STUDIES IN QUANTITATIVE PALEONTOLOGY: I. SOME ASPECTS OF THE THEORY AND PRACTICE OF QUANTITATIVE INVERTEBRATE PALEONTOLOGY

Benjamin H. Burma
University of Nebraska-Lincoln

Follow this and additional works at: http://digitalcommons.unl.edu/geosciencefacpub

Part of the Earth Sciences Commons


This Article is brought to you for free and open access by the Earth and Atmospheric Sciences, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in the Earth and Atmospheric Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
STUDIES IN QUANTITATIVE PALEONTOLOGY: I. SOME ASPECTS OF THE THEORY AND PRACTICE OF QUANTITATIVE INVERTEBRATE PALEONTOLOGY

BENJAMIN H. BURMA
University of Nebraska, Lincoln, Nebraska

ABSTRACT—For some time there has been a noticeable trend toward a more strictly quantitative outlook in paleontology. In view of many special circumstances encountered in dealing with fossil material, there is a rather extended consideration here given to the theoretical background of the quantitative method as applied to fossil invertebrates. First there is a consideration of the relations between sampling theory and paleontology. This is followed by a discussion of variability in populations and the problem of inference from sample to population. A consideration is next given to the species concept and a comparison of the units which can be used by the paleontologist to those used by the neozoologist. The problem of distinguishing between two species is then considered, this being done from several aspects. Synchronous species, a source of much difficulty, are discussed at some length and this leads into the problem of geographic gradients or clines. This in turn brings up the subject of infraspecific units and the possibility, and advisability, of recognizing them. There is also a short discussion of methods of presentation of data. The last part of the paper deals with the practical application of these various methods to the study of several representative kinds of invertebrate fossils. Of these, the blastoids are representatives of animals that build skeletons of plates, the pelecypods and brachiopods of those with two valves. There are two general types of spiral shells, that in which the initial portion is not enclosed within the later portion, and that in which it is. The gastropods serve to represent the first type, and the fusulinids the second.

It is concluded that quantitative methods are demanded by theoretical considerations and that they are thoroughly usable and practical.

THIS paper is the outgrowth of some ten years of work devoted to the problem of how to get the most, and the most accurate, information from paleontological materials. As a result of this study, it has been concluded that quantitative methods represent the best approach. However, a rather extended search of the literature has failed to reveal any statement of the theoretical and philosophical background for such studies which would also take into consideration the special problems encountered in dealing with invertebrate fossils. The growing trend toward the quantitative outlook in the field of invertebrate paleontology makes the need for such a statement imperative.

HISTORICAL REVIEW

It is beyond the scope of this paper to deal with the development of the quantitative attitude in science. In any branch, the qualitative outlook has been the mark of its youthful stages, and the assumption of a quantitative outlook has marked the beginning of maturity. Invertebrate paleontology to date has been almost entirely qualitative. A few students have made some use of quantitative methods but practically none has used them consistently or fully.

The British have probably used quantitative techniques most. Among the earlier workers was Carruthers (1910) who has been widely followed by other British coral students. His studies are, however, best termed semi-quantitative. A. E. Trueman (1922) used frequency graphs to show the evolutionary changes in Gryphaea. In 1924 (Trueman, 1924), he published a short paper on the species concept in which he concluded that the species name is best applied only to those specimens which agree with the characters of the holotype and that variations should be otherwise distinguished. This abiological concept of the species was used by Davies and Trueman in their 1924 publication on the Coal Measures lamellobranchs. This paper makes use of frequency
plots, scattergrams, and coefficients of correlation to show the variability of populations. In the late 1920's two papers appeared on the ontogenetic variation in gastropods (Stuart, 1927; Rowlands, 1928). These emphasized the importance of variation in ontogenetic development, but such a beginning does not seem to have been further developed. Waddington (1929) urged the use of more strictly quantitative methods in the study of ammonites but his recommendations likewise do not seem to have been much followed.

Willard (1930) published a short paper on the evolution of the Platystrophias. He used averages of characters but his data were not analyzed and variation was largely ignored. As a consequence his conclusions are open to question. Elias (1937) dealt with the late Paleozoic fenestrate bryozaons using some quantitative measures with, apparently, much success. Here again, however, the data were not analyzed and little attention was paid to variation. Leitch (1936, 1940) has written concerning the Coal Measures lamellibranchs. His 1940 paper, particularly, includes statistical analyses and is an important advance because it defines species in terms of variation.

Simpson has written a number of papers dealing with the quantitative aspects of paleontology, particularly vertebrate paleontology. His book on methods (Simpson and Roe, 1939) is invaluable to the interested paleontologist. Salmon (1942) has made some use of quantitative methods, but her treatment of the Mohawkian Rafinesquinas is essentially qualitative. Nicol (1944) has published a study of Elphidium made on a strictly quantitative basis. This paper, excellent in concept, is discussed at some length below. The year 1945 saw a number of such papers published. Bancroft's (1945) very important posthumous paper on brachiopods made full use of symbols in a quantitative manner, seemingly with great success. C. L. Cooper (1945) presented a quantitative approach to the study of moult-stages of ostracodes using ontogenetic formulae. Schenck (1945) attempted to study the shifting of populations during the late Tertiary on the Pacific coast with reference to their present distribution, but encountered difficulties due to lack of distribu-

tional and other data. Wood and Barnard (1946) have published a very exhaustive study of the variation of Ophthalmidium but their approach was essentially qualitative. Finally there is the recent paper of Jeffords (1947) on late Paleozoic lophophyllid corals which is tantalizing because of its suggestion of the possible applications of truly quantitative methods to the study of corals.

The above review makes no pretense to completeness but constitutes a fair sample of the papers and emphasizes the recency of such studies and the increasing amount of attention given them. Other papers are discussed later in this study.

OBJECTIVES

The present paper is an attempt to make a rather thorough inquiry into the theoretical and philosophical basis for quantitative work with fossil invertebrates, and into some typical applications of these methods. Portionately more space is devoted to the theoretical than to the purely practical aspects of the subject.

Only minimum attention is given to statistical methods in this paper. For persons without a mathematical background, Simpson and Roe's "Quantitative Zoology" is excellent as a reference. Those with a mathematical background will find Fisher's "Statistical Methods for Research Workers" useful. Methods discussed in these books are referred to throughout this paper. The theoretical discussions, however, are made as non-mathematical as is consistent with clarity and usefulness.

Great difficulty has been experienced in the organization of this paper. Most of the subjects discussed presuppose a knowledge of so many others that it has been impossible to avoid a great deal of cross-reference. Also the present study is concerned with species and infraspecific units only. Generic and higher units are more or less arbitrary groupings which are not particularly amenable to quantitative handling.

ACKNOWLEDGEMENTS

Financial assistance has been rendered this study by the Wisconsin Alumni Research Foundation of the University of Wisconsin and is gratefully acknowledged. I wish particularly to express my great debt
to Dr. Norman D. Newell who early encouraged me in these studies. Drs. Lewis M. Cline and M. L. Thompson have given me the benefit of stimulating criticism. Dr. K. J. Arnold has contributed much excellent advice concerning statistical methods but responsibility for their manner of use in this paper is entirely my own.

**Sampling Theory and Paleontology**

A fundamental factor in any consideration of the quantitative approach to paleontology is embodied in sampling theory. Before discussing some of the consequences of this theory, it may be well to review it briefly.

The primary interest in paleontology is, or should be, the study of fossil populations. However, it is quite impossible to study all the members of a population. Only part of the original population was preserved, of this some specimens have been destroyed and most of the remainder are unavailable because they are buried beyond reach. Also it is a sheer impossibility to carefully observe the enormous number of individuals that may be available. For these reasons, it is necessary to study a comparatively small number of specimens, a sample, selected from a population. From the study of this relatively small sample we seek to infer the properties of the population from which it was drawn, and further, to infer the probable character of other samples that may be drawn from this same population.

Obviously, the best sample would be one in which the entire range of variation present in the population is represented, and in the same proportions. If we were allowed a sufficiently large sample and could select the individuals in it, we might hope to come very close to this desideratum. In actual practice, neither of these two conditions can be met. Samples studied are, for one reason or another, mostly small, and clearly we cannot select specimens to show the nature of a population when it is that nature which we wish to determine. This means that samples, being unselected, may or may not correspond closely to the population. Under these circumstances, which are unavoidable and must be accepted, our inferences may be widely in error.

In practice, a sample is usually taken by going to an outcrop and picking up whatever specimens are available. Under the circumstances, this is the best method of sampling, for the sample will then be as nearly random as possible. (Actually, of course, the random-ness of the sample may be altered by conditions obtaining at the time of burial of the fossils. This is an inherent difficulty which cannot be overcome. Our object should be to avoid adding any additional selection.) The word "random" as used in this sense is rather difficult to define, and a rigorous definition will not be attempted here. Essentially, it means that, within available limits, no element of the population will be favored over another. For example, we will not study the larger members of a population and avoid the smaller ones. If we did, our sample would not be random, and our view of the population would be correspondingly distorted.

The fact that a given sample is random does not mean that any other sample drawn from the same population will have precisely the same characteristics. In fact, the opposite is true. If we have a population of 500,000 white and 500,000 black organisms (two independent characters) which are thoroughly mixed (a homogeneous population), a random sample of six would not invariably consist of three white and three black individuals. Actually, in the long run, a series of 64 samples of six specimens each would have this composition:

- 6 black .......................... 1 sample
- 5 black, 1 white ................ 6 samples
- 4 black, 2 white ............... 15 samples
- 3 black, 3 white ............ 20 samples
- 2 black, 4 white ............. 15 samples
- 1 black, 5 white .......... 6 samples
- 6 white ........................... 1 sample

Thus the proportions of two independent variables will reflect the composition of the entire population exactly in less than one-third of the samples consisting of six individuals. As the number of independent variables increases, this correspondence drops rapidly. If we take a sample of six from a homogeneous population of three independent variables, only about 1/11 of the samples will have the same compositions as the original population. The actual situation, however, is not quite so unsatisfactory because most characters of animals are not
independent but are correlated and they tend to vary together according to certain definite proportions. This means that an average sample will come closer to approximating the original population. The general conclusion, however, still holds true. Any two samples taken from a population cannot be expected to be exactly similar and probably they will not be.

To consider only one variable: A fundamental problem in paleontology is the comparison of two collections, or better, two samples. We wish to know, usually, if these two belong to the same species (might have been drawn from one population), or if they belong to different species (probably came from two separate populations). The fact that collections are samples of populations has, with few exceptions, been ignored and investigators have proceeded on the false and unwarranted assumption that collections are in fact perfect reflections of the populations they represent. We have just seen that the fact that two samples differ, even widely, does not necessarily mean that they came from different populations. On the contrary, we should expect that two samples from the same population will differ from each another. Many paleontologists, past and present, seem however to have proceeded on the assumption that the species is an entity which admits little or no variation. It is as though they subconsciously appealed to the old idea of “archetypes” and considered that there is some basic, metaphysical model to which the members of a species must conform. Thus, each time they obtained a group of specimens, or even a single individual which varied, often even slightly, from the holotype, they forthwith described another “species.” The analysis above, and that which follows, should indicate clearly that such procedure is, to put it most kindly, suspect. We can only conclude that all “species” created under the aegis of qualitative paleontology are to be viewed with great suspicion. Many are undoubtedly valid groups; many more were as certainly created on inadequate grounds. The mere fact that two collections, or two parts of a single collection differ in their composition is not an a priori reason for concluding that they were drawn from different populations.

Variability

Variability may be somewhat loosely defined as the tendency of one individual to differ from others in the same population. Its relative and actual amount will itself vary from character to character within a single species and from one species to another. Although the limits of variation are wide, we may expect that, in general, the maximum value of any character will be about 50 per cent greater than the minimum value for any particular growth stage. As an example, if the minimum width of a brachiopod at a certain growth stage is six mm., the maximum width may be expected to be about nine mm. It must be emphasized that this amount of variation is not unusual, but entirely normal. In fact, if the variability of a good-sized sample is too low, it is generally an indication that sampling has been inadequate or selective. In spite of these facts, it is commonplace for individuals at one end of a variable series to be separated from individuals at the other end, and perhaps both from the more abundant middle members of the sample, and considered to represent two, three or more “species” instead of one. There is no simple method for dealing with this problem. If the sample is sufficiently large, the intergradation of forms will usually be obvious, and generally there will be few representatives of either extreme with more and more intermediate specimens as the middle part of the range is approached (a normal distribution). If two well-marked frequency maxima are found, it may be an indication that two populations are represented in the sample. Only rarely, however, will paleontological samples be large enough to show this clearly. Usually sampling errors will confuse the situation so that no such simple separation is possible. An actual case is described below in the section on the Blastidea.

Usually it is well to make the preliminary assumption that only one species of any genus is present in a sample. In modern marine faunas two very similar species do not commonly exist side by side, although such associations are certainly known. This problem is considered further below.

With these factors in mind, let us see what kind of units we can recognize, regardless of the label we place on them, as opposed to
what kind of units we might like to recognize. Assume that we have collected a suite of some particular kind of fossil from a homogeneous population. This sample will have certain individual characteristics. It will consist of a certain number of individuals \( N \). It will have a certain arithmetic mean, or, simply, a certain mean
\[
\frac{\Sigma X}{N}
\]

It will show a certain variability, which almost certainly will not exactly duplicate variability of the population from which the sample was drawn. It will have a certain standard deviation
\[
\left( \frac{\Sigma (d^2)}{N} \right)^{1/2}
\text{ or }
\frac{\sqrt{\Sigma (X^2)) - (\Sigma X)^2}}{N}
\]
in which \( d \) is the difference between the measured value of a character in a single individual and the mean of the sample, or, in statistical language, the deviation from the mean.

So far we have considered only the sample. Our real interest, however, is, or should be, the population of which the sample is a part. In qualitative paleontology, inferences from sample to population are made on the basis of opinion. Unfortunately, opinions are widely variable, and still more importantly, one person has no way of checking another's opinion because it is wholly subjective.

There can be no effective argument against the injection of objective, quantitative methods into such a haphazard state of affairs. Fortunately, relatively simple methods exist for attacking this problem. The standard deviation \( \sigma \), just referred to, is a measure of the central tendency of a sample, i.e., the tendency for its members to group themselves about a mean. The smaller the value of \( \sigma \), the less variable the sample is and the larger its value, the greater the variability. Further, if the sample (and population) approaches a normal distribution, as most do, we may calculate the expected limits of variability to any desired degree of accuracy. For example, for a sample with mean, \( M \):

\[
M \pm 1\sigma \text{ includes } 68\% \text{ of the population}
\]

\[
M \pm 2\sigma \text{ includes } 95.5\% \text{ of the population}
\]

\[
M \pm 3\sigma \text{ includes } 99.7\% \text{ of the population}
\]

\[
M \pm 4\sigma \text{ includes } 99.994\% \text{ of the population}
\]

Thus, if the mean of a sample is 30 mm. and its standard deviation is 4 mm., only three individuals in a thousand of the original population should fall outside the range of 18 mm. to 42 mm. \((M \pm 3\sigma)\). Thus, with the usual margin for error that we must always leave when dealing in probabilities, we may say that our population varies in size from 18 mm. to 42 mm. for one particular growth stage. The value \( M \pm 3\sigma \) is most commonly used for this purpose and is generally adequate. Simpson (1941) offered some cogent criticisms of this method which are dealt with in the next section.

After inferring the limits of a population, we are next concerned with its mean. Although the mean of the sample, or the "typical specimens," is practically a fetish among many paleontologists, it is not as useful or characteristic a feature of the population as its variation. However, because of its wide usage and because it is more easily and widely understood than measures of variability, it is considered throughout this paper. Only by pure coincidence will the mean of a sample correspond exactly with the mean of a population. If, however, the standard deviation be divided by the square root of the number of individuals in the sample \( \sigma / \sqrt{N} \), a measure known as the standard error of the mean \( (\sigma_m) \) is obtained. If this measure be added to and subtracted from the mean in the same fashion as the standard deviation, we obtain the range within which the mean of the population may be expected to fall, to any desired accuracy. Thus, if \( M = 50 \) mm. and \( \sigma_M = 0.5 \) mm. \((3\sigma_m = 1.5 \text{ mm.})\), the chances are 997 in 1000 that the mean of the population will fall in the range from 48.5 mm. to 51.5 mm. The converse is also true: With the same sample, the mean of the population cannot be determined more closely than to say that it probably lies between 48.5 mm. and 51.5 mm. This is extremely important.
THE SPECIES CONCEPT

The definition of the word "species" has been the subject of a great deal of debate, much of it, unfortunately, rather fruitless. Perhaps the best recent discussion of the word and concept is by Mayr (1943, p. 102 et seq.) He concludes by defining the species as: "a group of populations which replace each other geographically or ecologically and of which the neighboring ones intergrade and interbreed wherever they are in contact or which are potentially capable of doing so (with one or more of the populations) in those cases where contact is prevented by geographical or ecological barriers. Or shorter, species are groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups." This definition seems to be an excellent one for the neozoologists. For the paleontologists, it is inadequate. Obviously we cannot view two similar specimens dead for millions of years and determine whether they belonged to interbreeding populations, potentially or otherwise. If those parts dealing with interbreeding are eliminated from Mayr's short definition, we are left with "species are groups of ... natural populations ...," which is hardly a helpful statement. Not being able to use a strictly biological definition of the species, we are forced, therefore to turn to a morphological definition with its attendant difficulties which Mayr has pointed out.

Most geneticists, zoologists, and paleontologists now agree that animal evolution has been an essentially continuous process unmarked by major saltations (Dobzhansky, 1941; Mayr, 1942; Simpson, 1944). As such, its course is better represented by a continuous line than by a series of en echelon line segments. This conclusion cannot be reviewed here but is accepted as the basis for the continuation of the present discussion.

Continuous evolution, however, introduces a difficulty when we come to define species. How can we distinguish stages in what is a continuous process? To set up a division at any point must be purely arbitrary. But species, as they have been distinguished, are set up as stages in this continuous evolutionary line. We are thus forced to the conclusion that "species" is a concept without objective existence when populations are viewed in a four-dimensional continuum.

On the other hand, we must also admit that the human mind is uncomfortable in dealing with continuities and is most at home with easily handled, discrete units. Our first effort should, therefore, be to decide how best to distinguish such units—which we will hereafter term "species." We have seen that we can infer the total range of the variation and that of the mean of a given population. Either of these ranges could be used for our present purpose but for reasons mentioned above, the range of the mean seems most suitable.

Let us now consider a phyletic line illustrated by the rising line in figure 1. In beds of age C we find fossils belonging to this evolutionary line. The mean of the sample falls, let us say, on the line at size 3.0 mm., and we find the standard error of the mean (σM) to be 0.50 mm. As we have already seen, we might expect the mean of the population to fall between 1.5 mm. and 4.5 mm. Now suppose we find fossils of the same evolutionary line in rocks of age D, the mean of its sample being 3.8 mm. and σM also 0.50 mm. According to our analysis of population C, its mean could be anything between 1.5 mm. and 4.5 mm. and the sample from D has its mean within this...
range. This indicates that populations C and D cannot be distinguished. In fact only in rocks older than age A and younger than age E would we expect to find samples which we can justify separating from population C. Such a unit may be designated a species insofar as paleontology is concerned. Thus, we might recognize species 1 with time range A to E and character range 1.5 mm. to 4.5 mm. and species 2 of time range E, to I and character range 4.5 mm. to 7.5 mm. Actually, as already intimated, the method above needs modification to be more accurate. First, since we are making a prediction and are working with only one sample, we must assume that any other sample we will compare to ours will be of the same size and have the same standard error of the mean. Since the two will belong to the same phyletic line, this last is not unreasonable. Under these circumstances, the expression

$$\sigma_d = \sqrt{\frac{N_1}{N_2} \sigma_{M_1}^2 + \frac{N_2}{N_1} \sigma_{M_2}^2}$$

reduces to

$$\sigma_d = \sqrt{2} \sigma_M.$$  

This last expression is then used instead of $\sigma_M$. Our expression for the range of the species then becomes,

$$\text{range} = M \pm 3\sqrt{2} \sigma_M^2.$$  

On this basis, we can objectively set up the limits within which we will refer several populations to the same species.

**ONE OR TWO SPECIES?**

A problem constantly confronting the paleontologist is that of deciding whether the collection (sample) he is working with belongs to some already described species or whether it is different. In the past, this question has been answered on a variety of bases, none of them particularly satisfactory. Commonly the decision was simply an opinion but such a haphazard method is to be defended only on the shaky grounds of expediency. Essentially our problem involves the comparison of two samples; one, our particular sample, and the other, the species with which we wish to compare it. Three general means of attack are open to us; a comparison of one character of each at a time; a comparison of two characters at a time; or a comparison of all relevant characters at once. These are discussed in order.

The few quantitative comparisons of invertebrate fossils have been made almost solely on the basis of single characters. There are several ways in which this can be done. An excellent one is that of the comparison of means by the following method:

$$\sigma_d = \sqrt{\frac{N_1}{N_2} \sigma_{M_1}^2 + \frac{N_2}{N_1} \sigma_{M_2}^2}$$

or

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{N_2} + \frac{\sigma_2^2}{N_1}}$$

whichever is most convenient. Next $d$ is obtained by subtracting the smaller arithmetic mean from the larger. Then $d$ is divided by $\sigma_d \cdot (d/\sigma_d)$. If the resultant is less than 2.0 ($P = 0.956$), the difference between the means is probably not significant, this probability increasing rapidly as the number decreases. If the answer is between 2.0 and 3.0 ($P = 0.955$ to $P = 0.997$), the difference is possibly significant. If the result is 3.0 or more ($P = 0.997$), the difference is almost certainly significant, the degree of probability increasing rapidly with the increase in the number. If the samples compared are small (say 20 or less), the quantity $t$ may be calculated according to

$$t = \frac{(M_1 - M_2) \sqrt{N_1 N_2}}{N_1 + N_2} \sqrt{\frac{M_1 \sigma_1^2 + N_2 \sigma_2^2}{N_1 + N_2 - 2}}.$$  

This quantity is then introduced into an appropriate table to evaluate its significance (Simpson and Roe 1939, p. 206). This last method is somewhat more lengthy but should be used if samples are very small or if they are only moderately large and significance is doubtful. In general, $d/\sigma_d = 3.0$ should be taken as the minimum for a significant difference.

The methods just described are perfectly valid but they ignore the fact that animals are differentiated from one another not by this or that character, but by the sum of many characters. Further, it takes no account of the ontogeny, whose importance is stressed in a section below. Quite apart from this, the method can be a source of
error among those who use statistical methods without understanding them. These persons might take two samples, each consisting of a growth series, and then compare the samples in toto by the method above. The mean in this case is the mean size of half grown specimens, the minimum size is that of the smallest of the youngest specimens, the maximum size is that of the largest of the oldest individuals. Such a procedure is meaningless and lacking in biological significance. If there is one dictum which should be established in quantitative paleontology, it is this: Comparisons of one character to be valid must be made at comparable growth stages and at comparable growth stages only. If the above method is used with circumspection and with an acute consciousness of its shortcomings, it will often be useful. It has the advantage of involving only short, simple calculations which, with a calculating machine will consume but little time. In general, however, it is preferable to use one of the other methods outlined below.

It is a decided improvement to compare two characters simultaneously, particularly if one is dealing with growth series. The computations involved are more complex and longer but by no means formidable. Let the two characters in question be $X$ and $Y$ respectively and the two samples 1 and 2. The coefficient of correlation ($r$) of $X$ and $Y$ is first calculated,

$$ r = \frac{\sum (d_y d_y)}{N \sigma_X \sigma_Y} $$

for each sample. Next, $\sigma_{d_y}$ is calculated from one of the following equations:

$$ \sigma_{d_y} = \sqrt{\frac{\sum (d_y d_y)(1-r^2) + N \sigma_X \sigma_Y (1-r^2)}{N_1+N_2+4}} \left( \frac{1}{N_1 \sigma_X^2} + \frac{1}{N_2 \sigma_Y^2} \right) $$

or

$$ \sigma_{d_y} = \sqrt{\frac{\sum (d_y d_y)(1-r^2) + \sum (d_y d_y)(1-r^2)}{N_1+N_2+4}} \left( \frac{1}{\sum (d_y^2)} + \frac{1}{\sum (d_y^2)} \right) $$

The second equation is somewhat preferable in dealing with raw data. Next, for each sample, there is calculated the value of $b_{yx}$ the regression coefficient, using either

$$ b_{yx} = \frac{\sum (d_y d_y)}{\Sigma (d_y^2)} \quad \text{or} \quad b_{yx} = r \frac{\sigma_y}{\sigma_x} $$

for the calculation, only simple arithmetic. The calculation is, however, lengthy and an electrically driven calculating machine is a necessity. Slide rules cannot be used except possibly as a check and longhand calculations would be prohibitively time-consuming. In spite of these limitations, the fact
that this method really compares specimens rather than characters makes it the method of choice in all doubtful cases.

At this point, a few parenthetical remarks concerning the use, and misuse, of statistical methods seem advisable. Statistical analysis is admittedly in bad odor with a great many paleontologists, and for a variety of reasons. One objection, which is constantly made, is that statistical methods may be a good thing for a number of specimens, but not with only a few. There are at least two fallacies evident here. First, there are statistical methods suitable for dealing with anything from 1 to 1,000 or more specimens. Secondly, anything which a person attempts to do with a small sample, which he could not do by statistical analysis (such as unduly exact discrimination of species), will probably be founded on error. Especially, with large samples, a person not using a quantitative technique may look for and find differences between the extremes of variation of two samples and then conclude that he has been able to distinguish between samples which statistical analysis would show to be the same. Such a conclusion is entirely unjustified. One of the chief services of statistical methods in paleontology is the erection of signposts saying "Danger! Your data are no good beyond this point!" Such a guide is extremely helpful in guarding against exaggerated notions as to how much can be done with a few specimens.

Still another consideration concerns the fact that statistics have so often been misused by those who have tried to apply these methods to paleontology. This is so either because of a lack of knowledge of the purpose and limitations of the methods used or because of a lack of appreciation of the philosophy behind them. Little can be gained by citing those who have used the right method in the wrong place, but they are all too numerous in geology as a whole and not only in paleontology. Perhaps even more serious has been the use of statistical methods by those without a grasp of its philosophy. This results in the use of statistical analyses as ends in themselves, or their use from some vague realization that such analyses should be made but with only a feeble attempt to relate them to the problem at hand.

Statistical methods are, first and foremost, tools and should always be used as such. They are also sharp tools, and if misused, they will do more damage than good. Used correctly, they are a powerful aid in attacking problems not otherwise soluble. No one, however, should undertake to use statistical methods unless he has a firm grasp of their uses and applications and a thorough understanding of the philosophy involved.

**HOLOTYPES AND HYPODIGMS**

On the basis of what has gone before, it should be clear that the larger the two samples are, the more precise their comparison can be, and, other things being equal, the closer their means can be together and still be significantly different. Quite apart from the mathematics involved, it must be clear that the larger the sample, the more nearly we can approximate the population, and the closer we can approximate the population, the closer it can be to another equally well known group and still be distinguished from it as a separate population.

Let us suppose that we have an unknown sample which we wish to compare to a previously known species. We will then want to compare our unknown to the largest possible sample of the described species in order that our comparison may be as accurate as possible. In all too many cases, the only data available are that for the holotype. There seems to be a widely current impression that the holotype of a species is the standard of comparison for a species. It is obvious where such a concept leads. Instead of an adequate sample of the known species, such as may have been originally available, we are asked to compare our unknown to a single specimen, the holotype. We are asked to make our comparison under the worst possible conditions. In fact, the only condition which could be worse would be to have no specimen at all.

The idea of the holotype as used in this sense, plainly shows the influence of the concept of archetypes already alluded to above. As a matter of fact, the holotype is a standard for nothing except itself. "The type is typical of nothing. It is only an indication of which groups of individuals must be associated with a particular specific name. It is the final court of appeal for purposes of
nomenclature only" (Williams 1940). The holotype is merely a specimen to which the name of the species must be attached no matter what the limits of the species are. In itself it tells us neither more nor less than any other specimen. G. G. Simpson has made a very pertinent suggestion in this matter. It is that the entire collection used in the description of a species be treated rather as collective types, a hypodigm. This collection would allow the comparison of known and unknown to take place under much better conditions.

RECOGNITION OF SYNCHRONOUS SPECIES

One of the most inherently difficult problems to which a paleontologist can apply himself is involved in the recognition of synchronous species, that is, species which belong to the same genus and which were in existence at the same time. It is also a problem whose implications have been blithely ignored by a majority of paleontologists.

The simplest aspect of this problem entails the recognition of young and mature individuals of the same species. Only too often these have been separated and made into two or more species. There is no "Open Sesame" for handling this problem. Fortunately, the form of the young animal is often preserved in the adult, either in the early stages of a spiral shell or in growth lines on the early portion of a shell or plate. If, as is often the case, the size of fossils belonging to one genus in a collection ranges from small to large, they should be tested to see if they form a growth series as discussed below under Ontogenies. The application of common sense to the problem should result in a ready solution. It should be remembered that the proportions of a young individual may be very different from those of the mature animal. For example, the ratio of the size of the head to that of the body is much greater in a human infant than in an adult. The chief thing to remember, however, is that big animals were little when they were young and it must be expected that such young individuals will be found.

A more difficult aspect of the problem is encountered in the recognition of true species in a collection from one restricted zone and locality. We are here confronted with the phenomenon of correlation. It does not seem to have been generally appreciated that many characters of animals are correlated with one another so that when one varies, others vary also, in accordance with certain definite ratios. For example, tall men have long arms and short men have shorter arms. If the ratio of height to arm length be calculated in each case, it will be found that the two ratios are about equal, that height and arm length are connected by a rather constant ratio. This is a matter of great significance in the recognition of synchronous species.

If we take a group of mature individuals from one locality and horizon and separate them into two groups on the basis of the size of one character, we will find that we have separated them on the basis of most other characters as well. Then if we are not familiar with correlation, we may conclude that we have two species, each with a definite set of characteristics.

To consider an example, Figure 2 shows the distribution of a growth series of an undescribed fusulinid from the Plattsmouth limestone. The radius vector is plotted against the volution number. Since this is a

![Graph of radius vector plotted against volution number of a Plattsmouth limestone Triticles.](Image)
spiral shell, a growth series is obtained and the measurements shown for the third volu-
tion, for instance, are made on the same individuals as those for the fourth volu-
tion, and so on. The diagonal center line connects the mean radius vectors of all volutions. We now note which individuals have a radius vector greater than the mean for the fifth volution and which less. We separate these two groups to form two samples and then recalculate the means of the radius vector for each volution. These new means are then plotted and connected by a line (fig. 3). The means of the radius vectors of each sample are now well separated over the whole area of the graph. If we now calculate the means of the height of volution and septal count for the same two samples and plot them (fig. 3 also), we find that these means are also separated by the same process, and in the same direction. The differences between these means are as great as are com-
monly used to distinguish between fusulinid species. We have, then, created two "species" where one grew before, and neither has any validity or reason for existence.

Here again there is no clear-cut method to avoid error. As was pointed out above,
ratio of character A to B of, say, 1:2 and the other part having the same ratio as 1:4, then we are on firmer ground in recognizing two species. In this situation, we are, in

and form characters depending on a combination of simple characters are more successful but still cannot be used blindly since a form character which seems to depend on

effect, working against correlation. This procedure is by no means free from criticism and each case must be examined solely on its own merits. In general, single characters of size, counts, and simple angular relations are the poorest, and indeed, for the most part unusable in splitting a sample. Shape

the interaction of several simple characters, may actually be found to be chiefly the result of a single size character. From all this it should be clear that the recognition of synchronous species is among the most difficult and uncertain tasks to which the paleontologist can apply himself.
CLINES

In the preceding section we considered the problem of recognizing species within a population existing at one place and time. Numerous other factors enter in if the population, still of one time, is spread over an area of notable extent. In addition to the question of "Are the differences statistically significant?", we must face the question of "Are the differences genetically significant?". To illustrate some of the difficulties involved we may use the data presented by David Nicol in his recent paper on West American species of the genus Elphidium (Nicol, 1944).

In this paper Nicol presented a study of what has been called Elphidium crispum on the Pacific coast of North America, and recognized in its stead E. fax fax and E. fax pingue from the Recent, E. fax barbarense from the Pleistocene, and E. excubitor and E. concinnnum from the Recent. We shall not concern ourselves with the Pleistocene form here. Of the rest, E. fax fax comes from the Straits of Juan de Fuca, Washington, Lat. 48° N; E. fax pingue from Monterey Bay, California, Lat. 36° 37' N; E. excubitor from Punta Penasco, Mexico on the Gulf of Lower California, Lat. 31° 21' N. and E. concinnnum from San Quentin on the Pacific side of Lower California, Lat. 30° 25' N. Nicol measured 200 specimens each of the subspecies of E. fax and a 200-specimen sample of E. excubitor and E. concinnnum combined. These data were analyzed and, from the results, he concluded that these are valid species and subspecies which are characterized by statistically significant differences.

If we examine these data, we find that, from north to south (fig. 4), the maximum diameter of the shells decreases with notable regularity. Similarly, but less regularly, the thickness, number of chambers in the last whorl, and the number of retral processes on the last chamber decrease with decreasing latitude (and, in general, increasing water temperature.)

This notable regularity of decrease of size with increasing temperature may lead us to wonder if we are here faced with a genetic difference, or one in which the primary control is environmental. To illustrate the reasonableness of this last supposition, let us turn aside for a moment.

The brine shrimps are small crustaceans, world-wide in distribution. They are found in saline waters ranging from almost fresh to those so concentrated that salts are precipitating. Early students noted a great

![Brine-shrimp figures](image-url)
variation in the form of these animals from place to place but a relatively constant form in any particular occurrence. As shown in fig. 5, the character of the distal end of the abdomen and the ratio of length of cephalothorax to abdomen are extremely variable, and almost every other external character is equally so. On the basis of these combinations of variations, the brine shrimps illustrated were placed in four species of Artemia in suitable environments, developed into over thirty "species" of the genus (Scott, 1924).

If this be true, and to such an extent, in animals of as high a degree of organization as arthropods and molluscs, it should not be surprising to find similar effects in protozoans.

In returning to Nicol's data, it is necessary first to examine his statistical conclusions.

![Graph](image-url)

**Fig. 6**—Abdomen to cephalothorax ratio of brine shrimps plotted against the salinity of the water in which they are found. Note how closely the points come to falling on a smooth curve. The forms with the largest and most hirsute tail-fans occur in the freshest water where buoyancy is most needed.

and two of Callaonella as shown. Unfortunately, it was then found that if the eggs of the brackish water *C. dybowskii* were reared in extremely saline water, *A. koppeniani* developed from them. It soon became clear that the six forms were one genetic group (species) of the same genetic constitution (genotype) whose external form (phenotype) depended on the environment in which it found itself (fig. 6). Similarly, eggs from a single individual of the pelecypod genus *Anomia*, have, on being hatched and reared

Though giving all due praise to Nicol for placing his study on a quantitative basis, we must conclude that the methods he chose to use in the analysis of his data were most unfortunate. Since the reasoning involved in these choices has been followed by others, it seems worth while to examine it further.

In his study Nicol used regression diagrams and by plotting one character against another derived "ontogenies." However, instead of using these growth curves for
comparisons, he used maximum diameters. "Maximum size was used rather than mean size because the smallest specimens were not available. Small shells of young individuals are more fragile and easily destroyed by waves and predatory animals. Without the smaller sizes, the location of the mean would naturally be affected. The maximum size, therefore, was the only stable measure that could be used in comparing species or communities within a species" (Nicol, 1944: p. 179). There are several serious misconceptions in these sentences. The central one concerns the idea of the mean size of a species. As Nicol uses the word "mean," it becomes the average size between that of the oldest and youngest specimens (see above, One or Two Species?). If, using this idea, we measure a random sample of the human male population, we would find that height varies from about 18 inches for infants to about 78 inches for large adults, and that the mean height of human males is about 48 inches, an obvious absurdity. The mean size of a growth series is not a point, but a line. There is no single mean size for an animal; there is a mean size for each growth stage. This is a point which Nicol, and certain others, have missed entirely and it brings down their whole analysis. The fact that the youngest members of a population are not available is certainly no reason for turning one's back on all but the largest. Furthermore, as has already been pointed out it is much more preferable and accurate to compare regression lines for two characters than it is to compare single characters. The maximum size, subject as it is to variations due to sampling (all extremes being more uncommon) is certainly not "the only stable character that can be used for comparing species." Rather, it comes nearer to being the most unstable. Nicol's estimates of significance, given on his pages 180 and 181, must therefore be rejected as being based on an erroneous supposition. These comments also apply with lessened force to his estimate of significance of ratios, though I agree with his conclusion that "Most of the ratios are not significantly different."

If, instead of using these unsuitable methods, we test for the significance of the differences of the regression lines, we come to quite different conclusions. In the case of *E. fax fax* and *E. fax pingue*: comparing the regression lines of the greater on the lesser diameters for both subspecies, $P = 0.70$ (Nicol's value, $P = 0.0001$ for both subspecies and both characters). Other characters give similar results. In other words, based on these two characters, the probability is 70 in 100 that the two samples do not differ significantly, that the differences observed could be due to sampling errors. On such a basis, there seems to be little reason for recognizing the two subspecies and we are left with simply *E. fax*. On the other hand, if we compare *E. excubitor* and *E. concinnum* taken together (data for the two are not separated in Nicol's paper) with *E. fax* for the significance of the difference of the same two regression lines, we find that $P = 0.0008$ (Nicol's values, $P = 0.0001$). This figure is decidedly significant, though rather less so than Nicol's calculations would indicate. Data are not presented which will allow us to decide whether *E. concinnum* differs significantly from *E. excubitor*, but on the basis of what is given, it would not be surprising if it did.

The question remains as to the meaning to be ascribed to the fact that *E. fax* and *E. concinnum* are statistically different. Nicol and others seem to proceed on the *a priori* assumption that a statistical difference necessarily means a real genetic difference. In the brine shrimps and in *Anomia*, however, greater differences are without genetic significance. If we examine the geographic distribution of these forms of *Elphidium*, we get a suggestion that the situation is similar to that of the brine shrimps and *Anomia*. As Nicol pointed out, there is a "break in the distribution of *E. crissum* just north and south of Point Conception." If we examine the water temperature of the area, shown on the right side of fig. 4, we see that this break corresponds almost exactly to a sharp drop in the temperature of the coastal waters of the same area. In view of the general correlation between size and other characters, and water temperature, as already noted, it does not seem likely that this is mere coincidence.

3 Both the above and what follows have been written solely as an illustration of principles. Under no circumstances is it to be viewed as a revision of the nomenclature of *Elphidium*. 
On the basis of the above factors, it would seem reasonable to conclude that the west coast Elphidiums of the *crispum* group are a genetic unit, and that their division into species is not warranted; that the variation found is a result of differences in water temperature which controls the form of the phenotype of a single unit. Such a unit fits well with the definition of a cline (Huxley, 1939), a population which varies progressively from one end of its range to the other. In this particular case, it seems likely that the variational control is primarily environmental rather than genetic. Strictly speaking, there seems little reason for attaching names such as *fux* or *concinnum* to non-genetic variants, whether such variants be considered as species, subspecies or varieties, and one name should be applied to the whole unit. Other, practical matters, however, enter into this problem which demand a consideration of infraspecific units before a decision can be reached in this case.

INFRASPECIFIC UNITS

If it is difficult for paleontologists to agree as to what a species is, it is even more difficult to get any agreement on the vexed problem of infraspecific units. There are three kinds in more or less regular use in paleontologic literature. One is based upon geographical variants of a species, another upon time variants and the third upon morphological variants at one locality. In addition, each of these may be subdivided into those in which genetic control is dominant and those in which environmental influences are more important. This makes a total of six infraspecific categories which might receive names. Paleontologists have two names for them, subspecies and variety, both having been used in various ways and interchangeably.

The International Rules recognize only sub-species, but they studiously, and perhaps properly, avoid definition of such a unit. Neozoologists, in general, use "subspecies" for geographic variants whose differences are presumably genetic. Unless paleontologists ignore the other categories, they might call all six of them "subspecies," even if this is not desirable. Also it is doubtful how many of these types of variation paleontologists can recognize.

Geographic variants. As shown above, there is a minimum range of variation of the mean \( \left( M \pm 3 \sqrt{2\sigma^2} \right) \) within which we can not validly differentiate between samples. Thus if the mean of a character of sample A is found to range from 15 to 20 mm. the mean of another sample, B, should lie outside this range if we are to conclude that sample B differs significantly from sample A.

A unit so differentiated has been termed a species above. If we recognize subspecies, their means would fall within the range of the mean of the species and the differentiation of subspecies would not be statistically valid. (It is assumed in the preceding statements that the sample for the species is a composite over its geographic range.) The only alternative is to extend the limits of the species so that units defined as above would be considered subspecies.

Another factor complicates the situation. So far we have assumed that we are dealing with synchronous populations (or a synchronous population). Generally, however, we cannot be sure that we have exact time equivalents in geology, particularly if notable distances are involved. With rare exceptions, such certainty does not exist, even to within a few thousand years. This means that in most cases, time and geographic variants are difficult to separate.

Environmental effects add to the difficulty. In modern faunas, environmental variants may be studied by breeding or other experiments as with the brine shrimps. Obviously, however, one cannot breed ammonites or fusulinids, but the problem is no less real. If such variants are suspected among fossils, one must search the environment for some parallel change. Nothing has yet been done along these lines except in the grossest manner. The difficulties are, of course, numerous. Many environmental factors left no known impress on the accompanying sediments. (It may be that this is mostly a matter of ignorance; if so, the fundamental research is yet to be done.) Diagenetic changes altered the composition and texture of sediments, and post-diagenetic changes may have changed them further almost beyond recognition. In the face of a problem so appallingly intricate it is small wonder that estimates of depositional environments have been qualitative.
Time variants. Many paleontologists have attempted to define subspecies as time variants. Evolutionary changes should, of course, show themselves in a succession of descendant populations. We are faced here, however, with the difficulty of distinguishing environmental and genetic variations and it is not likely that much can be done to separate them. In dealing with synchronous specimens there is some possibility that certain environmental influences can be recognized. In time series, it may be expected that genetic and accompanying phenotypic changes will have occurred even in a “constant” environment. If the environment changed we may expect that organisms evolved to adapt themselves to the changed environment. The difficulty in distinguishing phenotypic changes, produced by a change in the genotype, from phenotypic changes unaccompanied by genotypic change introduces much uncertainty.

Morphologic variants in local populations. Many paleontologists have dealt with individual variants as though they were infraspecific units. Genetically various though the individuals of a population may be, modern neozoologists are very hesitant about recognizing more than one infraspecific unit at any particular place. This is easily understandable. Since a species is an actually or potentially interbreeding population, the subdivisions also will be. If two morphologically distinct subdivisions of this sort come in contact it is to be expected that interbreeding would soon reduce them to a single unit. Thus a local interbreeding population will, in the aggregate, be genetically homogeneous, isolation or some such factor being necessary to develop subspecific or specific differences.

It is true that even minute environmental variations, particularly size variations, are very important in local populations. It is such variations that many paleontologists have termed subspecies and varieties.

Summary of infraspecific units. Certain conclusions may be drawn from the above considerations. First, there seems to be little reason to distinguish variants within a synchronous sample from one locality. Such variants are probably genetically homogeneous and there is no theoretical reason for a subdivision. Subdivisions based on sizes or counts are especially illusory as they can be adequately accounted for on the basis of variations in environment, amount of food, etc. The only result of recognizing subspecies under such circumstances is to burden the literature and no useful purpose is served.

There seems to be little value in recognizing time variants in units smaller than species. Probably we are dealing with evolutionary series and there is a minimum size to significant units into which they can be split. The size of these units has nothing to do with the label attached to them, whether it be species or subspecies and it seems most useful to consider the minimum unit a species. If someone chooses to term it a subspecies, this can hardly be designated an error, but there seems little reason for such a course.

On the other hand, the subspecies is a useful unit for distinguishing synchronous geographic variants. When such variation is recognized, the entire series can be considered a species and the validly recognizable subdivisions termed subspecies. Such variation may be primarily either genetic or environmental. In fossil material it may not be possible to distinguish these influences but in some cases the evidence may point more strongly to one than to the other. If the control is genetic, the geographic variants should be termed subspecies, as they would be by the neozoologist. If environmental control is primary, the significance is not the same, but because of uncertainty it seems best to term such units subspecies also.

In the case of the West Coast Elphidiums, it would seem best to consider Nicol’s valid subdivisions as subspecies. If the entire group is specifically distinct it could be termed E. fax with the subspecies fax, concinnum, and perhaps excubit.

Ontogeny and Heterogony

A feature noticeable in most paleontological literature is the lack of attention paid to a quantitative study of growth stages. Many persons have ignored them as though they did not exist. Consequently, little regard has been paid to sizes when comparisons were made between samples. If the various parts of an animal maintained a
uniform ratio with respect to each other during growth, this might not be so important, but generally the proportion between any two characters changes constantly during growth. For example, if the ratio of the length of the deltoid plate to the total height of Pentremites pyriformis is plotted against time, we get the graph shown in figure 7. In the youngest stage the deltoid is a little more than 0.1 of the height. As the blastoid grew the deltoid increased in size proportionately faster than the height and eventually reached a proportion of about 0.3 of the total height. In the final growth stages, however, the height increased at a somewhat greater rate so that the ratio falls to nearly 0.25. Consequently it is impossible to specify a certain deltoid to height ratio as a character distinguishing P. pyriformis. Neither can the size of the deltoid, nor the height, be used without qualification since both changed constantly during growth. Most invertebrates do not have a terminal size, as do many of the higher vertebrates (mammals); instead their parts continued growing throughout life but not necessarily at the same rate. Also the size attained by these animals at any given age is not constant.

Authors, in trying to escape these difficulties (too often dimly realized), have commonly used what they called "mature specimens" in setting up standards for a species. This is a comfortably vague term, but it is a qualitative term covering a good part of the life span of most animals and it is quite unsuited to precise work. Some authors, including the present one, have attempted to meet this problem by constructing "ontogenetic" curves for various characters and then specifying the size of the character at some particular growth stage. From what has gone before it should be clear that it is absolutely necessary so to specify the size of any character. However, so far as I know, no analysis of the method of doing this has yet been presented.

Ontogeny may be defined as the history of the development of an individual organism or of the individuals of a contemporary group. An ontogenetic curve is a graph showing this development visually in quantitative terms. In practice, of course, both are considered in terms of the development of individual characters. In either case, the essential feature of the whole idea is the change of a character or characters in terms of time. We can, for example, measure the height of a number of human males at birth and on their subsequent birthdays, and then

![Graph showing the deltoid to height ratio of Pentremites pyriformis as plotted against a "time" character (the standard radial).](image-url)

*Fig. 7—Graph showing the deltoid to height ratio of Pentremites pyriformis as plotted against a "time" character (the standard radial).*
construct a curve showing the relationship between time and height in human males. This would be a truly ontogenetic curve.

On the other hand, if we deal with fossil material, no such time scale is available. We have the hard parts of animals which died at different ages (stages of development). We know that, in general, size is directly correlated with age, but because of normal variation, a younger animal may be notably larger than an older one. (The influence of sexual dimorphism will not be treated here. In most of the usable fossil invertebrates it does not seem to be very important.) With these factors in mind, note how ontogenies of fossil animals have been reconstructed. In every case, one size, or count, or angular relationship is plotted against another such character. Thus we are not dealing with true ontogenies which are plotted against time, and also our curves may not even be an accurate secondary plot against time. We have, in fact, constructed curves of heterogony showing the relative development of two characters.

These considerations are illustrated by the following example. Suppose we have four specimens with the following characteristics:

<table>
<thead>
<tr>
<th>Specimen</th>
<th>Length</th>
<th>Height</th>
<th>Relative Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>

If height is plotted against length, as might be done in constructing the usual "ontogeny," we obtain the lower line shown in figure 8. If the relative ages are as indicated above, as they well might be, the upper line more nearly represents the true ontogeny. The trends of these two lines are quite different.

Another factor may be considered in relation to the actual case of Pentremites pyriformis illustrated in figure 7. It might be assumed that, on the basis mentioned in footnote 4, each division of the horizontal scale represents an equal amount of time and that there is a direct, uniform (in this case exponential) relationship between the length of the standard radial and time. Instead, let us suppose that the length of the standard radial is negatively accelerated with respect to time; that is, that each successive division of the horizontal scale of figure 7 represents less time than the one preceding it. If so and if the graph be adjusted to give a uniform time scale, the

![Fig. 8—Theoretical graph showing a possible relationship between ontogeny and heterogony in a series of individuals.](image-url)
commonly be considered an ontogenetic curve, but of course neither character is actually plotted against time. If, however, $Y$ were plotted against time and a true ontogenetic curve drawn, each point would lie within a certain vertical distance of the curve, this distance being determined by the inherent variability of character $Y$ at any particular age. If we knew the standard deviation of $Y$ from this curve, and drew through each point a vertical line of length equal to $Y \pm 3\sigma$, the ontogenetic curve would pass through all the vertical lines at one place or another ($P = 0.997$).

Character
STUDIES IN QUANTITATIVE PALEONTOLOGY

We can draw circles of a radius $3\sigma$ instead of horizontal and vertical lines. A moment's consideration will show that the length of the radius is of secondary importance so long as it is uniform and long enough to allow a line or simple curve to cut all the circles. If the radii are needlessly large, the two bounding curves will be far apart, but uniform with respect to the midline. If we construct suitable circles as in figure 11, and

**Fig. 12**—Graph showing the estimation of the relative age of individuals in a sample of *Pentremites pyriformis.*
proceed as above a midline curve is obtained expressing a relationship of $X$ and $Y$ to time. If we now drop perpendiculars from the plotted points to this curve, the intersections of these lines and the curve will be the most probable positions from which each specimen varied in $X$ and $Y$. These are points in time and their relative positions on the curve indicate the relative ages of the specimens. Finally, if we project these points onto an arbitrary scale on the $X$-axis, we can read off the probable relative age of each specimen studied.

An actual example is shown in figure 12, where height of the blastoid *Pentremites pyriformis* is plotted against the length of the standard radial. (These terms are defined below in the section on blastoids.) These two characters were selected because they show closer correlation than others and would probably give the best results. The circles drawn around the points have a radius arbitrarily chosen but probably approximating $3\sigma$. The original data indicate that the regression line is slightly curved and allowance was made for this in drawing the two outside curves, one as far to the upper left as possible, and still touching all circles, and the other as low as possible to the right. The time curve is midway between these and the original points were projected onto it. These final points may in turn be pro-

---

**Fig. 13**—Dispersion of points representing the height of *Pentremites pyriformis* plotted against the standard length compared to the dispersion resulting from plotting height against relative age.
JECTED ONTO AN ARBITRARY TIME SCALE ALONG THE X-AXIS.

OF COURSE SUCH A PROCEDURE CAN LAY NO CLAIM TO GREAT ACCURACY. IF WE WERE TO USE A DIFFERENT PAIR OF CHARACTERS, WE WOULD ARRIVE AT A SOMewhat, THOUGH NOT WIDELY, DIFFERENT SET OF RELATIVE AGES. ALSO THERE MAY BE A MULTIPLYING OR EXPONENTIAL FACTOR INVOLVED WHICH, ALTHOUGH NOT AFFECTING THE AGE-ORDER OF THE SPECIMENS, MIGHT GREATLY AFFECT THEIR SPACING.

WITH THESE RESERVATIONS RECOGNIZED, THE EFFECT OF PLOTTING SOME CHARACTER AGAINST RELATIVE AGES MAY BE COMPARED WITH PLOTTING THE SAME CHARACTER AGAINST ANOTHER. IN FIGURE 13 THE LENGTH OF THE DELTOID OF *P. PYRIFORMIS* IS PLOTTED AGAINST THE RELATIVE AGES (POINTS SHOWN BY SOLID CIRCLES) AND ALSO, ON A PRECISELY COMPARABLE SCALE, AGAINST THE STANDARD RADIAL (POINTS SHOWN BY OPEN CIRCLES). THE POINTS PLOTTED AGAINST RELATIVE AGE HAVE SOMewhat LESS SCATTER THAN THE OTHERS, AS MIGHT BE EXPECTED. WHETHER OR NOT THIS IS ADVANTAGEOUS MAY BE DETERMINED BY CONSIDERING THE REGRESSION LINES FOR THE TWO SETS OF DATA.

BY THE USUAL METHODS THE REGRESSION LINES ARE CALCULATED AS FOLLOWS: DELTOID = 0.38 (RELATIVE AGE) — 1.0 AND DELTOID = 0.77 (STANDARD RADIAL) — 2.5. THESE TWO LINES, SHOWN IN FIGURE 13 (DELTOID AGAINST RELATIVE AGE, CONTINUOUS LINE; DELTOID AGAINST STANDARD RADIAL, DASHED LINE), ALMOST COINCIDE AND ARE SO SIMILAR THAT THERE IS NO NEED TO COMPARE THEM STATISTICALLY. EVIDENTLY PLOTTING THIS CHARACTER AGAINST TIME HAS NOT CHANGED THE TREND OF THE RESULTING CURVE, AND, CONVERSELY, IT MAKES LITTLE DIFFERENCE WHETHER THIS CHARACTER BE PLOTTED AGAINST TIME OR ANOTHER CHARACTER SO FAR AS THE TREND OF THE REGRESSION LINE IS CONCERNED.

NEXT, THE DEVIATIONS FROM ITS REGRESSION LINE ARE CALCULATED FOR EACH SET OF POINTS, AND FROM THIS THE STANDARD DEVIATIONS ARE OBTAINED. FOR DELTOID AGAINST RELATIVE AGE, \( \sigma = 0.69 \text{ mm.} \), FOR DELTOID AGAINST STANDARD RADIAL, \( \sigma = 0.76 \text{ mm.} \). THESE STANDARD DEVIATIONS ALSO INDICATE THAT THE FIRST SET OF POINTS HAS SOMewhat LESS SCATTER THAN THE SECOND. HOWEVER, IF THE TWO ARE COMPARED, WE FIND THAT \( d/\sigma_a = 0.5 \) WHICH IS DECIDELY NOT SIGNIFICANT \( (P = 0.38) \). \( (d/\sigma_a = \sigma_1 - \sigma_2/\sqrt{\sigma_1^2 + \sigma_2^2} \) samples are of the same size.) thus, plotting of deltoid against time (or a time factor) has not significantly improved the scatter or, conversely, it makes little difference in the scatter whether the deltoid is plotted against time or another character.

OTHER EXAMPLES COULD BE CITED TO SHOW THE SAME RELATIONS, BUT THIS WOULD BE MERELY REPETITIOUS. THE CONCLUSION TO BE DRAWN IS THAT, UNDER CERTAIN CONDITIONS, WE ARE JUSTIFIED IN USING CURVES OF HETEROGENY AS ONTOGENETIC CURVES AS THEY PROBABLY ARE NOT GREATLY IN ERROR. THE CONDITIONS ALLOWED TO NEED SOME EXPLANATION,

IF THE NECESSITY FOR THE USE OF GROWTH SERIES IN PALEONTOLOGY BE GRANTED, IT IS DESIRABLE FOR THEM TO BE AS UNIFORM AS POSSIBLE WITHIN ANY PARTICULAR GROUP OF FOSSILS. THEREFORE IT IS NECESSARY TO SELECT ONE CHARACTER TO BE USED THROUGHOUT THE GROUP AS A "TIME" CHARACTER. FOR EXAMPLE, IN THE SECTION ON BLASTOIDS, THE SELECTED CHARACTER IS THE "STANDARD RADIAL." ALL OTHER BLASTOID CHARACTERS ARE THEN STUDIED IN RELATION TO THIS ONE. THIS "TIME" CHARACTER SHOULD HAVE CERTAIN ATTRIBUTES. IT SHOULD BE PRESENT AND MEASUREABLE IN ALL GROWTH STAGES; IT SHOULD CHANGE SLOWLY AND UNIFORMLY AND MEASURABLY DURING GROWTH; IT SHOULD, IN SO FAR AS POSSIBLE, BE FREE FROM GERONIC EFFECTS; IT SHOULD BE PRESENT IN A USABLE FASHION IN ALL MEMBERS OF THE GROUP; IT SHOULD NOT BE SUBJECT TO DRASTIC MODIFICATION IN ANY SPECIES. IT MAY NOT BE POSSIBLE TO SATISFY ALL THESE REQUIREMENTS FULLY, BUT EVERY EFFORT SHOULD BE MADE TO DO SO.

PRESENTATION OF DATA

THUS FAR EMPHASIS HAS BEEN PLACED ON THE THEORETICAL ASPECTS OF QUANTITATIVE PALEONTOLOGY. IF THE APPLICATION OF THESE METHODS TO ACTUAL FOSSIL MATERIAL IS ATTEMPTED, A VARIETY OF PROBLEMS ARISE, ONE OF THE FOREMOST OF WHICH IS THE PROBLEM OF DECIDING ON THE MOST SUITABLE WAY TO ORGANIZE AND PRESENT QUANTITATIVE DATA. THE FOLLOWING DISCUSSION IS DEVOTED TO THE GENERAL PRINCIPLES INVOLVED. ACTUAL EXAMPLES OF APPLICATION OF THE METHODS HERE DISCUSSED ARE PRESENTED IN THE SUCCEEDING SECTIONS.

LITTLE DIFFICULTY IS ENCOUNTERED IN THE PRESENTATION OF THE DATA OF QUALITATIVE PALEONTOLOGY. ONE SIMPLY DESCRIBES IN WORDS WHAT HE CONSIDERS TO BE THE IMPORTANT FEATURES OF THE FOSSIL IN QUESTION, AND PERHAPS SUPPLEMENTS THIS DESCRIPTION BY MEAS-
measurements of a few specimens.

If, instead, the study is to be quantitative, we are confronted by a mass of raw data both ponderous and indigestible. Our first concern is to reduce this mass of data to manageable and comprehensible proportions. This can be done partly by graphical and partly by mathematical means. The original raw data should in all cases be presented, however. This allows others to check the author's conclusions with an exactness and fairness not otherwise possible.

There are, essentially, three things we wish to know about any character: its size, its variability, and its relations to other characters. The size of a character cannot, of course, be given by any single number. It varies during growth in any individual and, at any particular growth stage, it varies between different individuals. The usual, and probably the best, way of dealing with this situation is to report the mean or average size. This is a measure which must be qualified. If the specimens dealt with are of essentially similar age, a simple situation but one to be avoided, the mean will give a logical measure of the central members of a population. However, if more than one growth stage is present in the sample, the mean is a biologically meaningless measure. This may be illustrated by a theoretical example. Suppose that we have a series of ten specimens whose respective lengths are 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 cm. It should be apparent that we are dealing with a growth series. If we calculate the mean of this series \( \frac{\sum X}{N} \), we obtain the figure 5.5 cm. which has practically no significance. It has precisely the same meaning as the statement that the average height of human males is that of a nine year old boy because his height becomes midway between that of an infant and an adult. It is necessary, therefore, to specify the average size of each growth stage. Since there is an infinity of growth stages for any species, we should have to specify an infinity of averages. The simple way of handling this situation is to give the equation of the line which passes through the successive averages. This is the regression line which should be used without exception in all cases involving growth series. Although a curved regression line may sometimes fit the data better, the equation for a straight-line regression should always be used, because there is no simple method for comparing two curved regression lines in a test for significance. As a convention, the regression should always be expressed as one of the dependent on the independent variable \( Y \) on \( X \).

The variability of a character is most simply stated in terms of the standard deviation. In a sample involving substantially one growth stage, this measure can be calculated from the deviations of the observed readings from the mean. In the case of growth series, however, the standard deviation must be calculated from deviations from the regression line in order to be meaningful. It can readily be observed that the absolute amount of the deviation from the regression line is less in the early than in the late growth stages but that the percentage or relative deviation remains much more constant. The standard deviation from the regression line should actually vary continuously during growth, being least in the youngest stages and, thereafter, steadily increasing. If the standard deviation be calculated as suggested above, we obtain not a variable but a single standard deviation. A moment's consideration will, however, show that we have, essentially, obtained the standard deviation of the latest growth stages where the absolute variability is the greatest. Of course, the total effect is to reduce the emphasis on the amount of scatter since the lesser scatter of the early stages is "averaged" in with the greater scatter of the later stages. It is believed, however, that a larger sample, consisting of all growth stages leads to better results in most cases than a smaller sample available for any single growth stage. One can plot the regression line on a logarithmic or semi-logarithmic graph and mark off the standard deviation at its upper end either above or below the regression line. A line is then drawn through this point and parallel to the regression line. The difference between these two lines at any particular growth stage will then give a usable value of the standard deviation for that stage.

The relations between characters may be shown by ratios, the usual way, or by plotting one character against another. In either case, the data are best presented
graphically. A table of numerical figures does not in itself mean very much but if presented graphically, measurements can be comprehended at a glance.

These last considerations raise the question of the most suitable manner in which data may be presented graphically. The most generally usable type of graph is the line graph. Such types as the pie graph, bar graph, etc., may sometimes be useful but they have no such general application. However, a line graph may be constructed in a number of different ways. The most common type of line graph is arithmetic in which the coordinates are spaced at equal intervals. This type is widely used but is not necessarily the best for biological data, particularly those involving growth series. Most ontogenetic curves are found to conform more or less well to a curve of the type \( Y = a + bX^k \). If this type of curve be plotted on an arithmetic scale, it will often be found to rise with undue steepness in the later growth stages, a familiar property of exponential curves. Further, the absolute variation is so much greater in the later than in the early growth stages that, relatively, the amount of variation shown by older specimens is overemphasized and that of the younger specimens underemphasized.

If, however, this curve is plotted on a logarithmic scale, it will usually be found to approach a straight line. Those which are curved on a logarithmic plot are affected by a number of factors not particularly pertinent to the present discussion, but they are more manageable than those resulting from arithmetic plots. The variability is nearly constant on a logarithmic plot; that is, the lines containing the variability are almost parallel rather than widely divergent as in arithmetic plots. The total effect is that the young and the old stages are given equal weight on the graph which is surely desirable.

In some cases, the dependent variable may be plotted against the whorl or volu- number, as in cephalopods or fusulinids. In such fossils, the spiral formed by the shell is a logarithmic (exponential) function and thus the successive whorl numbers will be spaced logarithmically. In these cases, the graph should be semilogarithmic, the dependent variable being plotted on the logarithmic and the independent variable (whorl number) on the arithmetic scale.

In some instances, triangular graphs may be used to good advantage. These are discussed below in connection with specific problems.

**PRACTICAL APPLICATIONS OF QUANTITATIVE METHODS**

The remainder of this paper is devoted primarily to applications of the methods explained above. In addition, certain developments and consequences of the theoretical considerations are brought out. Five types of animals are discussed: the blastoids typifying those whose skeletons are made up of separate plates, the pelecypods and the brachiopods typifying those which have two valves, the gastropods typifying those whose shell is spiral and whose protoconch is usually not preserved, and the fusulinids whose shell is spiral with an internal protoconch (proloculum). The great majority of fossil invertebrates fall into one or another of these shell types.

**Blastoids**

The material used for this study consists of a collection of *Pentremites* from the Paint Creek formation of Chester age, collected from an outcrop 1½ miles west of Floraville, Illinois. One part of this collection consists of the characteristic and easily separable *P. pyriformis*. Data concerning this form have been used above. The rest of the sample consists of specimens usually referred to *P. godoni*.

The study of blastoids has been almost entirely qualitative. The first problem in this study, therefore, was the setting up of a standard system of measurements. The one selected is suitable to the forms studied but it might have to be modified slightly for other genera.

The blastoids are invertebrates whose skeleton is composed of discrete plates. Such forms usually carry no record of their early growth. The blastoids studied do show growth lines, but the difficulties involved in trying to reconstruct growth series from them are almost insuperable. For all practical purposes, one must rely on young and old specimens to construct such series. This emphasizes the necessity of complete and
Fig. 14—Ratios of various characters of *Pentameres* *pyriformis* as plotted against the standard radial. Each has been multiplied by a constant factor as indicated. 1—number of side plates to length of ambulacrum in mm. (X4); 2—azygous basal to deltoid (X6.7); 3—height to standard radial (X7); 4—side plates to height in mm. (X6); 5—ambulacrum to standard radial (X11); 6—base of radial to standard radial (X12); 7—ambulacrum to height (X14); 8—deltoid to ambulacrum (X2.75); 9—deltoid to standard radial (X16); 10—deltoid to height (X16).

thorough collecting of such forms. (Ulrich made complete collections of some Mississippian blastoids and then described the young forms as separate species.)

Only in post-Meramec rocks are blastoids usually numerous enough for large samples to be collected and, therefore, small samples must be dealt with in many cases. This does not mean that statistical methods cannot be used, but rather that the need for them is increased.

The plate structure of blastoids is quite stable. Therefore, the size and relationships of the plates may be used in studying blastoid species. The “time” character, as usual, is one of primary importance. It was finally decided to use a measurement of the radial plate for this, the independent variable. The radial plate shows a minimum effect of geronticism, and it is as easy to measure as any other plate. A measure of the total height of the calyx would be easier to obtain but it is a composite of many other measures and thus unsatisfactory for this purpose. The particular measure chosen was the length of the outer margin of the left ray of the anterior radial from the lowest part of the base of the deltoid to the top of the basal plate. This measure is termed the standard radial. Other measures are as follows: base of the radial, from the base of the anterior ambulacrum to the top of the basal below along the surface of the plate; length of the ambulacrum, from the lowest part of the notch of the anterior radial to a height equal to the left anterior spiracle; number of side plates, counted on the left side and including the spiracle; the length of the azygous basal, in its longest vertical dimension; the height, from the base of the calyx to the top of the mouth; and the thickness, from the surface of anterior ambulacrum horizontally through the specimen at its thickest part to the surface of the posterior interambulacrum.

The ratios between individual characters are often good specific “characters” themselves and ten such ratios for *P. pyriformis* are shown in figure 14. Each curve is nearly a straight-line up to a point where the standard radial length is about 8 mm. At that stage, they tend to change slope markedly and level off, and the ratios become more nearly constant. This is an indication of maturity and *P. pyriformis* may be considered to become mature when the standard radial reaches this length. At the point where the standard radial is about 11 mm. long most of the curves show a rather
definite reversal of trend. This size, therefore, may be considered to mark the beginning of geronticism. The fact that all of these ratios change greatly during growth is very clear.

While measuring the specimens of P. godoni, it became evident that the sample of over forty blastoids could be split into two parts, one characterized by a somewhat stellate horizontal section associated with a rather low deltoid to standard radial ratio, and another with a more rounded section.

<table>
<thead>
<tr>
<th>Character</th>
<th>P. godoni A</th>
<th>P. godoni B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>( Y = (1.235, 1.410, 1.615)X^{1.146} )</td>
<td>( Y = (1.202, 1.436, 1.713)X^{1.185} )</td>
</tr>
<tr>
<td>Thickness</td>
<td>( Y = (1.233, 1.552, 1.948)X^{1.060} )</td>
<td>( Y = (1.400, 1.640, 1.930)X^{1.068} )</td>
</tr>
<tr>
<td>Base of Radial</td>
<td>( Y = (0.356, 0.487, 0.670)X^{1.072} )</td>
<td>( Y = (0.387, 0.507, 0.665)X^{1.077} )</td>
</tr>
<tr>
<td>Deltoid</td>
<td>( Y = (0.341, 0.428, 0.540)X^{1.202} )</td>
<td>( Y = (0.353, 0.480, 0.662)X^{1.301} )</td>
</tr>
<tr>
<td>Azygous Basal</td>
<td>( Y = (0.185, 0.262, 0.384)X^{1.100} )</td>
<td>( Y = (0.195, 0.270, 0.362)X^{1.222} )</td>
</tr>
<tr>
<td>Ambulacrum</td>
<td>( Y = (0.901, 1.052, 1.233)X^{1.218} )</td>
<td>( Y = (0.905, 1.157, 1.475)X^{1.217} )</td>
</tr>
<tr>
<td>No. of Side Plates</td>
<td>( Y = (3.45, 4.21, 5.17)X^{1.057} )</td>
<td>( Y = (3.92, 4.97, 6.29)X^{1.048} )</td>
</tr>
</tbody>
</table>

and a markedly higher deltoid to standard radial ratio. The first of these will hereafter be referred to as P. godoni A and the second as P. godoni B. It was believed that the two characters used could lead to a valid separation and not a spurious one of the type referred to in the foregoing section on the recognition of synchronous species. The stellate or rounded section has no obvious relationship with the deltoid to standard radial ratio and a correlation between the two might be genetic rather than structural.

When the two sub-samples were plotted on a graph, however, it became evident that they are not at all well separated. Figure 15 is a graph for the base of the radial, which is thoroughly typical of the other characters, plotted against the standard radial. (Data for P. pyriformis are included for comparison.) The two sub-samples overlap almost completely but each of the curves for P. godoni A is below the comparable ones for P. godoni B. Next, these differences were tested for significance. A comparison of the regression lines indicated that the difference is not significant. Because of the constant relationships of the two sets of curves, this test, which compares only a pair of characters may not be conclusive. Therefore, the more decisive, but more laborious test of significance, that of multivariate analysis, was made. Thus all seven investigated characters of each sub-sample were compared simultaneously to give as nearly as possible a comparison of the fossils as a whole. It revealed that there is less than one chance in one hundred that the sub-samples, drawn from a population with a variability of the two combined, would differ as much as they actually do. Therefore, P. godoni A and P. godoni B. probably do differ significantly, and apparently represent two species. The table below lists the equations of the regression lines of these two species:

![Fig. 15—Comparison of the ontogenetic development of the base of the radial of Pentremites pyriformis, P. godoni A and P. godoni B. The outer pair of each triad of lines defines the probable limits of variation of that particular population.](image-url)
$Y$ is the dependent variable, the character listed to the left in the table, and $X$ is the independent variable, the length of the standard radial. In each equation $X$ has three coefficients. The second represents the regression line. The others represent the limiting curves of minimum and maximum variability and are spaced $\pm 3\sigma$ from the regression curve. All equations are based on measurements in millimeters. Thus with

2. On the slide rule, place the sliding index over the value of $X_1$ on the appropriate log log scale.
3. Move the right or left index of the $C$ scale over this.
4. Move the sliding index over to the exponent of $X$ (this is the exponent of $X$ in the regular regression equation) on the $C$ scale.
5. Read the answer from the appropriate log scale. These operations have raised $X_1$ of step 2 to a power, $X_1^b$.
6. Set the sliding index over the value of $Y_1$ on the $D$ scale.
7. Move the value of $X_1^b$ (step 5) on the $C$ scale over $Y_1$.
8. Read the result on the $D$ scale under the right or left index. This is the value of the coefficient $a$ in $Y = aX^b$.
9. The equation for the line in question is, then, $Y = aX^b$, where $a$ is obtained from step 8, and $b$ is the exponent of $X$ in the regression equation.

A further indication of the justification for separating $P. \text{godafone}$ $A$ and $P. \text{godafone}$ $B$ is to be found in a triangular graph. Such graphs are a simultaneous plot of three characters but are peculiar in that the measurements of a character must be expressed as a per cent of the sum of all three. Thus it is essentially a graph of proportions. The standard radial, deltoid, and height are plotted against one another in figure 16. The pattern formed by each set of points does not depart widely from circularity. (For the sake of clarity, the individual points are not shown, but only the lines enclosing the areas occupied by the points.) The patterns of $P. \text{pyriformis}$ and $P. \text{godafone}$ are completely separated. The pattern of $P. \text{godafone}$ $A$ and $P. \text{godafone}$ $B$ overlap considerably, but indicate variation that seems to be distinctly different.

**Brachiopods**

The brachiopods, like the pelecypods which are considered in the next section, are characterized by having two valves. Each is actually a spiral shell with a very rapid rate of expansion (Thompson, 1942). This is so great that methods used with such spiral shells as the fusulinids are not suitable. The valves often have growth lines so that growth series for many characters, partic-
ularly external ones, can be reconstructed from a few specimens. Changes occurring in the growth of other characters, principally internal ones, cannot be so determined, however. This is probably of small importance in the study of species since internal changes occurring in the growth of other characters, principally external ones, can be reconstructed from a few specimens. Changes occurring in the growth of other characters, principally internal ones, cannot be so determined, however. This is probably of small importance in the study of species since internal Des Moines series of Livingston County, Missouri. Four characters were studied, the width (at the hinge line), the length, the number of lirae per mm. at the anterior end of the center of the ventral valve, and the number of spines on one side or the other of the beak. Growth lines are absent or discontinuous, so that it was necessary to rely on ordinary growth series. A total of 18 specimens was studied. The ontogenetic development of the length and the number of cardinal spines are shown graphically in figure 17. The number of lirae per mm. is not noticeably related to size, but varies from $5\frac{1}{2}$ to $6$ per mm.

This group of specimens, which show no signs of being heterogeneous, may be compared to several species and subspecies of Chonetina described by Dunbar and Condra (1932) whose descriptions are qualitative,
though perhaps better than average.

The most noticeable characteristic of the Exline Chonetina (hereafter referred to as Chonetina X) is its small size at maturity. The largest one measured has a width of less than 9 mm., and it seems doubtful that it attains a width of 10 mm. The length reaches 4.5 mm. and probably does not exceed 5.0 mm. All the other described forms are notably larger, varying from a “mature” width of 12 mm. in C. flemingi plebia to 22 mm. in C. flemingi alata. The number of cardinal spines of Chonetina X does not seem to exceed four on each side of the beak but other forms seem to have 5 to 10 on each side. C. verneuilianus may have four but its spines are nearly parallel to the hinge and not steeply oblique as in Chonetina X. The number of spines of C. flemingi crassiradiata has not been reported. The number of lirae per mm. in Chonetina X is 5½ to 6; C. rostrata has 6 to 7; C. flemingi alata has 4 to 5, and C. flemingi plebia has 5½ to 6. Data for other forms are not available.

Chonetina X is most obviously differentiated from other forms by being smaller but size alone must always be used with considerable caution in separating species. The length and number of spines of Chonetina X also are decidedly less than in most others. Perhaps, therefore, these specimens are simply dwarfs of some recognized species or “subspecies”?

I do not know of any published consideration that is particularly helpful in pointing out means for identifying dwarfs, which are primarily phenotypic rather than genotypic, with larger specimens of the same species. Therefore, it seems best to assume that if certain specimens are actually dwarfs, normal size might be attained by continuation of growth at its previous rate. This would involve a straight-line extrapolation of the growth curves and such a procedure for Chonetina X seems justified. This has been done in the graphs of figure 18 where the length and number of spines are compared with information available for other Chonetinas. In both respects the range of Chonetina X seems to fall below those of other Chonetinas but if the curves for Chonetina X should turn upward somewhat (if the extrapolation should be curvilinear in the proper degree) some of these ranges might coincide. There is no indication, however, that the curves of Chonetina X would actually do so and, on the basis of available data, we are probably justified in concluding that Chonetina X of the Exline limestone is not a dwarf but a distinct, undescribed species.

Pelecypods

Pelecypods, like brachiopods, have a test consisting of two rapidly expanding spiral shells. Their measurements are somewhat
STUDIES IN QUANTITATIVE PALEONTOLOGY

The length is the maximum dimension parallel to the hinge line, the height is the maximum dimension perpendicular to the hinge line, the angle \( \alpha \) is measured between the hinge line and the umbonal ridge, and the thickness is measured across the maximum inflation of the two comya which he examined by statistical methods and concluded to be distinct. Such a graph compares three characters instead of two. Although much used in the physical sciences they have not been employed in biology to any great extent. Construction is as follows: The measures of the three charac-

![Fig. 19—Triangular graph of the length, height and distance to maximum down-bulge of four species of Anthracomya.](image)

valves. A suitable measure for the “time” dimension in the pelecypods is difficult to obtain. Length, height, and thickness are easily determined but all are commonly subject to gerontic effects. The angle \( \alpha \) seems to be less affected, but it cannot be measured accurately and in many forms cannot be measured at all. All things considered, the length probably represents the best compromise for this purpose.

A triangular graph presents the data of Leitch (1940) on four species of Anthracomya selected are added together for each specimen, and their ratios in per cent are calculated. For example, character \( a = 2 \) mm., \( b = 3 \) mm., and \( c = 5 \) mm.; \( a + b + c = 10 \) mm.; \( a = 20\% \), \( b = 30\% \), \( c = 50\% \) of the whole. These percentages are plotted on the graph which thus shows proportions rather than absolute sizes. In figure 19 the characters used for Anthracomya salteri, A. modiolaris, A. adamsi and A. dolabrata are length, \( L \); height, \( H \); and distance from the anterior end of the shell to the point of
maximum ventral down-bulge, \( D \). A boundary line encloses the field defined by each set of points. Each species clearly occupies a rather definite area, although they overlap to a greater or less extent, and this indicates that we are probably dealing with distinct species.

The distribution fields of the blastoids are more or less circular (fig. 16) but for these pelecypods they are distinctly elon-

![Figure 20](image)

**Fig. 20**—Graph of the per cent length used in Fig. 21 plotted against the length for four species of *Anthracomya*.

gated. This elongation is related to the mode of growth because if the per cent of length is plotted against length (fig. 20), a definite trend in one direction is shown which shifts with growth and this change in proportions with growth is reflected in the triangular graph by an elongation of the fields.

The fields of *A. salteri* and *A. dolobrata* are elongated in one direction but those of *A. modiolaris* and *A. adamsi* are elongated approximately at right angles. Although the samples of the last three species are not large enough to permit great certainty, the available information indicates that the two groups have different patterns of development. The interpretation to be placed on this is uncertain. Possibly the two groups belong to different phyletic lines but data are not sufficient for a thorough investigation.

The triangular graph seems to have interesting possibilities in the study of fossils. It should not be used to reach definite conclusions, but it may be very helpful in indicating possible relationships and differences.

**Gastropods**

Spiral shells may be divided into two general groups, the planospiral, represented by the fusulinids and ammonites, and the helicospiral which is characteristic of most gastropods. In the latter group the protoconch is external and unprotected and for this reason it is commonly broken or worn away. In helicospiral shells, therefore, it is generally not possible to make measurements from the beginning point of growth as it is in the fusulinids. It is desirable, however, to use a whorl number as the independent “time” variable, but because the protoconch usually is not available as a beginning, it is necessary to set up another zero point. This may be done as follows: The position where the height of a whorl is exactly three mm., measured parallel to the general shell surface, is found and marked. This is designated whorl number zero. A line passing through this point is then drawn on the shell in the plane which passes through its axis of coiling. This line crosses successive whorls in the direction of the aperture which are designated No. \( -2 \), \( 2 \), etc. and similarly whorls in the direction of the apex are designated No. \( -1 \), \( -2 \), etc. This sets up a “time” dimension comparable to that in the other groups of fossils. Selection of a 3 mm. whorl height as the zero point is satisfactory for medium sized shells. A different height may be chosen for groups of larger or smaller species but this should be kept constant within a genus.

The present study is based upon specimens of *Lymnaea stagnalis appressa* Say which were aquarium reared by Dr. Lowell E. Noland of the Zoology department of the University of Wisconsin. Each died a natural death and the size range is typical, therefore, of “mature” individuals. All have the protoconchs preserved and the system of
whorl designation explained above can be checked against the absolute whorl number.

(measured from the external suture to the axis of coiling) and length of body chamber from its maximum downward extension to

These snails are devoid of ornamentation except for growth lines. The following characters, generally measurable in gastropods, were used: whorl height, radius vector the suture above, parallel to the axis of coiling. The number of actual whorls present in 12 specimens varies from 5½ to 7½ and variation in a natural population would be

![Graph showing the range of variation of the whorl height and radius vector as plotted against the whorl number in Lymnaea stagnalis appressa.](image-url)
at least as great. The designated zero volu-
tion occurs at 3 volutions in two specimens, 
at 3½ in seven, at 3½ in one, and at 3½ in two
This indicates a satisfactory whorl height to
volution number correlation. Figure 21 is a
graph of the whorl height and radius vector
plotted against the assumed volution num-
ber. Figure 22 is a graph of the height of the
body whorl plotted against the volution
number at the aperture. All show practically
straight-line relationships on a semilog-
arithmetic plot. This indicates that the
study of gastropods by these methods is
entirely practical.

Fusulinids

The fusulinids have, as a group, been
treated in a quantitative manner more
consistently than any other group of fossils.
White, Skinner, Dunbar, Condra, Henbest,
Newell, Keroher, Merchant, and Burma,
to mention only a few American authors,
have all used more or less quantitative
methods in dealing with them. These
methods have been used in a rather timid
manner, however, and a qualitative attitude
toward the data has always been evident.
For example, the writer (Burma, 1942)
published a paper on the *Triticites* in which
a quantitative handling of the data was at-
tempted. This attempt was half-hearted,
however, and deservedly failed. An effort
was made to define species on the basis of
their variation, which was certainly worth
while, but the variation was handled in an
essentially qualitative manner. Also, no
effort was made to compare species except
in a qualitative manner and, consequently,
it seems certain that some of the species
erected were distinguished on inadequate
grounds and are probably invalid.

The methods of measuring fusulinids
seem adequate for most purposes. The fol-
lowing data are commonly reported: half
length (or total length), radius vector (or
total width), form ratio, tunnel angle, wall
thickness, height of volution, septal count
per volution, and diameter of proloculum.
Each measure is reported in relation to
volution number, which is another charac-
ter reflecting the rate of expansion of the
spirotheca. The spiral is essentially log-
arithmic and therefore the length of the
spirotheca in successive volutions is de-
scribed by an exponential function and the
volution number should be plotted on an
arithmetic scale in making graphs. This
character is universally used as a "time"
character in fusulinids and seems to be well
fitted to the purpose.

The half length is measured on axial sec-
tions along the axis of coiling from the center
of the proloculum to the extremity of a
volution. Some authors have used the total
length and there is no *a priori* reason for
selecting one rather than the other except
that the half length is preferable if the form
ratio is to be considered.

The radius vector can be measured on
either axial or sagittal sections. In actual
practice, it should be measured on both,
the first being used to compute the form
ratio, and the second for reporting the char-
acter itself as explained below. In either
case the radius vector is the measure of the
logarithmic spiral of the spirotheca from
the center of the proloculum to the exterior
of a volution perpendicular to the axis of
coiling. Some authors have reported the full
width which is the sum of a radius vector of
one volution and that of another 180° from
it. The radius vector is more significant bi-
ologically and mathematically than the total
width and there are other reasons for pre-
ferring it.

The form ratio is the ratio of the half
length to the radius vector (or total length
to total width). It is usually reported as a
single figure, e.g., 2.0 meaning that the half

![Figure 22](image_url)

*Fig. 22—Graph of the height of the body whorl of *Lymnaea stagnalis* appressa* plotted against the
whorl number at the aperture.*
length is twice the radius vector. A total length to width ratio has been used by some authors but this is a compound measure which is less significant biologically. The length is a valid character, and may be measured at any growth stage but the width is the sum of the radius vectors of two growth stages half a volution apart. If is a compound measure, being a function of the width of the tunnel and the radius vector. As such it is undesirable, but it is so firmly entrenched that it seems hopeless to attempt to substitute the width of the tunnel, a more meaningful measure, for it. The wall thickness can be measured on either axial or sagittal sections. For the sake

![Diagram](image)

**Fig. 23**—Diagrammatic sagittal cross-section of a fusulinid showing various planes in which an axial section might be cut.

species are to be compared by multivariate analysis, the form ratio should not be used, as this would introduce duplication of data, increase the complexity of calculation and gain nothing.

The tunnel angle is the angle subtended by lines joining the sides of the tunnel with the center of the proloculum. It can be measured only on the axial sections. This of accuracy (see below), it should always be measured on sagittal sections.

The height of volution is the distance from the outside of one volution to the outside of the next. It gives no information not implicit in the radius vector. Although almost universally reported, its usefulness is very limited. Like the septal count it should be measured on sagittal sections.
The diameter of the proloculum is a very useful specific character but because few sections are perfectly centered it is difficult to use in a multivariate analysis.

As indicated above, the accuracy and reliability of measurements obtained from axial and saggital sections differ greatly. Saggital sections show the entire coil of the sprotheca and measurements made on them can be referred to a definite volution and thus to a definite growth stage. The axial sections, however, are unoriented. They may coincide with an infinity of planes each of which cuts the sprotheca in a different place. Thus in figure 23, any line segment labeled "2" might be reported as the height of the second volution. Thus measurements made on axial sections can not be accurately located and they may be expected to show greater, though spurious, variation than do those on the saggital sections. In a sample their planes may or may not be evenly distributed through 180 degrees. In one sample such sections might cluster around the orientation shown by plane C in figure 23, and in another around plane A. The tunnel angles of these two samples would not be comparable because they would show development at stages 135 degrees apart and they might erroneously indicate a nonexistent difference.

As an example comparison is made between actual measurements of Triticites secalicus and T. primarius. Measurements of the former were made from topotype specimens, and of the latter from specimens on the slides used in the original description of the species by Merchant and Kerohr (1939). The following comparisons are for the fifth volution:

<table>
<thead>
<tr>
<th>Axial sections</th>
<th>T. primarius</th>
<th>T. secalicus</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radius Vector</td>
<td>M = 0.788 mm.</td>
<td>M = 0.613 mm.</td>
<td>d/σ = 3.77 (0.67)</td>
</tr>
<tr>
<td></td>
<td>σ = 0.121 mm.</td>
<td>σ = 0.990 mm.</td>
<td>P = 0.999 (0.497)</td>
</tr>
<tr>
<td>Half length</td>
<td>M = 2.338 mm.</td>
<td>M = 1.589 mm.</td>
<td>d/σ = 4.49 (1.39)</td>
</tr>
<tr>
<td></td>
<td>σ = 0.415 mm.</td>
<td>σ = 0.413 mm.</td>
<td>P = 0.9999 (0.835)</td>
</tr>
<tr>
<td>Tunnel Angle</td>
<td>M = 46.0°</td>
<td>M = 53.1°</td>
<td>d/σ = 6.02 (2.92)</td>
</tr>
<tr>
<td></td>
<td>σ = 5.0°</td>
<td>σ = 10.35°</td>
<td>P = 0.99999 (0.997)</td>
</tr>
</tbody>
</table>

As usual, a value of d/σ ≥ 3.0 or more is considered to be significant. (It is a better measure here but leads to the same results.) In the axial sections the radius vector indicates the probability (P) that the two samples are different is 0.999 but the probability indicated by the saggital sections is only 0.497. Thus the axial sections greatly exaggerated the difference between these species probably as a result of inaccurate orientation.

Since the radius vector for the axial sections should show the same significance as the radius vector for the saggital sections, we may subtract 3.10 from the axial significance to reduce it to 0.67 as in the saggital sections. If we subtract the same amount from the significance figures for the half length and tunnel angle, their significance, shown in parentheses, is 1.39 and 2.92 respectively. (This method of treatment is very approximate, but it is probably as accurate as the data warrant.) These figures suggest that the radius vectors, half lengths, and septal counts of T. primarius and T. secalicus do not differ significantly but that the tunnel angles and wall thicknesses do. It is very common for different stratigraphic occurrence of fusulinids to be made the basis of different specific names. These two species, barely separable morphologically, are separated by eight cyclothems and the Missouri Virgil unconformity.

Such similarity of widely separated Triticites suggests the advisability of checking some species which occur closer together. For this purpose, T. collus of the Cement City limestone and T. caccus of the Argentine limestone, separated by only one or two cyclothems and both described by me
in 1942, were compared by multivariate analysis and the difference between them was found to be completely insignificant. These results are as important as they are unexpected. Although *T. collus* and *T. cacces* have similar measures, they differ from one another as much as a great many other fusulinid species. Consequently the entire classification of fusulinids on the species level appears to be questionable and almost certainly the fusulinids have been over-finely split.

**CONCLUSION**

The foregoing outline of quantitative invertebrate paleontology does not cover this field thoroughly and much remains to be done. However, some conclusions stand out clearly. Paleontological collections are samples. For valid results they must be recognized as such and treatment on any other basis is inadequate and likely to produce erroneous conclusions. Paleontology should be concerned with genetic units and eternal vigilance is required to prevent the recognition of morphological sections of intergrading populations as valid units. Fossils should be studied as growing, developing organisms, for otherwise comparisons will be made on very dubious grounds. Species are variable and this variability is an important characteristic of species. Certainly it is time for paleontologists to cease merely labeling specimens and to become paleobiologists.

**BIBLIOGRAPHY**


