, symbolized as GA_x and GA_y for the x and y scales, respectively, and to enter these in the appropriate places on the form. In this example, both midpoints equal one. Next, the grouping intervals, symbolized as I_x and I_y, are noted and entered in the appropriate places on the right-hand side of the form. Both intervals in this case equal one. Other midpoints and grouping intervals define other scales. The values 50.5 and 2, for example, characterize the middle of the three printed scales. Each pair of measurements on table 5 is entered by a tally mark in the appropriate box on the correlation form.

2. The total frequency in each column is then entered in the line of boxes labeled f_x. Next, the total frequency in each row is indicated in the f_y column. Finally, the total fre-

[1] The correlation form employed here has been slightly modified from one originated by P. S. Burnham and his associates at the Student Appointment Bureau, Yale University, New Haven, Connecticut.

TABLE 5

MEASUREMENTS (IN MILLIMETERS) OF LENGTH AND WIDTH OF A SAMPLE OF *Pholidostrophia gracilis gracilis*

(Lower Ferron Point shale, abandoned quarry at Rockport, Sec. 6, T. 32 N., R. 9 E., Alpena County, Michigan.)

Width	Length
14	10
15	12
15	13
16	12
16	13
17	12
17	13
17	13
17	13
17	14
17	14
17	14
17	14
17	14
18	13
18	13
18	15
18	15
18	15
18	15
18	15
18	16
18	16
18	16
19	14
19	15
19	15
19	16
19	16
19	17
19	17
19	18
20	15
20	15
20	15
20	16
20	16
20	16
20	17
20	17
20	18
21	15
21	16
21	17
21	17

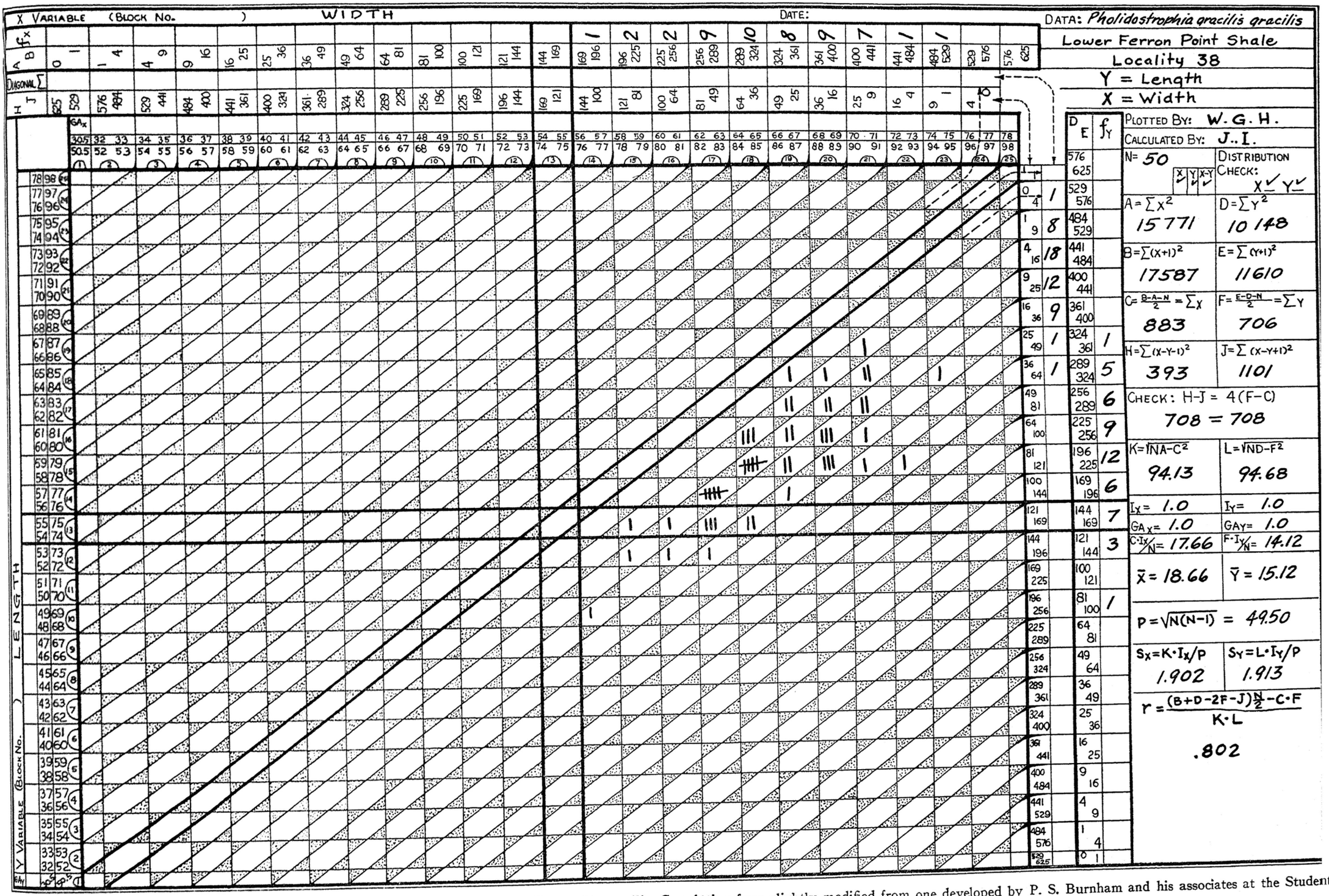

FIG. 7. Calculation of the basic bivariate statistics for a sample of *Pholidostrophia gracilis gracilis*. Correlation form slightly modified from one developed by P. S. Burnham and his associates at the Student Appointment Bureau, Yale University, New Haven, Connecticut. For explanation, see text.

TABLE 5—*continued*

Width	Length
21	18
21	18
21	19
22	15
23	18

quencies along the diagonal paths are entered in the appropriate places along both the top and the side of the form in the boxes labeled Diagonal Σ.

3. From the total x, y, and diagonal frequencies just recorded, the quantities A, B, D, E, H, and J are then calculated, as follows: The quantity A is the sum of the products of the total frequency of each class times the corresponding A factor listed in the adjacent box. Thus

$$A=(1\times169)+(2\times196)+\cdots+(1\times484)$$
$$=15{,}771.$$

This quantity is entered in the appropriate box on the right-hand side of the form. Similarly, the quantities B, D, E, H, and J are computed. Expressed in words, these operations appear to be complex and tedious. Fortunately, the actual work proceeds rapidly, especially if an automatic calculator is available. Such machines, it should be noted, are designed to cumulate automatically the sum of a series of products. Moreover, as the numbers involved are generally quite modest in size, it is possible to calculate the pair of quantities A and B simultaneously. This may be done by entering the A factor on one side of the keyboard and the B factor on the other and multiplying both by the appropriate total frequency. A similar "short cut" can be used for the pairs D-E and H-J.

4. Next, the quantities C and F are computed from the previously calculated quantities. The quantity C, for example, is

$$\frac{B-A-N}{2}.$$

Hence

$$C=\frac{17587-15771-50}{2}=883.$$

5. As a check on the work already performed, $H-J$ should equal 4 $(F-C)$.

6. The quantities K and L are then computed according to the formulas indicated. K, for example, is given as

$$\sqrt{NA-C^2}.$$

Hence

$$K=\sqrt{(50)(15771)-(883)^2}=\sqrt{8861}=94.13.$$

Note that the entire operation indicated under the radical can be performed directly on any calculator with a provision for negative multiplication. Square roots are conveniently taken from a table (e.g., Comrie, 1930).

7. The operations indicated by the expressions

$$C\cdot I_x/N$$

and

$$F\cdot I_y/N$$

are next performed. For the former,

$$(883)(1)/50=17.66.$$

8. The means of x and y are then found by the addition of the quantities in the two boxes immediately above the space for the means. Thus

$$\bar{x}=1.00+17.66=18.66.$$

9. The standard deviations of x and y are found as indicated by the expressions

$$K\cdot I_x/P$$

and

$$L\cdot I_y/P$$

where

$$P=\sqrt{N(N-1)}.$$

In this example,

$$P=\sqrt{(50)(49)}=49.497.$$

Appropriate values of P may of course be tabled for the range of sample sizes likely to be encountered, so that this work is done once for all. Thus

$$s_x=(94.13)(1)/49.497=1.902.$$

10. The correlation coefficient, r, is next computed according to the expression

$$r=\frac{(B+D-2F-J)\frac{N}{2}-C\cdot F}{K\cdot L}.$$

In this example,

$$r=\frac{(17587+10148-1412-1101)25-(883)(706)}{(94.13)(94.68)}=\frac{7152}{8912}=.802.$$

Note that the entire numerator may be computed in one set of operations on most automatic calculators.

11. Repeat steps 6 to 10 as a check.
12. Note the observed range of x values. In this example

$$OR_x=14\text{–}32.$$

In this way, the basic bivariate statistics (N, $\bar{x}$, $\bar{y}$, s_x, s_y, r, OR_x) are provided with a minimum of effort. A more elaborate characterization would include other statistics calculated as follows:

$$a=\frac{s_y}{s_x}=\frac{1.913}{1.902}=1.006^{1}$$

$$\sigma_a=a\sqrt{\frac{1-r^2}{N}}=(1.006)\sqrt{\frac{1-(0.802)^2}{50}}=(1.006)\sqrt{0.00714}=0.085$$

$$b=\bar{y}-\bar{x}a=15.12-(18.66)(1.006)=-3.65$$

$$D_{y\cdot x}=\frac{100s_{y\cdot x}}{\bar{y}}=\frac{100s_y\sqrt{1-r^2}}{\bar{y}}=\frac{(100)(1.913)(0.598)}{15.12}=\frac{114.4}{15.12}=7.6$$

$$\sigma_{D_{y\cdot x}}=\frac{D_{y\cdot x}}{\sqrt{2N}}=\frac{7.6}{10}=0.76$$

$$D_d=100\sqrt{\frac{2(s_x^2+s_y^2)(1-r)}{\bar{x}^2+\bar{y}^2}}=100\sqrt{\frac{2(7.277)(0.198)}{348.2+228.6}}=100\sqrt{\frac{2.882}{576.8}}=100\sqrt{0.00500}=7.1$$

Example 2. Samples of Various Subspecies of *Strophodonta extenuata* Shown to Differ Significantly in Growth Patterns Relating Length and Width

Data in table 6 provide a partial statistical characterization of four samples of different subspecies of the Devonian brachiopod *Stro-*

[1] Because a is the tangent of the angle of slope, the angle may be readily computed as 45°10′. Knowing one point on the line ($\bar{x}$, $\bar{y}$), one can then plot the growth line without further calculation.

TABLE 6

Bivariate Statistical Characterization of Four Subspecies of *Strophodonta extenuata*

(x=width; y=length; measurements in millimeters.)

Statistic	*S. e. bellensis*[a]	*S. e. extenuata*[b]	*S. e. ferronensis*[c]	*S. e. rockportensis*[d]
N	50	49	49	40
$\bar{x}$	20.50	18.15	25.17	21.05
$\bar{y}$	16.54	15.95	20.48	18.75
s_x	3.387	4.091	3.637	4.528
s_y	3.458	3.770	3.745	4.954
r	0.864	0.953	0.905	0.899
OR_x	6–14	6–15	8–18	7–20
a	1.021	0.922	1.030	1.094
σ_a	0.073	0.040	0.063	0.076
b	−4.39	−0.78	−5.45	−4.28

[a] Upper Bell shale, abandoned quarry at Rockport, Sec. 6, T. 32 N., R. 9 E., Alpena County, Michigan.
[b] Lower Ferron Point shale, abandoned quarry at Rockport, Sec. 6, T. 32 N., R. 9 E., Alpena County, Michigan.
[c] Upper Ferron Point shale, SE ¼, Sec. 18, T. 32 N., R. 9 E., Alpena County, Michigan.
[d] Rockport Quarry limestone, abandoned quarry at Rockport, Sec. 6, T. 32 N., R. 9 E., Alpena Co., Michigan.

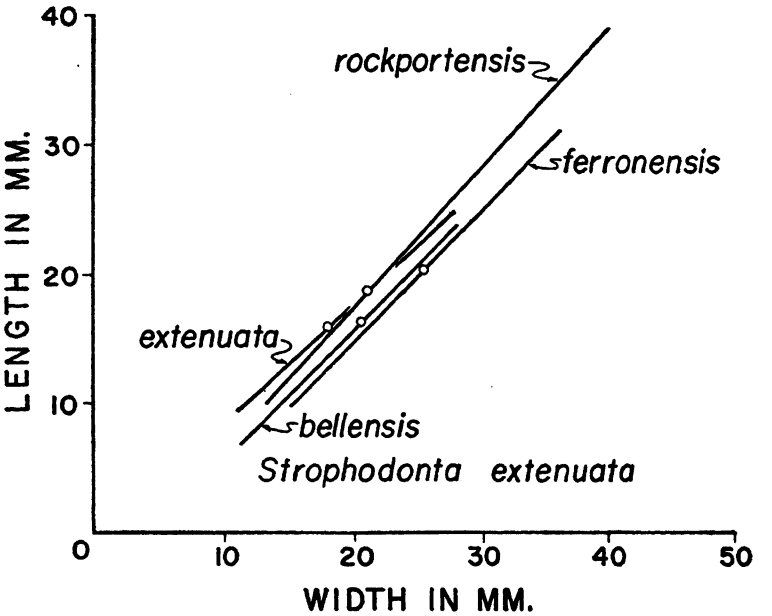

FIG. 8. Samples of four subspecies of *Strophodonta extenuata* characterized by reduced major axes relating length and width. Length of lines corresponds to observed range. Points represent joint means. Data given in table 6.

phodonta extenuata. Growth lines relating width and length constructed from these data are presented in figure 8. The use of these data in statistical discrimination is illustrated below.

EXAMPLE 2A. *Strophodonta extenuata rockportensis* AND *Strophodonta extenuata ferronensis:* DIFFERENCE SHOWN BY INSPECTION

From an inspection of growth lines representing samples of these two subspecies, and from a consideration of the sample size and dispersion, it is evident that the difference between the two growth patterns is too great to have arisen by chance. Formal statistical tests are therefore unnecessary.

EXAMPLE 2B. *Strophodonta extenuata extenuata* AND *Strophodonta extenuata rockportensis* SHOWN TO HAVE DIFFERENT SLOPES

From an inspection of figure 8 it is evident that the two samples in question differ slightly in the slope of the growth line relating length and width. It is not clear, however, that the observed difference in slope is significant. With the use of the test previously described,

$$z=\frac{a_1-a_2}{\sqrt{\sigma_{a_1}^2+\sigma_{a_2}^2}}=\frac{0.922-1.094}{\sqrt{(0.040)^2+(0.076)^2}}=\frac{-0.172}{\sqrt{0.00738}}=\frac{-0.172}{0.0859}=-2.00.$$

Because the absolute value of z is greater than 1.96, the observed difference is considered to be significant at the 5 per cent level.

EXAMPLE 2C. *Strophodonta extenuata bellensis* AND *Strophodonta extenuata ferronensis* SHOWN TO HAVE THE SAME SLOPE BUT DIFFERENT POSITION

Testing for slope difference, as above, one has

$$z=\frac{a_1-a_2}{\sqrt{\sigma_{a_1}^2+\sigma_{a_2}^2}}=\frac{1.021-1.030}{\sqrt{(0.073)^2+(0.063)^2}}=\frac{-0.009}{\sqrt{0.00930}}=\frac{-0.009}{0.0964}=-0.093.$$

TABLE 7

BIVARIATE STATISTICAL CHARACTERIZATION OF TWO SUBSPECIES OF *Pholidostrophia gracilis* (MEASUREMENTS IN MILLIMETERS)

	x=Width y=Length	x=Width y=Thickness	x=Log-Width y=Log-Thickness
P. gracilis nanus[a]			
N	50	50	50
$\bar{x}$	14.12	14.16	1.149
$\bar{y}$	11.12	2.22	0.331
s_x	1.409	1.405	0.0463
s_y	1.206	0.465	0.0933
r	0.793	0.383	0.609
OR_x	11–16	11–16	1.04–1.20
a	0.856	0.331	2.015
σ_a	0.073	0.043	0.226
b	−0.97	−2.47	−1.984
P. gracilis gracilis[b]			
N	50	50	50
$\bar{x}$	18.66	18.66	1.268
$\bar{y}$	15.12	3.96	0.579
s_x	1.902	1.902	0.0477
s_y	1.913	1.009	0.1010
r	0.802	0.546	0.580
OR_x	14–23	14–23	1.15–1.36
a	1.006	0.530	2.117
σ_a	0.085	0.063	0.243
b	−3.65	−5.93	−2.105

[a] Upper Bell shale, abandoned quarry at Rockport, Sec. 6, T. 32 N., R. 9 E., Alpena County, Michigan.
[b] Lower Ferron Point shale, same locality as above.

A statistically significant difference in slope is therefore not demonstrated. Inspection of figure 8 indicates that the significance of positional difference may be tested at the point $x_0=\bar{x}_1$ (where $\bar{x}_1$ indicates the mean of the sample of *S. e. bellensis*). Thus

$$z=\frac{\bar{x}_1(a_1-a_2)+(b_1-b_2)}{\sigma_{a_2}(\bar{x}_1-\bar{x}_2)}=\frac{(20.50)(-0.009)-4.39+5.45}{(0.063)(20.50-25.17)}=\frac{-0.18-4.39+5.45}{(0.063)(-4.67)}=\frac{0.88}{-0.294}=-2.99.$$

As $z>2.58$, the observed difference in position is taken to be statistically significant at the 1 per cent level.

EXAMPLE 3. SAMPLES OF SUBSPECIES OF *Pholidostrophia gracilis* SHOWN NOT TO DIFFER SIGNIFICANTLY IN LENGTH-WIDTH AND THICKNESS-WIDTH GROWTH PATTERNS

The data in table 7 provide a partial statistical characterization of two subspecies of the Devonian brachiopod *Pholidostrophia gracilis*. Growth lines constructed from these statistics are plotted on figures 9 and 10 together with the observed points.

EXAMPLE 3A. *Pholidostrophia gracilis gracilis* AND *Pholidostrophia gracilis nanus* SHOWN NOT TO DIFFER IN GROWTH PATTERNS RELATING LENGTH AND WIDTH

Let sample 1 be *P. gracilis nanus* and sample 2 be *P. gracilis gracilis*. In testing for slope difference, one has

$$z=\frac{a_1-a_2}{\sqrt{\sigma_{a_1}^2+\sigma_{a_2}^2}}=\frac{0.856-1.006}{\sqrt{(0.073)^2+(0.085)^2}}=\frac{-0.150}{0.112}=-1.34,$$

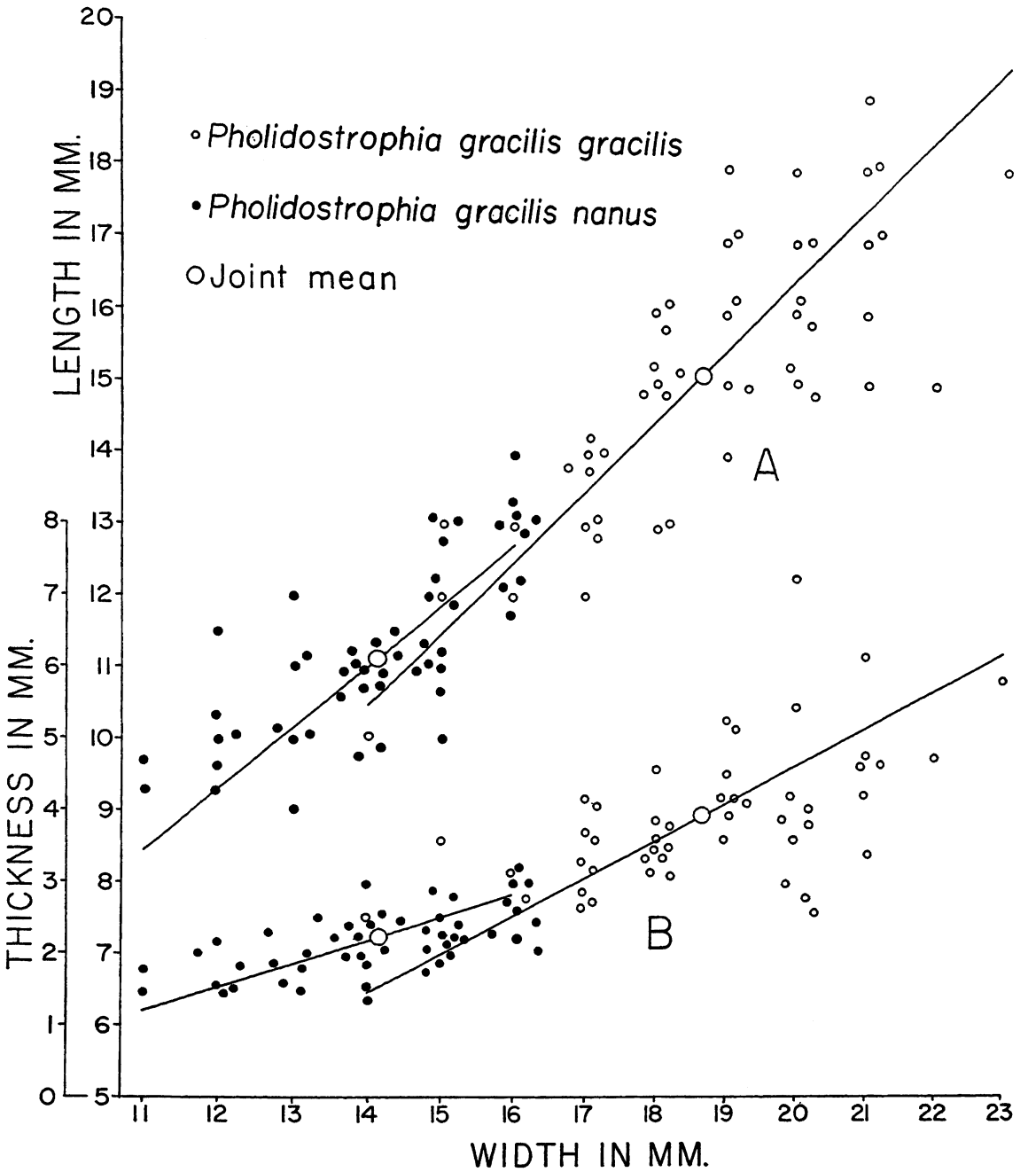

FIG. 9. Samples of two subspecies of *Pholidostrophia gracilis* characterized by reduced major axes calculated for linear measurements. *P. g. nanus* from the upper Bell shale. *P. g. gracilis* from the lower Ferron Point shale. Data given in table 7. A. Growth pattern of length and width. B. Growth pattern of thickness and width.

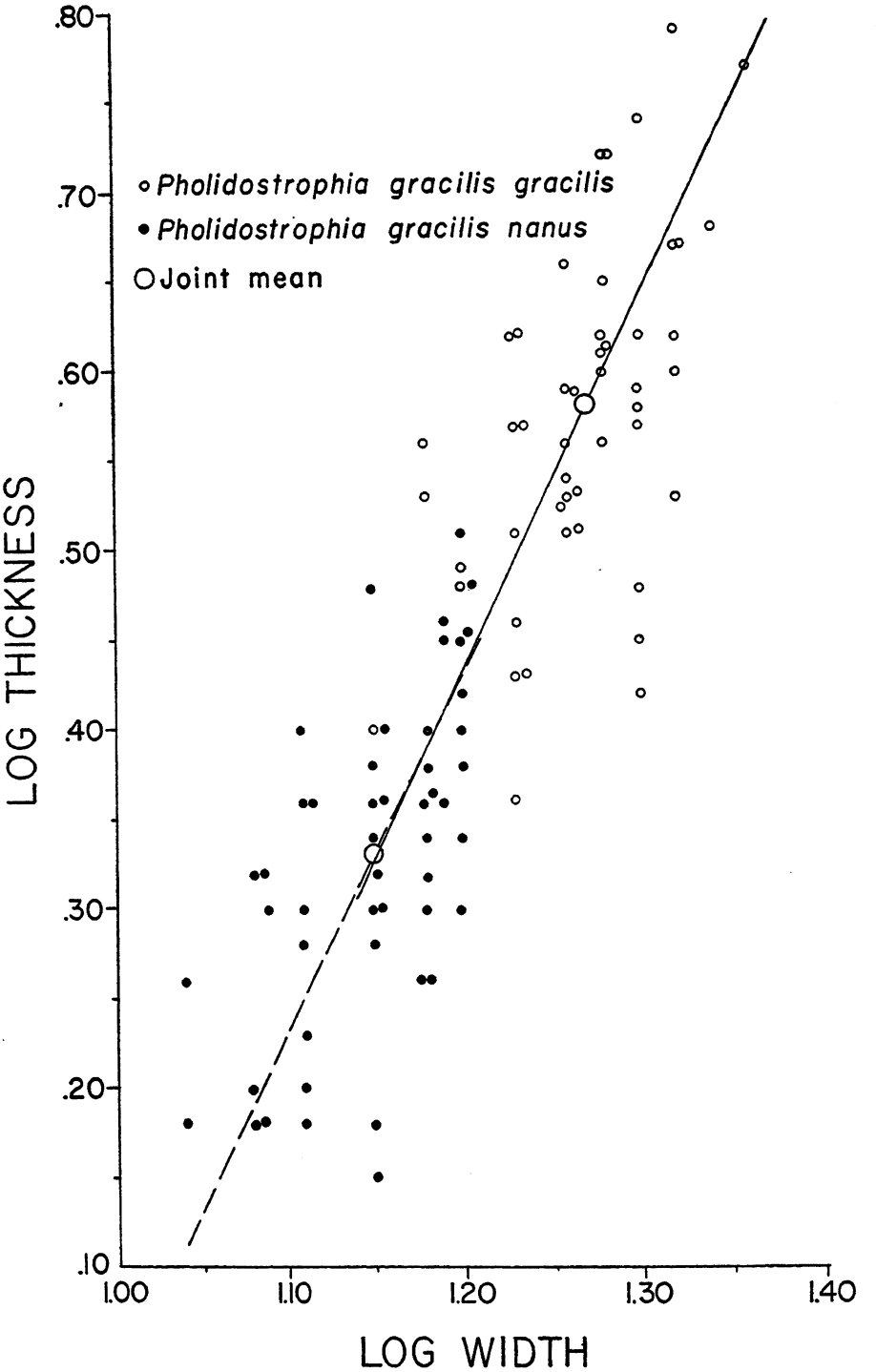

FIG. 10. Samples of two subspecies of *Pholidostrophia gracilis* characterized by reduced major axes calculated for log-thickness and log-width. Points correspond to figure 9B. Data given in table 7.

and the observed difference is not significant. An inspection of figure 9A indicates that it is reasonable to test for positional differences at $x_0 = \bar{x}_1$. Thus

$$z = \frac{\bar{x}_1(a_1 - a_2) + (b_1 - b_2)}{\sigma_{a_2}(\bar{x}_1 - \bar{x}_2)} = \frac{14.12(0.856 - 1.006) - 0.97 + 3.65}{(0.085)(14.12 - 18.66)} = \frac{0.56}{(0.085)(-4.54)} = \frac{0.56}{-0.386} = -1.45$$

and the difference in position is not shown to be significant at the 5 per cent level.

EXAMPLE 3B. *Pholidostrophia gracilis gracilis* AND *Pholidostrophia gracilis nanus* SHOWN NOT TO DIFFER SIGNIFICANTLY IN GROWTH PATTERNS RELATING LOG-THICKNESS AND LOG-WIDTH

An inspection of the scatter of points representing the original data on thickness and width (fig. 9B) indicates that there is a considerable difference in the linear trends of the two samples. This difference is statistically significant. From the over-all nature of the trend, however, it is apparent that growth is allometric in this pair of characters. Statistical analysis therefore requires that logarithms of the original data be employed. Data so transformed are plotted on figure 10 together with the calculated growth lines. The test for slope is then carried out as usual.

$$z = \frac{a_1 - a_2}{\sqrt{\sigma_{a_1}^2 + \sigma_{a_2}^2}} = \frac{2.117 - 2.015}{\sqrt{(0.243)^2 + (0.226)^2}} = \frac{0.102}{\sqrt{0.110}} = 0.31$$

The difference in slope is not significant. Testing for positional differences at $x_0 = \bar{x}_1$ (with *P. g. gracilis* taken as sample 1), one has

$$z = \frac{\bar{x}_1(a_1 - a_2) + (b_1 - b_2)}{\sigma_{a_2}(\bar{x}_1 - \bar{x}_2)} = \frac{(1.268)(2.117 - 2.015) - 2.105 + 1.984}{(0.226)(1.268 - 1.149)} = \frac{0.129 - 2.105 + 1.984}{(0.226)(0.119)} = \frac{0.008}{0.027} = 0.30,$$

and the difference in position is not significant.

Because the data at hand do not indicate a significant difference between *P. g. gracilis* and *P. g. nanus* in growth patterns relating length and width or thickness and width, the conclusion is justified that (for these characters) the only demonstrable differences are directly or indirectly related to differences in absolute size.

REFERENCES

BURMA, B. H.
1948. Studies in quantitative paleontology, I. Some aspects of the theory and practice of quantitative invertebrate paleontology. Jour. Paleont., vol. 22, pp. 725–761.
1949. Studies in quantitative paleontology. II. Multivariate analysis—a new analytical tool for paleontology and geology. *Ibid.*, vol. 23, pp. 95–103.

CAZIER, M. A., AND A. L. BACON
1949. Introduction to quantitative systematics. Bull. Amer. Mus. Nat. Hist., vol. 93, art. 5, pp. 347–388.

COMRIE, L. J.
1930. Barlow's tables of squares, cubes, square roots, cube roots, and reciprocals of all integer numbers up to 10,000. Third edition. London, 208 pp.

DELEERS, CHARLES, AND ANDRE PASTIELS
1952. Contribution à l'étude biométrique de *Lingula mytilloides* Sowerby du Westphalien de la Belgique. Assoc. Étude Paléont. Stratigraphie Houillères, Publ., no. 12, 67 pp.

DIXON, W. J., AND F. J. MASSEY, JR.
1951. Introduction to statistical analysis. New York, 370 pp.

HUBBS, C. L., AND CLARK HUBBS
1953. An improved graphical analysis and comparison of series of samples. Syst. Zool., vol. 2, pp. 49–56.

HUBBS, C. L., AND ALFRED PERLMUTTER
1942. Biometric comparison of several samples, with particular reference to racial investigations. Amer. Nat., vol. 76, pp. 582–592.

HUXLEY, JULIAN S.
1932. Problems of relative growth. London, 276 pp.

KERMACK, K. A.
1954. A biometrical study of *Micraster coranguinum* and *M.* (*Isomicraster*) *senonensis*. Phil. Trans. Roy. Soc. London, ser. B, vol. 237, pp. 375–428.

KERMACK, K. A., AND J. B. S. HALDANE
1950. Organic correlation and allometry. Biometrika, vol. 37, pp. 30–41.

KLAUBER, L. M.
1943. Tail-length differences in snakes with notes on sexual dimorphism and the coefficient of divergence. Bull. Zool. Soc. San Diego, no. 18, pp. 4–60.

KOTAKA, TAMIO
1953. Variation of Japanese *Anadara granosa*. Trans. Proc. Palaeont. Soc. Japan, new ser., no. 10, pp. 31–36.

KRUSKAL, W. H.
1953. On the uniqueness of the line of organic correlation. Biometrics, vol. 9, pp. 47–58.

MAYR, ERNST, E. G. LINSLEY, AND R. L. USINGER
1953. Methods and principles of systematic zoology. New York, 328 pp.

MILLER, R. L.
1949. An application of the analysis of variance to paleontology. Jour. Paleont., vol. 23, pp. 635–640.

OLSON, E. C., AND R. L. MILLER
1951. Relative growth in paleontological studies. Jour. Paleont., vol. 25, pp. 212–223.

PASTIELS, ANDRE
1953. Étude biométrique des Anthracosiidae du Westphalien A de la Belgique. Assoc. Étude Paléont. Stratigraphie Houillères, Publ., no. 16, 56 pp.

REEVE, E. C. R.
1940. Relative growth in the snout of anteaters. Proc. Roy. Zool. Soc. London, ser. A, vol. 110, pp. 47–80.

REEVE, E. C. R., AND J. S. HUXLEY
1945. Some problems in the study of allometric growth. *In* Clark, W. E. Le Gros, and P. B. Medawar, Essays on growth and form presented to D'Arcy Wentworth Thompson. Oxford, pp. 121–156.

SIMPSON, G. G.
1941. Range as a zoological character. Amer. Jour. Sci., vol. 239, pp. 785–804.
1943. Criteria for genera, species, and subspecies in zoology and paleozoology. Ann. New York Acad. Sci., vol. 44, pp. 145–178.
1945. The principles of classification and a classification of mammals. Bull. Amer. Mus. Nat. Hist., vol. 85, 350 pp.

SIMPSON, G. G., AND ANNE ROE
1939. Quantitative zoology. New York, 414 pp.

SNEDECOR, G. W.
1946. Statistical methods. Fourth edition. Ames, Iowa, 485 pp.

TEISSIER, GEORGES
1948. La relation d'allométrie: sa signification statistique et biologique. Biometrics, vol. 4, pp. 14–48.

WANG, Y.
1949. Maquoketa Brachiopoda of Iowa. Mem. Geol. Soc. Amer., no. 42, 55 pp.

WILKS, S. S.
1951. Elementary statistical analysis. Princeton, 284 pp.

ZUCKERMAN, SOLLY
1950. A discussion of the measurement of growth and form. Proc. Roy. Soc. London, ser. B, vol. 137, pp. 433–523.