INTRODUCTION TO QUANTITATIVE SYSTEMATICS

MONT A. CAZIER AND ANNETTE L. BACON

BULLETIN
OF THE
AMERICAN MUSEUM OF NATURAL HISTORY
VOLUME 93 : ARTICLE 5 NEW YORK : 1949

INTRODUCTION TO QUANTITATIVE SYSTEMATICS

INTRODUCTION TO QUANTITATIVE SYSTEMATICS

MONT A. CAZIER

Chairman and Associate Curator
Department of Insects and Spiders
The American Museum of Natural History

ANNETTE L. BACON

Santa Barbara, California; formerly Scientific Assistant
Department of Insects and Spiders
The American Museum of Natural History

BULLETIN
OF THE
AMERICAN MUSEUM OF NATURAL HISTORY
VOLUME 93 : ARTICLE 5 NEW YORK : 1949

BULLETIN OF THE AMERICAN MUSEUM OF NATURAL HISTORY
Volume 93, article 5, pages 343–388, figures 1–12, tables 1–7

Issued July 8, 1949

Price: $.50 a copy

CONTENTS

INTRODUCTION

ALTHOUGH MANY STUDENTS of systematic zoology know that the exact methods of mathematics have made available a relatively new and refined tool (statistics) with which increased accuracy of systematic interpretation can be obtained, few have yet availed themselves of the advantages offered by the application of quantitative analysis to systematic problems. Many conclusions that are biased by the personal opinions of the systematists can be eliminated or improved upon by the application of statistical methods to biological and morphological data. This does not imply that biologists should give up their studies of the intricacies of living organisms and become mathematicians or statisticians, or that systematic problems can be entirely solved or verified by statistical methods. On the contrary, there is nothing to replace the biologists' familiarity with and understanding of the variation of living organisms, the many factors influencing this variation, and the intricate ramifications. The study of figures and formulas pertaining to biology will not make a biologist of a mathematician. This in part accounts for the inability of those trained strictly in mathematics to appreciate and to interpret correctly biological problems by using mathematical tools alone. It is, however, this variation in living organisms that often necessitates the use of statistical methods. The statistician, through his studies on variation, variability, and chance phenomena, has made available this more exacting and inquiring method of analysis which, however, does not replace qualitative observation and knowledge of biological data but serves to supplement them. It remains, then, for the biologist to apply this method whenever possible in conjunction with the fundamental principles of biological differentiation in the interpretation of systematic problems.

Since genetics, in its present state of fluctuation and rapid development, is of little direct practical use to the systematist, morphological studies correlated with measurable biological and ecological differences have served by necessity as a basis for the analysis of populations. However, many theoretical features of genetics can be used profitably to supplement the morphological analysis which at the present time is the only practical means of studying indirectly the underlying fundamental genetic changes. Inasmuch as these morphological differences are used as indirect indicators of genetic differences, it is necessary that they be analyzed as accurately as possible, thus requiring the aid of statistical analysis in many cases.

In order to appreciate properly the value of statistics in systematics it is necessary to realize from the outset that virtually every character used to separate organisms can be evaluated mathematically. Differential characters such as morphology, biology, ecology, etc., can all be expressed in terms of numbers which can then be used statistically to prove the reliability of observed differences in these characters. Once these numerical measures of the observed differences have been made, it is the function of statistics to aid in proving their significance and reliability, in ascertaining how far divergence has gone, in selecting those characters that show the highest degree of plasticity (variation) and, inversely, those characters that are relatively stable. These measures when made on morphological characters can be correlated with relevant biological and distributional data in aiding the systematist to establish criteria for the separation of genetically distinct populations.

Quantitative measures are much more refined than qualitative expressions where individual judgment automatically ranks paramount. How much better it would be, for instance, if such commonly used and misinterpreted stock phrases as: medium sized, narrow; length 8 to 12 mm., width 2 to 3 mm.; were expressed as definite ratios or as variables with given observed limits accompanied by calculated population limits. Descriptions would cease to be necessarily vague and largely meaningless and would assume a valuable and reliable place in the definition of organisms.

The speculative nature of qualitative interpretation should be enough to cause the systematist to inquire into any method that might reduce the great odds operating against him. The investigator never knows the limits of variation in the biological unit he is studying; he knows only the limits of his

sample, which was taken from the unknown total population. From this sample he makes various conclusions pertaining to the total population about which he actually knows very little. It is, therefore, of prime importance to evaluate as precisely as possible the information gained from this sample and to utilize the most accurate means of interpreting the characteristics of the population. In many instances this may be done best by statistical analysis of the sample. Even then, any conclusions drawn are only probabilities and never certainties, but an estimate of the reliability of these conclusions is obtained by the application of these statistical methods.

That statistical methods have a definite place in systematic zoology has been proved by actual application in such fields as herpetology, ichthyology, mammalogy, ornithology, and paleontology. The lethargic response of systematic entomology to these methods is possibly the result of several factors. Among these we might mention that there are few workers in this field, as compared with most of the others, in proportion to the enormous number of described and undescribed species. The great need for descriptive work and the complexities resulting from this have drawn attention away from more refined studies. In many of the attempts that the writers have found in the literature to apply statistical methods to systematic entomological problems, the workers have developed the data only to the point of constructing frequency distributions, which, if unaccompanied by the proper calculations of significance and probability, add little or nothing except in the convenience of expressing the usual qualitative description.

Unfortunately, the books and papers dealing with statistical methods impress the average biologist as being technical and forbidding, as indeed many of them are. To those who are well trained in mathematics, the acquisition of statistical technique offers no particular difficulty. To many otherwise capable students, however, either because of inadequate preparation in mathematics or because their preparation is not recent, the application of statistical methods to biological data is more than ordinarily difficult. It is this need for simplification of explanation, illustration, and selection of pertinent statistical formulas that prompts the writers to compile this paper dealing with quantitative methods. The formulas are taken from the literature and are arranged, in the writers' opinions, in the order of their application in the systematic analysis of any group of organisms, being subject to modification in special cases. Many complicated features of statistics, such as theory and derivation of formulas, are purposely omitted. The methods and formulas presented are exemplified by data obtained in a study of the beetle genus *Omus*.

Biological workers with but little mathematical experience who wish to study statistics in greater detail than can be offered in this paper might begin with a book on statistics in general, such as the one by Thurstone (1928) which explains the various fundamental formulas in an easily understandable manner. For the application of these formulas to general biological data, the worker will receive invaluable aid from Simpson and Roe (1939) and Snedecor (1946). The systematist is referred especially to a series of excellent articles written by Klauber (1937–1941) dealing with the application of statistical methods in the solution of various systematic and biological problems in herpetology.

The nature of the systematic problem largely determines the methods to be followed and the formulas to be applied. This fact makes it rather difficult to present any fixed outline of procedure that will apply to all systematic problems and is possibly the reason why the writers have been unable to find such a scheme in the literature. Without attempting to discuss the many complications involved in constructing such a scheme or the possible ramifications that might arise during an investigation, the writers are presenting an organized outline of procedure that, it is hoped, will enable the systematist to solve his more common difficulties. In doing this it has been necessary to make a selection of the methods used, and the reader should bear in mind the fact that those given here are not the only ones that could be used nor are they always the shortest. An attempt has been made, however, to select the most accurate and newer methods available, whereas those that largely duplicate the procedure and are incidental to the solution of

the problems or are primarily of mathematical interest have been omitted.

Before we proceed with this outline of statistical application, it is necessary to give the reader some idea of its limitations and adaptabilities. Nearly all systematic problems resolve themselves sooner or later into the differentiation of biologically distinct organisms on the basis of morphological characters. These characters may be divided into two distinct groups: first, those based on structures that do not grow or change during at least a portion of the life of the organism; and second, those structures that continue to grow throughout the life of the organism. The methods outlined here are inadequate in the solution of problems dealing with characters that are subject to change in organisms that continue to grow throughout life. For methods applicable to these problems the reader should refer to Klauber (1937–1941) or to the standard texts dealing with growth ratios. In insects and many other organisms, the structures of any one stadium are subject to virtually no change in an individual and are therefore readily analyzed statistically. It is to data of this kind that the procedure in this paper properly applies.

Numerous individuals have given unstintingly of their time and knowledge to aid in this project. Without shifting any of the responsibility for the material herein contained, the writers would like to express their appreciation to the following: Drs. F. A. Beach, C. M. Bogert, E. O. Essig, L. M. Klauber, E. Mayr, C. D. Michener, J. A. Oliver, L. V. Searle, G. G. Simpson, and H. T. Spieth.

STATISTICAL METHODS

TERMS

SAMPLE: The actual group of specimens of any particular taxonomic unit that is available to the systematist. There may be many samples representing any specific or subspecific population.

POPULATION: Any closely allied (morphologically, biologically, etc.) group of individuals. It includes all the existing individuals of that group, including those that are unobtainable for analysis. Unless used with a general meaning, the term should be accompanied by the appropriate modifier to indicate the presumed taxonomic status of the population (specific, subspecific, etc.).

OBSERVED SAMPLE RANGE: The actual total amount of variation in a character, that is, the difference between the maximum and minimum individuals in a sample.

CALCULATED POPULATION RANGE: The range of the variation in the indefinite total population represented by the sample, that is, the sample mean plus and minus three standard deviations of the sample. This range theoretically includes approximately 100 per cent of the population providing that the frequency distribution of the sample approximates the pattern of the normal curve.

VARIATION: The difference in measurement values among individuals belonging to a single sample.

VARIABILITY: The proportional relationship between the range of variation and the mean size of the character.

SAMPLING

The specimens available to the systematist that represent the biological unit being considered form the sample. For example, a sample of a genus would be composed of the available samples of the species belonging to that genus; specific samples comprise the available individuals of the species and of any component subspecies. It is with these samples that the systematist attempts to delimit and to classify the unknown and always unavailable population. The relationships between these samples, irrespective of their sizes, and the total population out of which they were collected are always somewhat uncertain. Inasmuch as these samples are only representatives of the population, any conclusions drawn from them regarding similarities or differences are only personal judgments or statistical estimates as to the probability of like similarities and differences existing in the population from which the samples were drawn. If taxonomic studies are to have any real value the systematist must use these sample estimates only in so far as they furnish information about the population. From this it can be seen that it is necessary for the systematist to be able to evaluate his sample carefully and to know some of its desirable characteristics and the various methods of improving it.

Numerous samples taken throughout the total geographical area occupied by each taxonomic unit being studied are desirable for the analysis of that unit. Unfortunately, most systematic work begins with samples that are already in collections and that therefore can be improved upon only with great difficulty, often necessitating the collection of new samples in the field or the borrowing of material from other institutions or individuals. However, in order to be able to segregate and subsequently to evaluate samples properly, it is necessary to know something about their three main characteristics: homogeneity, size, and bias.

HOMOGENEITY

Every possible specimen should be collected and segregated into samples, preliminary to statistical analysis, according to the following factors:

LOCATION: Specimens making up each sample should be from a single location and preferably from a small area. The size of the area depends on the taxonomic unit.

ENVIRONMENT: The specimens should be from similar climatic, edaphic, and biotic situations.

TIME: They should have been taken within a limited range of time so as to preclude possible confusion resulting from contamination by seasonal variance.

AGE: They should all be of approximately the same age or in the same stage of development.

SEX: They should be of the same sex (a

sample for each sex) to avoid the effects occasioned by the existence of sexual dimorphism.

Size

The size of the sample necessary to give a reliable approximation of the population depends largely upon the range of variation of the character being studied; the larger the range of the character, the larger the sample necessary to approximate the population extremes (Klauber, 1941, p. 33). In zoology sampling is generally limited by the available material in collections so that in most cases all adequately labeled specimens should be used. Samples comprising 300 or 400 specimens are often too large to be handled easily and have been shown to add little to estimates based on 100 specimens. Single specimens often add considerable to our knowledge but should not be used in forming conclusions unless treated with the proper small (single) sampling formulas, and even then great caution should be exercised. It has been shown by statisticians and illustrated by Klauber (1941, p. 54) that samples of as few as five specimens have been sufficient to indicate considerable sample variation in some cases and that the increase in the observed range in the sample does not maintain a proportional increase after 25 specimens have been used. For instance, in the examples cited by Klauber, 25 specimens indicated almost the same range as 100, and 200 very little more than 100, there being no increase in the variation between 200 and 500. For all practical purposes, samples of at least 15 to 25 specimens may be used with good results, but samples of 50 to 100 specimens are more desirable. Small samples are useful especially when a large sample of the same taxonomic unit collected in another area is available for comparison. When estimates based on a sample of fewer than 15 specimens are used, it is wise to increase the odds against the systematist and thereby increase somewhat the probabilities of occurrence of specimens beyond the observed limits of the sample. This will be discussed later in the section dealing with small samples.

Bias

The personal opinions and desires of the individual systematist may enter into the selection of the sample in such a way as to make any conclusions based on it unreliable and incorrect. If any essential variation of the population cannot be inferred from the sample, owing to selection of elements out of or into it by the worker, the sample is biased and should not be used. Bias is often difficult to detect and is a constant hazard in both qualitative and quantitative methods. Random sampling is probably the only practical way to eliminate bias of this type, but in systematics it is impossible to get a random sample since the samples and specimens are withdrawn from the population permanently so that any sample taken subsequently to the first specimen is automatically biased in that chance selection cannot operate. The collector cannot change to another geographical locality in the hope of obtaining a previously unsampled and therefore unbiased sample from the same population because samples from each locality differ to a greater or less degree, as will be shown later in this paper. Also, in biological sciences it is impossible to get a random sample of a group unless samples are taken and replaced each season. One can never be sure that all the variations of a given population are in existence during any one season and therefore available for random sampling. The extremes may die at any time or be replaced by more extreme variations. Systematic work is therefore confined to "random" samples that are not statistically random, taken from the population in a given area at a given time. However, this fact does not invalidate the use of statistics on biological data as it has been shown that most samples approach closely the normal curve which is based statistically on random sampling.

The collector should take any and all specimens irrespective of their influence on his preconceived conclusions. If it is necessary to eliminate specimens in the laboratory (subsampling), it should be done entirely by various means of chance selection. No extreme variants should be arbitrarily excluded from the sample unless they can be shown to represent malformed specimens, or other species or subspecies. The extreme variants are those in which the systematist is especially interested as they influence his conclusions more than do those specimens between the ex-

tremes, because of their effect on the mean value and on the observed and calculated ranges of variation.

SELECTION OF CHARACTERS

When the systematist has before him a number of samples that appear to be homogeneous, adequate, and unbiased, the next logical step in the analysis is to select the characters to be measured. The following two requirements should be satisfied when the selection is made.

Evolutionary Significance

The systematist should attempt to select characters whose evolutionary trends appear to be governed by more fundamental genetic changes. A complex structure, which is almost certainly governed by the behavior of numerous genetic elements, is more desirable than one governed by a single gene and is more apt to indicate profound genetic change. Single gene differences are important, however, in intraspecific studies. Numerous divergent characteristics should be selected and measured and the more divergent of these used in conjunction with biological differences in establishing the status of the population. If the status of the population is fixed through the study of characters that are not the most divergent available, the conclusions are subject to question, as the most divergent genetic change and therefore possibly the isolating mechanism (or its indicator) between the populations may have been ignored.

Character selection should not be limited to the morphological features of the organism, although they are the ones most generally available for use. More fundamental differences between populations are often found in their biology, ecology, physiology, psychology, etc., and characters based on these may be employed in the same way as long as they are measurable. Where these features are found to be more divergent than the morphological differences and are not due to nongenetic conditioning, they should be used in establishing the status of populations and the morphological differences used as supporting characters.

Demarcation

In measuring organisms, accurate demarcation of the characters is one of the most important features. The most accurate and reliable data are obtained from measurements of structures that have definite limitations or "landmarks" from which the measurements can begin and end. Indistinct landmarks should be avoided as should those that do not persist throughout the group being measured.

MEASUREMENT

Systematic studies of organisms involve primarily estimates concerning the probable relationships between or among various samples of the populations. These estimates are usually made after the similarities and differences among the organisms are studied, the similarities being too often neglected. All systematists are aware of the existence of variation in biological material and therefore employ various types of measurements in attempting to express this variation and the differences among allied samples. Even though this variation is often easily visible, it may be deceptive and frequently leads the uncritical worker into numerous avoidable pitfalls. Differences in samples that appear obvious to one worker are often shown to be of little value when more critical studies are made on the same samples or when additional samples are studied. Many of the difficulties in modern systematics are encountered in the measurements and the interpretations of the differences found. This is especially true where these differences are expressed in descriptive words and where no definite limits of variation have been recorded in the sample or calculated for the population. The most successful method of avoiding these common errors made by the qualitative systematist is to measure the variation in terms of definite units rather than merely presuming the limits of variation.

There are essentially two problems in measuring. The first is due primarily to the frequent failure of qualitative examination to reveal the most reliable characters by which two distinct groups of organisms may be separated. The second is due to the success of qualitative examination in locating this character but its inability to express and determine the reliability of these observed differences. The discovery of the more divergent

characters that are not obvious on preliminary examination involves the measuring and evaluation of all characters that show divergence between one sample and another and is therefore one of the most difficult problems encountered. Recently evolved groups are especially well represented in the enigmas of this sort, as are those groups in which subspeciation is a common occurrence. The proper evaluation of observed qualitative differences depends entirely upon the individual worker and his desire to determine the trend of evolution more accurately and to establish uniformity in nomenclature by applying refinements of measurement and analysis.

Once the sample is ready and properly sexed, the characters to be measured are selected, and the scale of measurement and landmarks determined, the systematist is confronted with the problem of how to go about the all important task of measuring with the greatest efficiency. From this standpoint, there are two features to be kept in mind: first, the method of handling the specimens, and second, the method of handling the raw numbers.

In respect to handling the specimens, several steps should be carried out in order, as follows:

Each specimen should be distinctly marked (numbered or lettered) so that it can be restudied at any time, thus making it possible to check any suspicious measurements.

Only a limited number of measurements should be made on any one specimen at a given time. By this it is meant that it is better to make one measurement throughout the entire sample and in all samples at one time in order to stabilize the measurement for all individuals and to avoid possible changes in personal opinion regarding landmarks. The method of measuring and a description of the landmarks should be carefully recorded.

Types of Measurements

In systematics we are dealing with two distinct types of measurements. The first type is the count of structures that are present in definite numbers and not in fractions, such as four antennal segments. This type also includes frequencies in which the actual counts vary in whole numbers, only the average of which may be a fraction. For example, individuals of a species may deposit from 16 to 21 eggs, the average number for the species being 18.5 eggs, which obviously cannot exist as a fraction. For convenience these are called discontinuous measurements, since no fractions actually exist that connect the whole numbers.

The second and more frequently used type is the measurement of dimensions of structures. Such measurements as volumes, angles, and time have only limited systematic application, and although areas often appear to be useful diagnostic characters they have the serious disadvantage of being difficult to calculate accurately, especially if the shape is irregular. The linear measurement of a structure is the most accurate and useful one to the systematist. This type of numerical data, the measurement of a structure to the nearest unit on the scale used, is sometimes called continuous measurement, as fractions of units of measure are involved. For example, while a structure measured on a millimeter scale may be said to be 3 mm. in length, this means only that it is nearer to 3 mm. than to 2 or to 4 (2.5 mm. to 3.4 mm.), but if the same structure were measured on a finer scale it might be found to be 3.3768 mm.

Requirements of Good Measurements

The following six mathematical requirements should be considered in measuring characters:

Unit of Measure: Inasmuch as most systematic descriptions use the metric system it is advisable to continue this. In the case of virtually all insects or small organisms it is advisable to use millimeters or fractions thereof when additional division is necessary. Accuracy of measurements in millimeters or less can be obtained through the use of eyepiece micrometers or screw micrometer eyepieces.

Standardization: The scale to be used and the refinement necessary to give the desired results should be selected at the beginning of the analysis and carried without change throughout the problem. If two or more characters having different ranges of variation (1 mm., 10 mm., 20 mm.) are to be compared in the analysis, it is advisable to measure all of them by the minimum scale

in order to facilitate interpretation of the results.

ACCURACY: Since it is possible to measure a structure not only too finely but also not finely enough, some idea should be had as to the most desirable practical limits. Large-scale measurements fail to break the character into enough parts to indicate smaller variational trends, whereas very fine measurements not only introduce an increased possibility of mechanical error but also give a greater number of figures, which increases the difficulties in handling them statistically. After a certain point is reached, additional refinement fails to produce added accuracy because of increased difficulty in making the measurement. A practical rule for determining the most useful scale has been advanced by Simpson and Roe (1939, pp. 28, 29) as follows: Knowing the characters to be measured, the systematist should select the smallest of these and then adopt a unit of measure that is contained within the range of variation of this character (largest minus smallest) at least 16 and up to 24 times. A convenient average is 20 times, that is, about one-twentieth of the range of the minimum character which has probably the smallest range of all the characters. In application, this means that a character with a range of 1 mm. should be measured with a scale divided into one-twentieth- (.05) mm. units; a range of 2 mm. into one-tenth- (.1) mm. units; 3 mm. into three-twentieths- (.15) mm. units; 10 mm. into one-half- (.5) mm. units; etc. Measurements in these units will, in a majority of cases, provide a maximum of useful zoological information and will suffice for statistical purposes. If an adequate series (15 to 100 specimens) is not available to determine the probable range of the minimum character, a useful rule is to record to three digits.

SIGNIFICANCE: Care should be taken in the selection of the characters to be measured to avoid those that might be subject to growth or individual distortion. The structure to be used should be reasonably well related to any other with which it might be compared or used as a ratio.

NUMBER OF MEASUREMENTS: The number of measurements necessary to portray adequately differences in various structures depends largely upon the character being measured. Some differences in shape of structures may require a large number of measurements on the same structure to give the proportions desired; for others these proportions may be obtained by a simple length and width measurement. Care should be taken to obtain enough measurements to express the differences between the characters. It should be kept in mind that it is better to make too many measurements than too few.

BIAS: The factor of bias, whether intentional or unintentional, should always be taken into consideration and avoided as much as possible. It is often desirable to have another person check at least a few of the measurements in order to test for unintentional bias. Another method by which bias can be eliminated in the individual is to avoid looking at locality labels (in the case of mixed samples) or attempting to draw conclusions before all the measurements have been made. This will eliminate favoritism in the measurements on the part of the systematist towards any desired goal. Remeasurement after a period of time has elapsed will often reveal unintentional favoritism in the first measurements. Cross checking is not applicable in most biological studies but does eliminate bias and inaccuracy in statistical application, as will be shown.

MECHANICS OF RECORDING

When large series of numbers are available for statistical treatment it is very desirable to have them arranged in convenient form. Not only will this speed up statistical operations but it will reduce the ever present possibility of making mistakes in transcription.

The numbers obtained can best be handled in the following manner:

Separate data sheets, divided into the proper number of columns (depending on the number of measurements to be made) should be kept independently for the males and females of each sample. Each sheet should contain the complete data for the sample, that is, locality, date, sex, and specific or subspecific name if possible. Also, the scale used in making the measurements should be noted. Each specimen should be listed according to its number or letter and all of its measurements given opposite this number. Proper headings should accompany the measurements. These

should be arranged so that all measurements on a single structure are grouped together.

PRELIMINARY TREATMENT OF RAW DATA

Having obtained measurements of the desired characters, the systematist is confronted with an additional problem. Should the raw data be used directly, or would it be more advantageous to find the proportional comparisons of one structure to another before treating them statistically? Special attention should be given this consideration as this is a point in the procedure where a subjective mistake can invalidate the entire analysis.

Before an evaluation of continuous measurement differences is made, it is well to point out that discontinuous measurements (number of segments, etc.), as they are not influenced by size variations, may be used directly rather than in ratios. It is also possible that even though a character is continuous (length, breadth, etc.) it too could be used directly if there were no great size differences between individuals within each sample. However, in most cold-blooded animals there is usually a rather extensive size range in any sample, presumably because of direct environmental effects on the individuals, and it is therefore advisable to eliminate this variable by using an expression of proportion (ratio) whenever possible.

If it be assumed that the various selected characters on the samples have been measured and that it is now desirable to determine the most divergent characters, it is convenient to place the raw data sheets of the samples side by side in order to make the selections from them. For illustrative purposes, actual sample data on two allied beetle populations are presented in table 1.

If ratio analysis be disregarded for the moment, several important differences between these two samples can be observed in the raw numbers. E.L. in sample B is uniformly much larger than E.L. in sample A, and such data

TABLE 1

MEASUREMENTS ON 15 SPECIMENS IN EACH OF TWO ALLIED BEETLE SAMPLES
(Unit of measurement: 1/20 (.05) mm.)

Specimen	Sample A						Sample B					
	T.L.	T.W.	F.M.	H.M.	E.L.	E.W.	T.L.	T.W.	F.M.	H.M.	E.L.	E.W.
1	84	110	88	74	216	142	90	98	84	54	252	137
2	84	101	84	66	200	141	98	110	92	60	265	150
3	84	106	89	74	210	145	98	106	91	63	263	145
4	85	107	88	72	203	141	99	107	92	60	260	146
5	85	109	91	76	209	148	93	105	90	60	250	140
6	88	115	92	76	212	146	100	106	90	59	262	146
7	86	108	90	72	212	143	102	113	92	62	260	148
8	87	107	90	70	209	144	98	106	91	62	261	147
9	85	111	91	70	198	141	103	109	94	64	276	150
10	78	101	84	68	194	131	103	108	91	64	262	150
11	78	102	84	70	206	143	97	110	92	64	268	152
12	90	114	92	74	211	150	103	115	98	64	269	151
13	86	110	89	72	208	144	92	103	86	58	246	137
14	86	110	92	72	214	142	96	106	92	64	257	148
15	92	112	93	75	216	146	100	106	90	61	259	145

T.L., Thoracic length along mid line
T.W., Thoracic width at widest point
F.M., Width of front margin of thorax
H.M., Width of hind margin of thorax
E.L., Elytral length
E.W., Elytral width at widest point

might therefore be used as raw numbers. E.W., T.W., and F.M. appear to be very nearly alike in both samples and by themselves are not so useful as E.L. in the analysis. T.L. in sample B averages somewhat larger than T.L. in sample A, but their distributions overlap as both have measurements in the low 90's. H.M. in sample B is uniformly smaller than H.M. in sample A, but the difference between them is small. Thus there are only two differences (E.L. and H.M.) in the raw data that appear to separate all specimens in these two samples.

A further examination along different lines discloses valid ratio differences between these two samples as follows: A study of the measurements of each individual specimen in sample A shows that F.M. is equal to or greater than T.L. for each specimen (84-88, 84-84, 90-92, etc.), while in sample B, F.M. is uniformly smaller than T.L. (90-84, 98-92, 103-98, etc.). This means that there is actually a difference between these two samples in the proportion of F.M. to T.L. Similarly, T.L. in sample A is smaller than T.L. in sample B, while H.M. in sample A is larger than H.M. in sample B. Thus there is another difference in proportion between these two samples on the basis of T.L. and H.M. Since E.L. in sample B is much larger than E.L. in sample A, it would be advantageous if some measure could be found in sample A that was larger than the same measure in sample B so as to accentuate this difference by expressing them as ratios. This exists in H.M., so that the combination of H.M. and E.L. as proportions gives a very large and reliable difference between the two samples—sample A having a short E.L. and a wide H.M., whereas sample B has a long E.L. and a narrow H.M. Similarly, but with differences of less magnitude, E.L. could be combined with all other characters as proportions that would express varying degrees of difference in shape between the two samples.

Ratios

From this discussion, the value of ratios is obvious: the use of them greatly increases the number of characters available, thus giving a more detailed description of the samples by expressing proportion. Ratios are especially valuable in insect studies as the largest specimens generally have the largest structures (size), even though the proportions of these to other structures may be the same as in smaller specimens. Inasmuch as an attempt is being made to select characters which show as little variation as possible to represent the sample, it is often advantageous to take the ratios between two characters in order to rule out variation that accompanies uncontrollable size differences. If, for instance, in a given species there is one specimen 24 mm. long by 8 mm. wide and another 18 mm. long by 6 mm. wide, the variation in length is from 18 to 24 mm., and in width from 6 to 8 mm. The proportional relationship of the width to the length in these two specimens, 24/8 and 18/6, shows, however, that in both specimens the length is three times greater than the width. Thus, actual size is ruled out, and the shape character is expressed as a ratio of one dimension to another.

Ratios are in widespread use in zoology as they express characters that are of fundamental importance. This is especially true in systematic studies of groups where the chief differences are in shape. There are, however, a number of undesirable features in the use of ratios that should be kept in mind. These do not detract from their usefulness but are concerned with the interpretation of the results based on them. One of their most confusing characteristics is that they bear no resemblance to the original figures that can be detected from simple inspection of the ratio figure. Two or more specimens having the same ratio value may differ in the size of the original measurements. For example, a length recorded as 1.0 mm. is known to be somewhere between .95 and 1.04 mm. on the continuous millimeter scale, a simple and obvious relationship which is not true of a ratio recorded as 1.0. The actual measurements involved in a 1.0 ratio might be 1.0/1.0 mm., 2.5/2.5 mm., 4.8/4.8 mm., etc., each of these measurements being individually variable as noted above. The real ratios of lengths recorded as 1.0 each might be anywhere between .91 (.95/1.04 mm.) and 1.09 (1.04/.95 mm.). Ratios are sometimes more variable than the dimensions on which they are based. Thus, if the lengths of homologous structures in a given sample vary from 0.9 to 1.1 mm., and the widths also from 0.9 to 1.1

mm., the possible length-width ratios vary from 0.8 (0.9/1.1 mm.) to 1.2 (1.1/0.9 mm.).

Ratios may be expressed numerically in a number of different ways: as unreduced ratios of the actual measurements (5 mm./10 mm.); as fractions (½); as quotients (0.5); as percentages (50 per cent); and as quotients multiplied by a constant (see Simpson and Roe, 1939, p. 13). In descriptive work it is best to use unreduced ratios where exact proportions are desirable; when an expression of the approximate ratio is desired, fractions can be used; when statistical treatment is to be done, quotients are of most use.

Since the ratio between E.L. and H.M. in table 1 appears to be the proportion expressing the greatest divergence between these two samples, these will be developed as an illustration. The first problem confronting the systematist in the development of ratios is that of deciding which character should serve as the dividend (numerator) and which as the divisor (denominator). In systematics it is most practical and elucidating to express the larger of the two structures as being proportionally greater than the smaller. In the present example, E.L. would therefore be divided by H.M. The figures obtained will give the proportion of H.M. to E.L. so that in descriptive work this character will read that the elytra are so many times longer than the width of the hind thoracic margin. Table 2 illustrates the actual ratios and the difference between these ratios in the two samples.

In sample A, specimen 1, the elytra are 2.92 times longer than the width of the hind thoracic margin, whereas in sample B, specimen 1, the same structures have a proportional relation of 4.67. The great advantage of this ratio over the raw number can be seen by referring to the actual measurements. The minimum value of E.L. in sample B is 246 and the maximum value of E.L. in sample A is 216, giving a difference of 30 points between these two samples for E.L. The minimum ratio value of E.L./H.M. in sample B is 4.02 and the maximum ratio value in sample A is 3.03. This gives a comparatively greater difference than that obtained by using the raw data directly.

TABLE 2

DEVELOPMENT OF RATIO VALUES FOR E.L. AND H.M. IN TWO SAMPLES OF THE BEETLE GENUS *Omus* FROM TABLE 1

Specimen	Sample A			Sample B		
	E.L.	H.M.	E.L./H.M.	E.L.	H.M.	E.L./H.M.
1	216	74	2.92	252	54	4.67
2	200	66	3.03	265	60	4.42
3	210	74	2.84	263	63	4.17
4	203	72	2.82	260	60	4.33
5	209	76	2.75	250	60	4.17
6	212	76	2.79	262	59	4.44
7	212	72	2.94	260	62	4.19
8	209	70	2.99	261	62	4.21
9	198	70	2.83	276	64	4.31
10	194	68	2.85	262	64	4.09
11	206	70	2.94	268	64	4.19
12	211	74	2.85	269	64	4.20
13	208	72	2.89	246	58	4.24
14	214	72	2.97	257	64	4.02
15	216	75	2.88	259	61	4.25

FREQUENCY DISTRIBUTION

Thus far systematic analysis has proceeded on the assumption that there are various divergent characters that can be used to distinguish samples despite individual variation within each sample. Since the pattern and

extent of this individual variation determine the reliability and stability of the characters, it is necessary for the systematist to make measurements in an attempt to establish this pattern for the characters being used. The quantitative data obtained by measuring comprise a series of numbers which are largely without meaning or significance by themselves until they have been arranged and classified in some orderly way. The next task that confronts the systematist, then, is the organization of his numerical data by grouping the measurements into classes. This may be done by means of a frequency distribution table which gives a summary of the variation in the character. A frequency is the number of observations or measurements that fall into any one defined class, and a frequency table is a list of these classes with the frequencies in each. Such tables form the basis of almost all important numerical observations in zoology and are a necessity in adequate quantitative systematic analyses.

TABULATION

The procedures for tabulating the frequency distribution of a character come under four main heads, which are given in the following order of application.

The determination of the range of variation, that is, the interval between the largest and the smallest measurements of the character. This may be easily obtained by subtracting the smallest measurement from the largest and adding one unit of measurement.

The division of this range into convenient steps or classes for tabulating the frequencies. The size of the class, or the class-interval, depends largely upon the range of the character. As a general rule, the class-interval should be about one-twentieth of the total range of the character in all samples. If, for example, the width of a structure varies from 10 to 30 units, this range of 21 units could be easily handled in 21 classes with a class-interval of one unit. However, in cases where a character has a range of 100 or even 200 units, it is more convenient to make the intervals five or 10 units apiece in order to have fewer classes. When class-intervals of more than one unit are used, the midpoint of the interval is used for tabulation and calculation, but the upper and lower limits of these classes should also be given in frequency distribution tables to avoid any overlapping of classes. In systematic analyses where several samples are to be compared, it is very important that the same class-interval and class limits be used for the same character in each sample in order to facilitate comparisons and to make possible the derivation of rough estimates of variation from these tabulations.

The construction of a frequency distribution graph is the plotting of the separate measurements or counts within their proper classes. Graph paper should be used; the midpoints or limits of the class-intervals (the classes) of the measurement for the particular character are placed along the bottom, reading from left to right, and the frequency scale is placed on the side, reading vertically from the bottom up. The separate items (from the original data sheets) are then entered as checks or dots, filling in the squares corresponding to the vertical frequency scale.

The following example, using the data for E.L./H.M. of sample A in table 2, will illustrate the processes. Minimum ratio is 2.75, maximum is 3.03. Unit is .01. Range is 3.03 minus 2.75 plus .01, or 29 units. Fifteen classes are nearer the optimum number of 20 than 29

TABLE 3

FREQUENCY DISTRIBUTION GRAPH FOR E.L./H.M., SAMPLE A, TABLE 2

Freq.															
3															
2					x	x				x					
1	x		x	x	x	x	x	x	x	x		x	x		x
Class-intervals	2.75–2.76	2.77–2.78	2.79–2.80	2.81–2.82	2.83–2.84	2.85–2.86	2.87–2.88	2.89–2.90	2.91–2.92	2.93–2.94	2.95–2.96	2.97–2.98	2.99–3.00	3.01–3.02	3.03–

classes. Range of 2.75 to 3.03 is therefore divided into 15 classes with class-intervals of .02 each. The ratios from table 2 are shown arranged in a frequency distribution graph in table 3. From this table it can be seen that there is one specimen with a ratio for E.L./H.M. of 2.75–2.76 (midpoint is 2.755), none with 2.77–2.78, one with 2.79–2.80, etc.

Frequency distribution graphs plotted in this manner are advantageous in that they give the systematist a graphic view of approximately how closely the polygon based on the sample data will fit a normal curve (to be discussed later).

The construction of a frequency table. It is unnecessary and in fact unwise for the systematist to illustrate his results only by the use of rough or smoothed frequency polygons, as these are inadequate for most scientific purposes since they require additional statistical treatment for systematic interpretation. The results of the frequency distribution graph can be most adequately portrayed by the use of the actual figures in a frequency table which enables subsequent workers to treat the raw data in a different manner if this seems desirable. The tabulation of the frequencies of each class can be done directly from the frequency graph. Using the same data as in table 3, the frequency distribution table is constructed as shown in table 4.

Before statistical analysis can begin the systematist must obtain measures of the central tendencies and of the extent of variation in the frequency distributions of the samples. The measures of central tendency needed are the arithmetic mean, the median, and the mode. If the frequency distribution is symmetrical ("normal"), all three of these measures of central tendency will fall on the same point on the range scale. If the distribution is skewed, they will fall at different points, the median usually between the mode and the mean and nearer the mean. These measures are of importance when dealing with skewed curves, but the mean is especially important throughout virtually all the remaining procedure.

Arithmetic Mean

The arithmetic mean is the most commonly used average and will be referred to hereafter simply as the mean (M). In calculating the mean an attempt is made to get a single number that will represent that point on the frequency distribution where the average individual is to be found. It may be defined simply as the sum of the separate measurements in a series, divided by the number of individuals in that series. When the data have been arranged in a frequency table, it is much more convenient to use the table for calculating the mean than to refer back to the original data. The procedure is to prepare a work sheet with headings as shown in table 5. Then in the first column record the class limits, in the second the midpoints of the classes, and in the third the corresponding frequencies for each class as was done in the construction of a frequency distribution table. The sum of these frequencies is the total number of individuals in the sample, hereafter referred to as N. The figures in the fourth column are the products of the frequencies times the values (midpoints) of the respective classes. The sum of these products divided by N is the arithmetic mean, usually calculated to one more decimal place than are the original data.

Thus, in table 5 the class limits of the 15 classes of .02 each are in column 1, the midpoints of these classes in column 2, the fre-

TABLE 4

Frequency Distribution Table for E.L./H.M., Sample A, Table 2, Taken from Frequency Distribution Graph, Table 3

Class Limits	Midpoints	Frequencies
2.75–2.76	2.755	1
2.77–2.78	2.775	0
2.79–2.80	2.795	1
2.81–2.82	2.815	1
2.83–2.84	2.835	2
2.85–2.86	2.855	2
2.87–2.88	2.875	1
2.89–2.90	2.895	1
2.91–2.92	2.915	1
2.93–2.94	2.935	2
2.95–2.96	2.955	0
2.97–2.98	2.975	1
2.99–3.00	2.995	1
3.01–3.02	3.015	0
3.03–3.04	3.035	1

quency of occurrence of each ratio in column 3, and the products of the frequencies times each respective class midpoint in column 4. The addition of the figures in column 3 gives the total number of frequencies which checks with the number of individuals in the sample (15). Then, by dividing the sum of column 4 (43.305) by N (15), a mean value of 2.887 is obtained which indicates that the average individual is located on the range scale at 2.887; or, expressed in another way, the average individual of this sample has an E.L./H.M. ratio of 2.887 (the elytra are on the average magnitude. In other words the median is the value of that member of a series that has as many individuals larger than it is as it has smaller. In a series composed of an odd number of individuals, the middle individual can be found by dividing N plus 1 by 2, and the median is the value of this individual. When the series is composed of an even number of individuals, the median is the value halfway between the values of the two middle individuals. For the data given in table 5, the middle individual is 15 plus 1 (16) divided by 2, or the eighth. By adding the frequencies in

TABLE 5

DERIVATION OF THE MEAN FROM THE FREQUENCY OF TABLE 4

Class Limits	Midpoints	Frequencies	Frequencies × Values	
2.75–2.76	2.755	1	2.755	
2.77–2.78	2.775			
2.79–2.80	2.795	1	2.795	
2.81–2.82	2.815	1	2.815	
2.83–2.84	2.835	2	5.670	
2.85–2.86	2.855	2	5.710	
2.87–2.88	2.875	1	2.875	Median = 2.875
2.89–2.90	2.895	1	2.895	Mean = 2.887
2.91–2.92	2.915	1	2.915	
2.93–2.94	2.935	2	5.870	
2.95–2.96	2.955			
2.97–2.98	2.975	1	2.975	
2.99–3.00	2.995	1	2.995	
3.01–3.02	3.015			
3.03–3.04	3.035	1	3.035	
		N = 15	43.305	

for this group of 15 individuals 2.887 times as long as the hind margin of the thorax is wide). Going back first to the original ratios which were recorded to the nearest one-hundredth, 2.887 becomes 2.89, and then referring to the establishment of the classes, based on class-intervals of .02 each, in which all individuals with ratios of 2.89 and 2.90 were put in the class labeled 2.895, we find that this average individual falls in the 2.895 class.

MEDIAN

This is the second measure of central tendency and may be defined as the value (the measurement, count, ratio, etc.) of the middle individual in a series arranged in order of column 3 from either end, the eighth individual is found to be in the 2.875 class. Thus the median for this sample is 2.875. This method of determining the median is adequate for most statistical analysis. (For a more detailed discussion, see Simpson and Roe, 1939, p. 94.)

MODE

This measure of central tendency is defined as the value of the range scale at which the frequency distribution graph is highest, or, in other words, the value of the class containing the greatest number of individuals. This measure can be roughly determined by inspection of the frequency graph or table. Thus in table 5 the mode could be 2.835,

2.855, or 2.935, the values of the classes with the highest frequencies, there being two individuals in each of these three classes. Small samples or the use of too many classes is apt to give erratic distributions, and it may be difficult in these instances to select the mode.

Standard Deviation

From the preceding discussion it has been seen that it is possible to establish definite points along the range scale by computing the measures of central tendency of the frequency distribution of the sample. In systematic work it is far more important, however, to know the extent of the variation in order to be able to delimit the samples accurately. The measures of central tendency are definite points on the range of variation scale, whereas range and standard deviation are measures of distance along this scale.

The most adequate and reliable measure of variation is the standard deviation, which is based on the value of the mean and the amount of variation from the mean of the individuals in the sample. Although the observed range of a sample expresses variation, it cannot be used to estimate population variation without further statistical treatment, as it represents only the actual range of the sample and gives no indication of the expected range in additional samples from the same population. By using standard deviation, the systematist is able to take into consideration the specimens he does not have at the time of the analysis, and he is therefore making his analysis of the population more inclusive and accurate.

In statistical terms, standard deviation (*S.D.*), also called sigma (σ), is defined as the square root of the sum (Σ) of the deviations (from the mean) squared, divided by N, i.e.,

$$S.D. = \sqrt{\Sigma d^2/N}$$

From the standpoint of the systematist, it represents a mathematical expression of the range of variation of the measured (or counted) characters. This serves as a basis for calculating the range of the total population and as a means of obtaining the percentage of individuals that might be expected to occur outside as well as inside the observed sample range and the calculated population range or at any point on this range. The necessity for the use of the standard deviation in systematic analyses has probably been overlooked many times owing to the failure of the systematist to realize that he is dealing with only a sample of the total population, that the variation of this sample is, almost always, less than that of the total population, and that the range of variation of a sample usually falls within that of the total population.

The following method of obtaining the standard deviation applies to samples containing 15 or more individuals. The methods of obtaining an estimate of the standard deviation for samples of two to 14 specimens and for single specimens will be discussed later.

Inasmuch as a number of the quantities obtained in the derivation of the mean are used in calculating the standard deviation, a short cut is possible in that both of these important measures can be derived on the same work sheet. Also, since the systematist may not be versed in mathematics, it is desirable to have a check on all figures to eliminate mathematical errors as much as possible. For this purpose he may use the modified Tryon-Searle "Form for mean and standard deviation," on which these two measures can be derived and automatically checked for accuracy. The following are the directions for the modified Tryon-Searle form, and the example used as an illustration in figure 1 is taken from an *Omus* sample composed of 96 specimens whose elytral width varied from 105 to 135 units of measurement.

Class limits	Scale	Tally	f	d	fd	fd^2	d′	fd′	fd'^2	Pop.
				+11			+12			
				+10			+11			
138-139	138.5			+ 9			+10			+3 S.D.
136-137	136.5			+ 8			+ 9			
134-135	134.5	I	1	+ 7	+ 7	49	+ 8	+ 8	64	
132-133	132.5			+ 6			+ 7			+2 S.D.
130-131	130.5	IIII	4	+ 5	+20	100	+ 6	+ 24	144	
128-129	128.5	卌	5	+ 4	+20	80	+ 5	+ 25	125	
126-127	126.5	卌 卌 卌	15	+ 3	+45	135	+ 4	+ 60	240	+1 S.D.
124-125	124.5	卌 卌 IIII	14	+ 2	+28	56	+ 3	+ 42	126	
122-123	122.5	卌 卌 卌	15	+ 1	+15	15	+ 2	+ 30	60	MEAN
A 120-121	120.5	卌 卌 卌 III	18		Σ=+135		+ 1	+ 18	18	MODE
A′ 118-119	118.5	卌 I	6	- 1	- 6	6		Σ=+207		
116-117	116.5	IIII	4	- 2	- 8	16	- 1	- 4	4	-1 S.D.
114-115	114.5	卌	5	- 3	-15	45	- 2	- 10	20	
112-113	112.5	IIII	4	- 4	-16	64	- 3	- 12	36	
110-111	110.5	III	3	- 5	-15	75	- 4	- 12	48	-2 S.D.
108-109	108.5			- 6			- 5			
106-107	106.5	I	1	- 7	- 7	49	- 6	- 6	36	
104-105	104.5	I	1	- 8	- 8	64	- 7	- 7	49	-3 S.D.
				- 9			- 8			
				-10			- 9			
				-11			-10			
				-12			-11			
			N=96		Σ=- 75	754		Σ=- 51	970	
I=2					Σd=+60	Σd^2		Σd′=+156	$\Sigma d'^2$	

1. Σd + N = 60+96 = 156 = Σd′ ✓
2. $\Sigma d^2+\Sigma d+\Sigma d'$ = 754+60+156 = 970 = $\Sigma d'^2$ ✓

3. c = Σd/N	= + .625	7. c′ = Σd′/N	= + 1.625
4. Ic	= +1.25	8. Ic′	= +3.25
5. A	= 120.5	9. A′	= 118.5
6. Ic+A	= 121.75 = Mean	10. Ic′+A′	= 121.75 = Mean ✓
11. $\Sigma d^2/N$	= 7.8542	14. $\Sigma d'^2/N$	= 10.1042
12. c^2	= .3906	15. c'^2	= 2.6406
13. $(\Sigma d^2/N)-c^2$	= 7.4636 = σ'^2	16. $(\Sigma d'^2/N)-c'^2$	= 7.4636 = σ'^2 ✓
		19. I^2	= 4
17. $\sigma'=\sqrt{\sigma'^2}$	= 2.73	20. $I^2\sigma'^2$	= 29.8544
18. $I\sigma'$	= 5.46 = σ = S.D.	21. $I\sigma'=\sqrt{I^2\sigma'^2}$	= 5.46 = σ = S.D. ✓

FIG. 1. Modified Tryon and Searle form for mean and standard deviation.

1. Determine the range scale.

A. Range: Subtract the smallest measurement from the largest and add one unit of measurement to find the range. Thus, $135-105+1=31$, so the range in this example is 31 units long.

B. Class-interval (size of classes): In this form the maximum number of classes that can be used is 24, so the range must be divided into 24 or fewer classes. In order to determine the size of the classes divide the range by 24 and raise the quotient to the next highest whole number. Since the range is 31 in this example, the class-interval would be 31/24 units expressed as the next highest whole number, or 2.

C. Number of classes: Divide the range by the class-interval and raise the quotient to the next highest whole number to find the number of classes required for the given frequency distribution. Thus, with a range of 31 and a class-interval of 2, this would be 31/2 expressed as the next highest whole number, or 16. The data extending over a range of 31 units of measurement can now be placed in 16 classes of 2 units each which will fit on the chart. The systematist should be sure to comprehend both ends of these class limits when entering the midpoint in order to avoid letting any classes overlap, and it may be found that entering the class limits in column 1 will be of great help. For example, 134–135 are the class limits of the class whose midpoint is 134.5, 132–133 for 132.5, etc.

Place the value of the class-interval in the box "I" at the bottom of the "Class Limits" column; in this example, 2.

2. Enter the midpoints of the classes in the "Scale" column.

In entering the classes, place the one with the largest value at the top of the sheet and center the classes approximately with reference to A and A' so that there will be about as many classes above A in the column as below. This facilitates later comparisons. In the example illustrated in figure 1 with 16 classes, the eighth class, 120.5, is placed opposite A, 118.5 opposite A', with seven classes above and seven classes below.

3. Determine the distribution of frequencies ("Tally" and "f" columns).

A. Distribute the values of the individual measurements in the "Tally" column. It is always better to take these from the original data sheets rather than from a frequency distribution graph in order to avoid errors in transcription. If individual entries are used in this column, it will be found that they will form a rough frequency distribution graph, which may help to visualize the arrangement of the specimens.

B. Add these items for each class and enter them in the "f" or frequency column. This corresponds to a frequency distribution table.

C. Add the frequencies in the "f" column and enter this sum in the box at the bottom, N = the total number of specimens or observations in the sample; in this case, $N=96$.

D. At this point the median and the mode can be determined as previously outlined, if these measures are desired. Median of the example is the value of the 48.5th individual which is in the 122.5 class; mode is the value of the class having the largest frequency which is the 120.5 class.

4. Compute the mean (M).

A. Multiply each f by its corresponding d to obtain the fd values. Thus, $7\times1=7$, $5\times4=20$, etc. Get the sums (Σ) of the positive and negative fd's separately, entering the totals in the spaces designated "$\Sigma+$" and "$\Sigma-$," respectively, in the "fd" column; thus, $+135$ and -75. Enter the algebraic sum of these two positive and negative totals in the "Σd" square at the bottom of the column; thus, $+135$ plus $-75=+60$.

B. Multiply each f by its corresponding d' value and enter in the "fd'" column. Thus, $8\times1=8$, $6\times4=24$, etc. Summate these as in step 4A to get $\Sigma d'$. Thus, $+207$ plus $-51=+156$.

C. Multiply each fd and fd' entry by its corresponding d and d' value and enter these products in the "fd^2" and "fd'^2" columns, respectively. Summate these products to obtain Σd^2 and $\Sigma d'^2$. Thus $\Sigma d^2=754$, and $\Sigma d'^2=970$.

D. At the bottom of the form, beginning with step 1, carry out the steps indicated through step 10. The answer obtained in step 1 should equal $\Sigma d'$, and in step 2 should equal $\Sigma d'^2$. If they do not, a mistake has been made in addition or multiplication. Check each fd, fd^2, fd', fd'^2, and the sums of these. The answer obtained in step 6 for the mean should be the same as the one obtained in step 10. If it is not, check the calculations (and the plus or minus values of c and c') in steps 3 to 10.

Follow the example through these steps: 1. Σd plus $N=60+96=156$. This checks with $\Sigma d'$ so check (X) in box. 2. Σd^2 plus Σd plus $\Sigma d'=754+60+156=970$. This checks with $\Sigma d'^2$ so check (X) in box. 3. $c=\Sigma d/N=60/96=+.625$. 4. Interval is 2. $Ic=2\times.625=1.25$. 5. A (approximate midpoint of "Scale" column) $=120.5$. 6. Ic plus $A=1.25+120.5=121.75$, which is the mean. 7 to 10. Complete in the same way as 3 to 6. As the answers to 10 and 6 agree, place check in box. The sample therefore has a mean value of measurement of 121.75 units.

5. Compute the standard deviation ($S.D.$).

The values for Σd^2 and $\Sigma d'^2$ having been obtained and checked for correctness, proceed to steps 11 to 21 which are used in the derivation and check of the standard deviation. The answer for 13 should be the same as for 16, and 18 the same as

21. This latter figure is the standard deviation (*S.D.*). If the answer obtained in step 21 does not equal that in step 18, check the calculations (and the decimal places) in steps 11 to 21. Barlow's tables or Davenport and Ekas' (1936, p. 196) are very handy for obtaining the squares and square roots in these steps. If the answer obtained in step 21 equals that of step 18, the process has been carried out correctly and the proper value for the standard deviation (5.46 in this example) obtained. This figure, as such, means very little to the systematist. However, its use as a measure of variation will be developed in the section dealing with the normal curve. The figures in light face in columns 1 and 2, as well as those in the "Pop." column, in figure 1 will be explained in the section on the Area of the Normal Curve.

Small Sample Standard Deviation

When only two to 14 specimens are available in a sample, the systematist should adopt the following method for obtaining the standard deviation to compensate for the small number of available specimens (Simpson and Roe, 1939, p. 205). The mean of the small sample is obtained in the usual way by dividing the sum of the measurements by N. Instead of using the formula for the standard deviation of

$$\sqrt{\Sigma d^2/N}$$

divide the sum of the squared deviation by $N-1$ before the square root is extracted, or

$$\sqrt{\frac{\Sigma d^2}{N-1}}$$

The deviations from the mean are found by subtracting the mean from each measurement. (Because of the small number of specimens in the sample, it is just as easy to work with ungrouped as with grouped data.) The sum of the deviations should equal zero. The deviations are squared and summated, and the sum is divided by $N-1$. The square root of the result is the adjusted standard deviation for small samples.

Single Specimen Standard Deviation

Obviously the methods thus far outlined are inapplicable to a single specimen as they depend on two or more observations. Estimates based on large samples are more accurate and therefore desirable in systematic work. However, it often happens in systematics that a species is represented in collections by a single specimen, the type, or perhaps there is only a single specimen available at the time for study. Naturally such individuals should be included in the study, and it is desirable to obtain an estimate of the probable variation in the population represented by this single specimen. It is possible that such estimates may be largely incorrect, but the odds greatly favor their being reasonably well within the actual population limits (Simpson and Roe, 1939, p. 214).

Referring to figure 6 and table 6 we see that within a range of $+2$ *S.D.* and -2 *S.D.* from the mean 95.45 observations out of every 100 will be found. Since the single specimen represents one part of the total distribution it can reasonably be assumed that, since 95 out of every 100 single specimen observations will fall within the limits of $+2$ *S.D.* and -2 *S.D.* from the population mean (based on the sample mean), this single specimen is within these limits. Naturally this assumes also that the specimen is within the limits characteristic of 95 per cent of a population that exists in nature.

In order to obtain an estimate of the population range represented by a single specimen, it is necessary to be guided by analogy with the range of the same character in other samples of the same or related populations in which there is an adequate number of specimens. In other words, it is assumed that the range of the single specimen sample is not greatly unlike that of larger allied samples. Thus, one might find that the variation of a length character in a population based on adequately represented allied samples is 12 units, with an *S.D.* of 2 units. With figure 2 as an illustration, in A the single available specimen has a value of 30 units. Using the *S.D.* of 2 units obtained from allied samples, if 30 is assumed to be the mean, -2 *S.D.* and $+2$ *S.D.* from the mean give an estimated range of from 26 to 34 units. Now, when the specimen is moved to a point $+2$ *S.D.* from the original assumed mean of 30, the range from this point (34 ± 2 *S.D.*) is found to be from 30 to 38 units, as shown in B. This indicates that if the population mean is at a point $+2$ *S.D.* above the measurement of the single specimen (30 units), or, in other words, if the single specimen fell at the lowest point

of the range limited by the mean -2 *S.D.*, 95.45 per cent of the observations would be between 30 and the maximum of 38 units if more specimens were available. If the same is done for -2 *S.D.* (as in C) the minimum expected range limit in 95 out of every 100 specimens is 22 units. Thus the expected population range as shown in D extends from 22 units to 38 units, a range of 16 units or ± 4 *S.D.* Within these limits one would expect to find at least 95 out of every 100 specimens of the population represented by a single specimen. E represents the expected mean range if additional sample specimens were found.

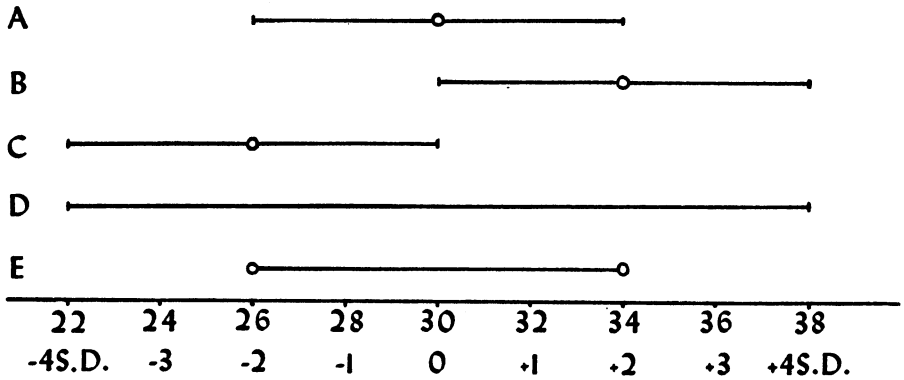


FIG. 2. The estimated population variation based on a single specimen.

In practice, the measurement or values of the single specimen plus and minus four standard deviations of the same character measurements on an allied sample gives the systematist the expected range of at least 95.45 per cent of the population represented by the single specimen. If the actual population mean happened to correspond closely with the observed single measurement, then only one specimen out of about 15,750 would be expected to fall outside these limits of plus and minus four standard deviations (see section on Probabilities). A very conservative systematist may wish to use six standard deviations instead of four, which would give the range of at least 99.73 per cent of the population. The probabilities are figured as in large samples, since the variation of a large allied sample was used to compute the variation of the single specimen sample.

NORMAL PROBABILITY CURVE

Having constructed the frequency distribution table of a particular character and ascertained the measures of central tendency (mean, median, mode) and of variation (standard deviation), the systematist is then confronted with the all important problem of the use and interpretation of these measures. However, before this can be adequately discussed, it is necessary that the systematist know something about the normal probability curve upon which the interpretation depends. Perhaps the simplest approach to an understanding of the normal curve is through a consideration of the elementary facts of probability. As used in statistics, and by adoption in biology, the probability of the occurrence of a certain measurement of a character may be defined as the expected relative frequency of occurrence of these measurements in a very large (infinite) number of observations. This expected relative frequency of occurrence, which when graphed is the normal curve, is based upon the knowledge of the conditions determining the probable chance occurrence, as in dice throwing, coin tossing, or upon purely empirical data. The normal curve is a symmetrical, bell-shaped curve (fig. 3) in which the mean, median, and mode have the same value. This type of curve has been carefully analyzed by statisticians. (For a more detailed discussion, see statistical texts such as Simpson and Roe, 1939, p. 129.)

This curve, as shown by the statistician in

his experiments on the operation of the laws of chance, serves to describe the frequency of occurrence of many variable data with a relatively high degree of accuracy. It is the widespread incidence of approximately normally distributed data in biology that accounts for the use of the normal curve in the interpretation of systematic problems. However, it must be remembered that the normal probability curve is based on chance, whereas many genetical features in animals are influenced by modifying factors and do not, therefore, necessarily conform entirely to chance pattern. Because frequency distributions based on biological samples are often abnormal to a greater or less degree, probably never completely following the normal pattern, a brief discussion of the two main types of abnormality and their systematic implications is here presented.

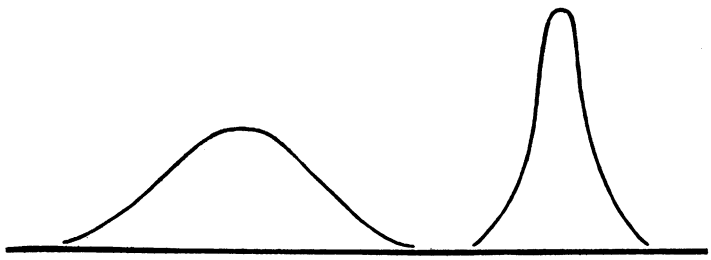


FIG. 3. The normal probability curve, showing two of its many possible forms.

SKEWNESS

In skewed curves, the highest point in the frequency curve is not in the center of the range, owing to an overabundance of individuals on one side or the other. If the overabundance is towards the larger values on the variation scale, the curve is positively skewed (fig. 4A); if towards the smaller values, it is negatively skewed (fig. 4B). Since the approximate mode is easily located by inspection of the frequency graph or table, the rough test for skewness, $Sk = (\text{mean-mode})/S.D.$, may be used in many cases. The more nearly the distribution approaches the normal curve, the closer together are the mean and the mode. When the mode lies beyond one standard deviation of the mean, the sample distribution is too greatly skewed to be used in estimating population variation by the normal curve method. (For a more detailed discussion and another formula, see Simpson and Roe, 1939, p. 143.)

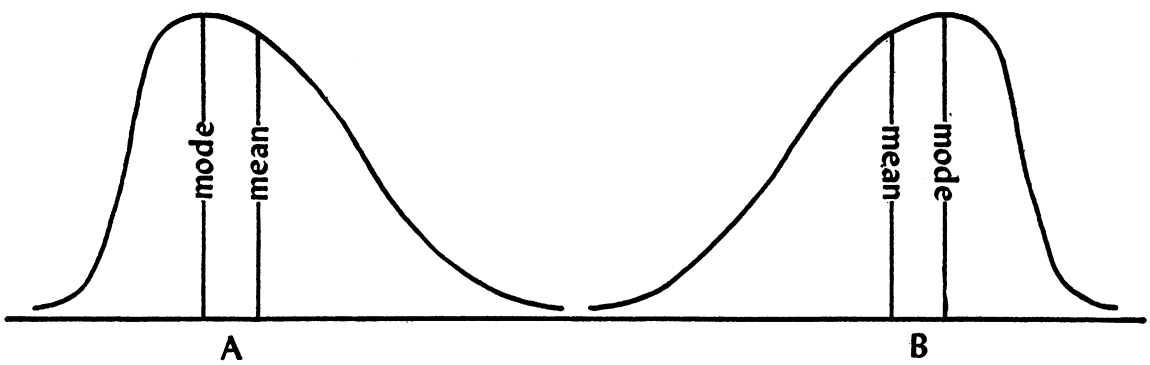


FIG. 4. Skewed curves. A. Positive. B. Negative.

KURTOSIS AND BIMODALITY

The leptokurtic and platykurtic curves (fig. 5) are symmetrical, and the values of the mean, median, and mode on the range scale coincide with those of the normal curves; but in the leptokurtic curve there are

proportionally too many individuals around the midpoint as compared to adjacent intervals, whereas in the platykurtic curve there are too few around the midpoint. The rather involved formula for testing for kurtosis may be found in Simpson and Roe, 1939 (p. 147).

The platykurtic (flat-topped) type of distribution often extends to a point where, instead of one mode near the middle of the curve, there are two separate modes. These are called bimodal curves and are often indicators of improper sampling technique. When the distribution is bimodal and this

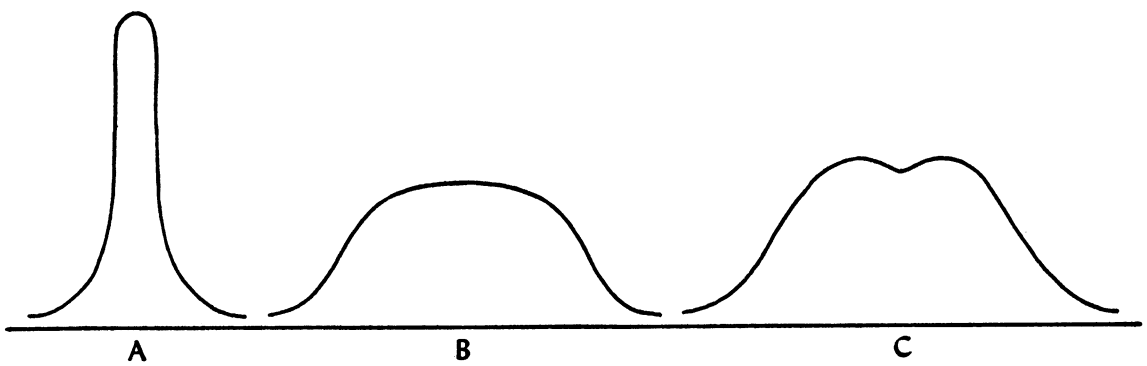


FIG. 5. Kurtotic curves. A. Leptokurtic. B. Platykurtic. C. Bimodal.

condition cannot be attributed to the small number of observations, it may be due to the heterogeneity of the sample, and the procedure for reëxamining it should be carried out. Figure 5C illustrates a bimodal curve which could be caused by sexual dimorphism in a character in which the samples were not analyzed separately. Two normal curves, one for each sex, were superimposed. Bimodality is evident when the difference between the two modes approaches twice the value of the standard deviation of the bimodal curve.

If, in the analysis of a character, the systematist finds that the frequency distribution of a sample does not approximate the normal curve, he should reëxamine it with the folowing points in mind:

SIZE OF SAMPLE: The larger the number of specimens, the more closely the sample should approximate the normal curve.

HOMOGENEITY OF SAMPLE: The presence of more than one taxonomic unit or sex in an improperly segregated sample is apt to cause skewness and bimodality.

RELIABILITY OF CHARACTER: A character that is individually unstable or of varietal status may cause skewness in a homogeneous sample.

TECHNIQUE: The size of the unit of measure, the size of the class-interval, and the method of grouping the original data into classes may give a false impression of skewness or kurtosis.

Having failed to detect the difficulty by reconsidering the above points, the systematist might find the solution in the more difficult problem of natural selection. By this we mean the causes, genetical, biological, or environmental, that might favor the survival of individuals on one end of the range. Thus, if the character were of vital importance in the survival of the organism, it might be possible for the optimum condition to be more closely approximate to the maximum lethal than to the minimum lethal requirements. This condition might conceivably exist in the case of static archaic species which, owing to their inability to change, are being eliminated by the action of unfavorable conditions primarily on one extreme of the character. Care should be exercised when making such deductions as it is difficult to ascertain an indicator of a genetically lethal condition. The skewed distributional pattern which occurs as a result of any limit in variation confined to one end of the range constitutes a difficult problem in statistical analysis. However, for all practical purposes it may be assumed that in most homogeneous samples the distribution of a measured character will approach the pattern of a normal curve and can be treated as such in the statistical analysis. The methods herein given cannot be applied to

data that do not follow closely the normal curve.

Area of the Normal Curve

For purposes of ascertaining the percentage of individuals in any given area of the normal curve, a scale based on units of standard deviation should be used in statistical systematics instead of the usual range scale expressed in units of measurement. Zero on the standard deviation scale is at the mean of the distribution, and the scale runs negatively to the left (lower values) and positively to the right (higher values). (Fig. 6.) This scale divides the total area of the normal curve into parts so that about 68.27 per cent of the observations fall within the range of +1 *S.D.* and −1 *S.D.* on each side of the mean; 95.45 per cent within +2 *S.D.* and −2 *S.D.*; and 99.73 per cent within ±3 *S.D.* from the mean. For all practical purposes the systematist may say that 100 per cent of an infinite number of observations on a variable character will fall within a range of ±3 *S.D.* from the mean of the sample. The normal curve, then, becomes the expression of the population from which the sample was taken, based on the mean and the standard deviation of the sample.

Figure 6 illustrates this use of the standard deviation in the sample given in figure 1 in which the mean is 121.75 units and the standard deviation is 5.46. When measured from the mean, +1 *S.D.* is at 121.75+5.46, or 127.21 units, and −1 *S.D.* is at 121.75−5.46, or 116.29 units; within these limits (116.29 to 127.21) 68.27 per cent of the observations of the total population would be expected to be found. Similarly, within ±2 *S.D.* of the mean, 121.75±(2×5.46), or from 110.83 to 132.67, 95.45 per cent of the observations would be found. Within the range of ±3 *S.D.*, or 105.37 to 138.13, 99.73 per cent of the observations would be found if the sample at hand properly represents the total population and approximates a normal curve. The method of ascertaining the probabilities of observations occurring outside these limits will be discussed in the section dealing with Probabilities.

The calculated range of the total population from which the sample was drawn is, therefore, 105.37 to 138.13 units. The next question that arises is, how does the available sample compare with this calculated population in variation. This question can be partially answered by placing the population range on the sheet with the sample range and frequencies, using the same class-intervals (fig. 1, last column). The maximum limit of the population range (138.13) will fall in the 138.5 class and the minimum (105.37) in the 104.5 class. The fact that the population has a wider range than the sample appears from the empty classes at the upper end of the distribution. A greater range for the population is to be expected in almost all cases. That the sample is asymmetrical appears from the fact that its observed range extends farther below the mean, nine classes, than above, six classes. This indicates that the sample is skewed to some extent, as most samples are by the effects of chance in sampling. In this case the mode (120.5) is less than the mean, so the skewness is positive but not

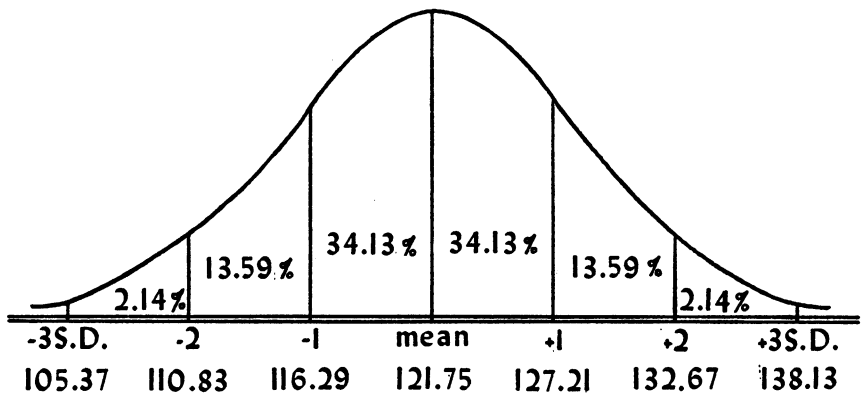


Fig. 6. Area of the normal curve. Based on data from figure 1.

enough (less than one standard deviation) to make this sample too skewed for estimating population variation by the above normal curve method.

Coefficient of Variability

It is often desirable in systematic work to compare the relative variability of a character in one sample with that of the same character in another sample, or to compare the relative variability between two different characters in the same sample. This cannot usually be accomplished by the direct use of the respective standard deviations, as these have been oriented around different mean values. For instance, a standard deviation of 10 units for a mean value of 100 units indicates little relative variability (one-tenth of average size), but if the mean value were 40 units, the relative variability would be large (one-fourth average size). Therefore, for the purpose of comparison, a measure is needed that takes into account the size of the mean as well as the variation of the characters to be compared. The Pearson coefficient of variation (more appropriately called coefficient of variability) is such a measure, expressing standard deviation as a percentage of the mean. The formula is 100 *S.D.*/mean.

As an example, assume that in one species the length of the thorax varies from 10 to 40 mm., with a mean value of 25 mm. and a standard deviation of 6.23 mm. In another species, the same character varies from 20 to 60 mm., with a mean of 40 mm. and a standard deviation of 10.63 mm. In order to compare these two samples to see which has the greater relative variability in this character and how much greater it is, the coefficient of variability is used, as follows. In sample A, $C.V. = 100 \times 6.23/25 = 24.92\%$. In sample B, $C.V. = 100 \times 10.63/40 = 26.58\%$. In other words, 6.23 mm. is 24.92 per cent of 25 mm., and 10.63 is 26.58 per cent of 40 mm. Had the differences between the standard deviations of the two samples been used, 6.23 and 10.63, it might have been concluded that sample A was about three-fifths as variable as sample B. However, this proportion being actually 24.92/26.58, sample A is about 93 per cent as variable as sample B in this character. Thus, there is really little difference in the relative variability when related to the size of the character in the two samples.

When the means of two compared samples are the same, a direct comparison of the variability can be made by comparing the values of the standard deviations.

This measure of variability can be used to great advantage in systematic work in disclosing the most variable samples within a population and therefore the samples that might contain hybrids. When correlated with geographical or ecological factors these samples might indicate subspeciation. Samples taken in zones of intergradation will tend to be more variable than those from non-intergrading zones because of the presence of these hybrids which have a larger number of different genes and therefore a greater range of variability in the characters.

Standard Errors

The statistical reliability of the measures of central tendency and of variation, in which the systematist is primarily interested, consists of a statement based on the standard error of these measures. In systematics, there are practically no possibilities of obtaining the entire population of an organism for study. The worker has available a given sample which he would like to use for estimating the particular population, but he would also like to know how the character he is using would be expected to vary if other samples were selected from the same population.

The systematist by now should realize that the total number of cases or specimens in the sample influences the various measures thus far obtained. He knows that the larger the sample the more nearly it probably approximates the population from which it was drawn, and that increases or decreases in the size of the sample influence the numerical values of such measures as the standard deviation and the mean one way or the other. The standard error is the same kind of probability estimate as the standard deviation and the same probability charts are used to determine the chances of additional observations occurring within a range expressed in units of standard error: 68.27 per cent will be within ± 1 *S.E.*, 95.45 per cent within ± 2 *S.E.*, 99.73 per cent within ± 3 *S.E.*, etc. It

is a means of ascertaining what the plus and minus variations from a given measure, such as the mean, standard deviation, and coefficient of variability, would be if other samples of similar size were taken from the same population; and what the probabilities are that other samples will be found whose measures of central tendency or of variation are at given distances outside the limits of these measures. Its numerical value varies inversely with the sample size, that is, the smaller the sample the larger the standard error for the measure, and vice versa. Also, the narrower the actual range of variation of a character, the more accurate will be the estimated variation of the population obtained from a small sample, since the chance that sample observations are far from the mean is obviously less when the range is small. The standard error of the standard deviation in such cases will also be small. Standard error does not correct mathematical errors made by the systematist.

The following formulas are of use to the systematist where a small difference in the value of the mean or standard deviation might cause a difference in interpretation and in the expression of the significance of certain measures. In many problems their most important use is in determining whether or not two or more samples are from populations that really differ in regard to a given measure, such as the mean or standard deviation, or if they are samples whose measurements occur within the expected chance fluctuations of a single population. Probable error and standard errors are measures of the same probable fluctuations in data but are based on slightly different formulas. Recent authorities (Simpson and Roe, 1939, p. 153) favor the standard error methods.

Standard error of the arithmetic mean: $S.E._M = S.D./\sqrt{N}$

Standard error of the standard deviation: $S.E._\sigma = S.D./\sqrt{2N}$

Standard error of the coefficient of variability: $S.E._V = C.V./\sqrt{2N}$

With the data in figure 1 ($N=96$, $M=121.75$, $S.D.=5.46$) as an example, the standard error of the mean would be $5.46/\sqrt{96}$, or .557, and the means of 99.73 per cent of additional samples of similar size from the same population would be expected to fall between $121.75+(3\times.557)$ and $121.75-(3\times.557)$, or between 123.42 and 120.08. The standard error would be larger if N were smaller, and the range within which the systematist might expect to find the means of other samples would therefore be greater. With a smaller number of individuals in the sample, the mean of the sample would not be expected to be so near the theoretical population mean and therefore not so accurate an indication of this population mean as a larger sample. Thus, if N were reduced to 10 in the above sample, the standard error of the mean would be 1.728 and the expected range would be from 116.57 to 126.93, a larger range within which the means of additional samples would fall.

In the same way the standard error of the standard deviation indicates the range within which the standard deviations of other samples from the same population would fall. The standard error of the standard deviation of the above example is $5.46/\sqrt{2\times96}$, or .394, and 99.73 per cent of the standard deviations of other samples from the same population would be between $5.46+(3\times.394)$ and $5.46-(3\times.394)$, or between 4.28 and 6.64.

From the standpoint of statistics, these measures of reliability are valid only if each sample is random; however, the advantages obtained from the use of standard errors far outweigh their biological disadvantages. In systematics it must be assumed that theoretically it would be possible to compare two random samples from the same population. Even though subsequent samples may not be completely random, the measures of standard deviation and coefficient of variability adhere closely to the theoretical limits of variation established by standard error methods. A given sample has a mean value and a standard deviation with which an attempt is being made to estimate the population range. These figures are accurate values for the particular sample but not necessarily for the entire population. Since this is true, it is advantageous that the systematist have some idea concerning the expected variation of these measures in a number of samples from the same population. These additional figures would provide an even closer estimate of the

variation of any one of these measures in the population.

The chief importance of these measures of reliability in systematics lies in their use in comparisons of calculated population estimates. If the calculated range and mean values for two samples seem to indicate that they are two different populations, then the standard error of these measures will indicate the expected sample fluctuations in each total population. They will therefore reveal the probabilities that these samples would overlap and approach the same values if many large series (samples) of each population were available. If the probability is very small (at least less than one chance in 741; see table 6, column 5) that a specimen belonging to one sample would fall within the calculated range of variation ($M \pm 3$ *S.D.*) of the same character in another sample it might be assumed that this would never occur. On the other hand, if by using the standard errors of the standard deviation and mean it was indicated that in other samples from each population one out of every 10 specimens belonging to one population would fall within the calculated range of the other, it might be concluded that the differences between the ranges for that particular character in the two samples were probably due to sampling or that the character is not necessarily a good indicator of completely divergent populations.

PROBABILITIES

The next problem in the analysis is the development of the method of obtaining, first, the probability of overlap between two separated population frequency distributions, and second, the degree of overlap in two overlapping population frequency distributions.

Large Sample Probabilities

In samples of 15 or more specimens, if it is assumed that the analysis has shown that the distribution of the characters approaches a normal curve, the calculated population range as represented by these samples may be obtained by subtracting $3 \times S.D.$ from the mean and adding $3 \times S.D.$ to the mean of each sample. For example, if the mean length of a character in one sample (A) were 125 units and the standard deviation were five units, and in another sample (B) the mean length were 160 units and the standard deviation were five units also, the calculated population range in A would be 110 to 140 units, and in B from 145 to 175 units. As shown in figure 7 the calculated population ranges of the two populations based on these two samples do not overlap. Since they are separated by only five units, the systematist desires to know what the chances are of getting an overlap between the two by increasing, theoretically, the size of his sample. In other words, what is the probability that A will overlap B, or vice versa; that is, what are the chances of finding a specimen belonging to population A that is larger than 145 units, the minimum calculated range limit of B, or a member of B that is smaller than 140 units, the maximum calculated range limit of A.

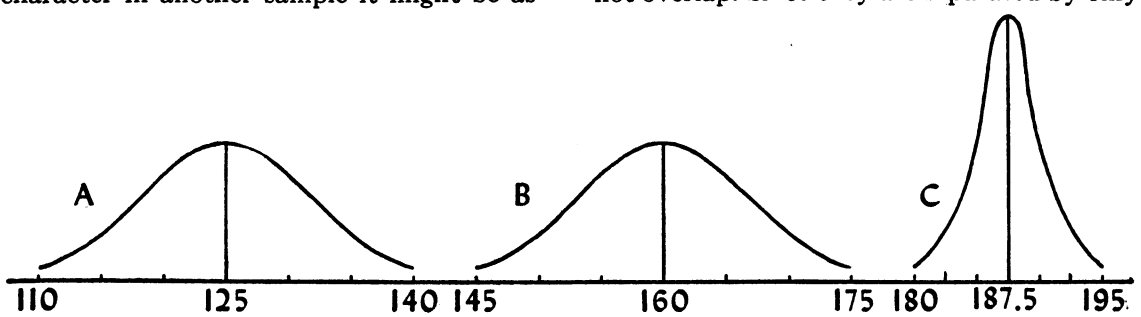


FIG. 7. Non-overlapping calculated population frequency distributions.

It has been previously pointed out that in the normal probability curve 99.73 observations out of every 100 will fall within plus and minus three standard deviations from the mean. Thus, in sample A, 49.865 observations out of every 50 that are 125 units long or over will be within the limits of 125 and 140 units. In B, 49.865 out of every 50 that are 160 units long or less will be within the limits of 145 and 160 units. If the ranges of A and

B were adjacent at 140 units (mean of B at 155 instead of 160 units), out of every 100 individuals (50 on each side of the mean) from either population, about 13/100 of one individual from each sample would be expected to fall within the calculated range of the adjacent sample. However, in figure 7 the distributions of A and B are separated by one standard deviation (5 units) of either A or B which in this example have equal standard deviations. What, then, are the probabilities of overlap with this increased distance between the frequency distributions? Statisticians have made available tables for determining these probabilities (table 6). At a distance equal to and more than four standard deviations from the mean, one can expect to get one specimen out of about every 31,500 specimens (column 5). Therefore in sample A, one would expect to get one specimen out of about every 31,500 specimens that would overlap B at 145 units, and inversely (since A and B have the same standard deviation) one out of about 31,500 of B that would overlap A at 140 units. If this difference between the distributions had been equal to two standard deviations (10 units) the probabilities would have been one out of about every 3,000,000 specimens of each that would overlap the other.

In the comparison of biological samples, the systematist is seldom if ever dealing with character frequency distributions that have exactly the same standard deviation. The complications resulting from unequal standard deviation values are made plain in figure 7B and C, in which B with a mean of 160 and a standard deviation of 5 is compared with C which has a mean of 187.5 and a standard deviation of 2.5 units. If these two distributions had been adjacent at 175 units (mean of C at 182.5), the probabilities of the larger values of B overlapping the smaller values of C would be equal. However, the probability that B will overlap more of the range of C is greater than that C will overlap more of B owing to the differences between the values of the standard deviations. For example, in figure 7, the two distributions B and C are separated by 5 units, or by one standard deviation of B or two standard deviations of C. With reference to table 6 for sample B at four standard deviations from the mean, one would expect to find one specimen out of about every 31,500 that would overlap C at 180 units. If C were to overlap B at 175 units, specimens would have to be beyond five standard deviations from the mean of C, since the standard deviation of C is 2.5 units and the distributions are separated at 175 units by 12.5 units on the minus side of the C mean. In further reference to table 6, it can

EXPLANATION OF TABLE 6

COLUMN 1: The range of variation divided into standard deviation units on each side of the mean.

COLUMN 2: "Fractional Parts of the Total Area . . . under the Normal Probability Curve, corresponding to distances on the baseline between the mean and successive points laid off from the mean in units of standard deviation" (Thurstone, 1928, p. 91). These are the figures found in statistical tables of areas of the normal or probability curve (such as in Davenport and Ekas, 1936). In this table, the total area is considered to be 100, so these figures are also percentages of the total area.

COLUMN 3: The systematist, concerned with the probability of overlapping between the ranges of two samples, is interested in the number of specimens that may fall on the other side of certain points expressed in standard deviation units on one side or the other of the mean; that is, those that are larger than the mean plus 3 *S.D.* or smaller than the mean minus 3 *S.D.* He assumes that 50 per cent of the specimens lie on the other side of the mean, that is, are smaller than the mean or larger. Thus, column 3 gives the percentage, or the number of specimens out of 100, that are within the standard deviation limit that he has set. This includes all the specimens on one side of the mean plus the specimens between the mean and a point expressed in standard deviation units on the other side of the mean.

COLUMN 4: The percentage, or the number of specimens out of 100, that may be expected to fall outside of the limit on one side of the mean; that is, in a sample of 100 specimens, 98 may be expected to be smaller than the mean plus 2 *S.D.*, and two may be expected to be larger. This may be expressed as 2 per cent, or two chances out of 100.

COLUMN 5: The percentages in column 4 expressed inversely; that is, there is one chance out of 44 that a specimen will be larger than the mean plus 2 *S.D.*

TABLE 6

TABLE OF PROBABILITIES

1	2	3	4	5
Standard Deviation Units	Percentage of Cases Between Mean and Plus or Minus *S.D.* Unit	Total Number Out of 100 Within *S.D.* Limit on One Side of Mean Plus Those on Opposite Side	Number Out of 100 Beyond *S.D.* Limit on *One* Side of Mean	Chances in this Number of Observations of Having 1 Specimen Outside the *S.D.* Unit Limit on One Side of Mean
Mean	00.000 00	50.000 00 or 50	50.000 00	2.0
0.1	3.983	53.983 54	46.017	2.2
0.2	7.926	57.926 58	42.074	2.4
0.3	11.791	61.791 62	38.209	2.6
0.4	15.542	65.542 66	34.458	2.9
0.5	19.146	69.146 69	30.854	3.2
0.6	22.575	72.575 73	27.425	3.6
0.7	25.804	75.804 76	24.196	4.1
0.8	28.814	78.814 79	21.186	4.7
0 9	31.594	81.594 82	18.406	5.4
1.0	34.134	84.134 84	15.866	6.3
1.1	36.433	86.433 86	13.567	7.4
1.2	38.493	88.493 88	11.507	8.7
1.3	40.320	90.320 90	9.680	10.3
1.4	41.924	91.924 92	8.076	12.4
1.5	43.319	93.319 93	6.681	15.0
1.6	44.520	94.520 95	5.480	18.2
1.7	45.543	95.543 96	4.457	22.4
1.8	46.407	96.407 96	3.593	27.8
1.9	47.128	97.128 97	2.872	34.8
2.0	47.725	97.725 98	2.275	44.0
2.1	48.214	98.214 98	1.786	56
2.2	48.610	98.610 99	1.390	72
2.3	48.928	98.928 99	1.072	93
2.4	49.180	99.180 99	.820	122
2.5	49.379	99.379 99	.621	161
2.6	49.534	99.534 100	.466	215
2.7	49.653	99.653	.347	288
2.8	49.744	99.744	.256	391
2.9	49.813	99.813	.187	535
3.0	49.865	99.865	.135	741
3.5	49.976 74	99.976 74	.023 26	4,300
4.0	49.996 83	99.996 83	.003 17	31,500
4.5	49.999 66	99.999 66	.000 34	294,000
5.0	49.999 97	99.999 97	.000 03	3,000,000
5.5	49.999 99	99.999 998	.000 002	50,000,000
6.0	49.999 99	99.999 999 90	.000 000 1	1,000,000,000

be seen that the possibilities of having specimens of C that overlap B at 175 units are one specimen out of about 3,000,000.

From this example it can be seen that when two frequency distributions with dissimilar standard deviation values are being compared, the probabilities as derived from the distribution having the larger standard deviation value will determine the greater probability of overlap between the two distributions. The distribution with the smaller standard deviation (with less variation) is less apt to overlap an adjacent distribution having a larger standard deviation (variation) than vice versa.

2.5 also overlap, but in this case the probabilities are unequal. The point $M-3$ *S.D.* of C is at 170 which is $M+2$ *S.D.* of B. From table 6 it is found that 98 per cent of B is distinct from C. But $M+3$ *S.D.* of B is at 175 which is $M-1$ *S.D.* of C, and thus only 84 per cent of C are distinct from B, or about 15 per cent of C are indistinguishable from about 2 per cent of B.

The probability that there is more overlap than that given in the calculated population range ($M \pm 3$ *S.D.*) may be obtained from table 6. In the example given in figure 8 there would be expected only one specimen of A out of every 31,500 (4 *S.D.* from mean

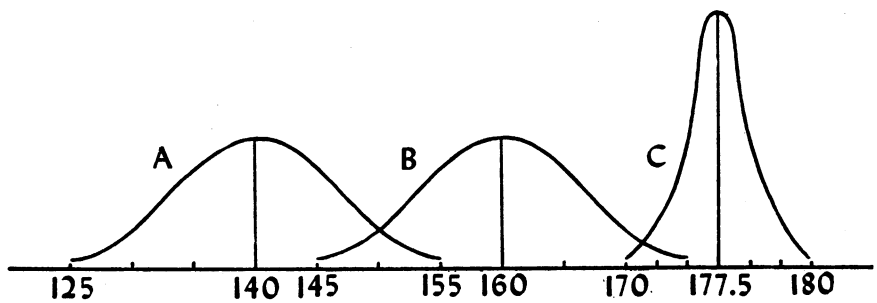


FIG. 8. Overlapping calculated population frequency distributions.

The problem of estimating the degree of overlap involves the same preliminary treatment as above: normalcy, mean, and standard deviation. The same examples as given in figure 7 are used except that A is moved up into the range of B, so that they overlap by two standard deviations each; mean of A at 140 units and standard deviation of 5 units, and mean of B at 160 units and standard deviation of 5 units also (fig. 8). In reference to the normal curve (fig. 6) each of these overlapping areas, from $+1$ *S.D.* to $+3$ *S.D.* in A and from -1 *S.D.* to -3 *S.D.* in B, contains about 15.73 per cent (13.591+2.140) of all the observations in each sample. In other words, about 84 per cent (50 per cent on one side of the mean plus 34 on the other (table 6) of the individuals in A and the same number in B can always be distinguished from one another and about 16 per cent of each cannot. In curves with the same standard deviations, the probabilities are always equal.

Sample B with a standard deviation of 5 and sample C with a standard deviation of

of A) that would fall beyond the mean of B (160 units) and thereby cause only 50 per cent of B to be distinguishable from A; and less than one specimen of C out of every 3,000,000 that would fall within $M+1$ *S.D.* of B (at 165) and thereby cause only 84 per cent of B to be distinguishable from C.

Where small samples with large standard errors of the mean have to be dealt with, the systematist should figure the population range estimates from the mean plus three times the standard error of the mean added in the direction of the mean of the sample being compared. For example, if in sample A of figure 8 the standard error of the mean (140 units) were ± 2 units, the calculated population range would be figured from 140+6 units, when A and B are being compared, as B is in the positive direction. This would give a population positive range of 146 to 161 rather than 140 to 155 and would therefore reduce the chances of A's being distinct from B, owing to the small number of specimens involved and the extent of the

sample range. Where the standard error of the standard deviation is large, this should also be taken into consideration and applied as in the mean and in addition to it if necessary. If the standard errors are small, they need not be considered, as the probability figures are rough enough estimates to take care of small deviations.

Small Sample Probabilities

Statistically, the probabilities of observations occurring outside the calculated population limits ($M \pm 3$ *S.D.*) are proportionally greater in samples of fewer than 15 specimens than in larger samples. It is therefore necessary to consult a special table for these probabilities. Table 7, adapted from Simpson and Roe (1939, p. 206), is sufficient for most systematic needs. In this table, *N* is the number of specimens in the sample, *S.D.* is the standard deviation unit of the scale from the mean of the sample, and at the bottom of each column are given the approximate probabilities for each column of standard deviation values. Thus, in a sample of five specimens, with a given standard deviation, the probabilities of finding additional specimens outside +2.1 *S.D.* (or −2.1 *S.D.*) from the mean are one in every 10 specimens. For a value of +4.6 *S.D.* or −4.6 *S.D.* from the mean for this same sample of five specimens, the probabilities of finding observations outside these limits are about one in 100. In larger samples (table 6) a deviation of 4.6 *S.D.* would give a probability that not more than one in about every 835,200 specimens would be outside these limits.

TABLE 7

Small Sample Probabilities

N	*S.D.* on Each Side of Mean			
2	6.3	12.7	31.8	63.7
3	2.9	4.3	7.0	9.9
4	2.4	3.2	4.5	5.8
5	2.1	2.8	3.7	4.6
6	2.0	2.6	3.4	4.0
7	1.9	2.4	3.1	3.7
8	1.9	2.4	3.0	3.5
9	1.9	2.3	2.8	3.3
10	1.8	2.3	2.8	3.3
11	1.8	2.2	2.8	3.2
12	1.8	2.2	2.7	3.1
13	1.8	2.2	2.7	3.1
14	1.8	2.2	2.7	3.0
15	1.8	2.1	2.6	3.0
Prob.	1/10	1/20	1/50	1/100

CORRELATION

Application of the preceding techniques enables the systematist to make certain comparisons both within and between samples in which characters have been measured and their variation expressed in terms of standard deviation. The relative variabilities of these measured characters are compared with one another by means of the coefficient of variability. Population ranges have been plotted and compared as to the amount of overlap. All through this discussion, and indeed throughout all systematic studies, there is a constant tendency to search for and to express the relationships between variables and to apply this relationship in systematic interpretation. The quantitative means of expressing the relationships between series of variables are by the use of correlations. The purpose of this section is to develop the method for determining simple correlation by means of scatter diagrams for the expression of character relationships, and to indicate how these diagrams may be used to express other relationships.

Statistical correlations are concerned with the relationships in a series of measured variables; a variable in this case is a character that, within a sample, has a series of values, such as length or breadth expressed in units of measurement, etc. If the relationship between two characters is such that, as the value of one of them increases the other increases also, the correlation is said to be positive. If, on the other hand, as the value of one of the characters increases the other decreases in value, the correlation is inverse or negative. In systematic studies it is important to know, for instance, if an increase in length of a structure is accompanied by a proportional increase or decrease in width, and to what degree they are correlated.

For ordinary purposes in systematic work it is unnecessary to go into the more complicated correlation tables and the calculation of correlation coefficients. In most practical work it is sufficient to plot a scatter diagram

and to get directly from it the desired information even though it expresses only roughly the relationship between the two variables. The only difference between a scatter diagram and a correlation table is that in the scatter diagram there is a point for every specimen, whereas in the correlation table only the total frequency in numerical form is entered in each square. The correlation coefficient is a numerical way of interpreting the scatter diagram, and the diagram often gives more information to the amateur statistician than the single numerical value of the coefficient. Correlation coefficients serve as measurements or indices of degree of relation and are therefore of most use when a great number of relationships are to be compared. The standard deviations of ratios may be used also, in that the smaller the standard deviation the more positive the correlation between the two variables in the ratio.

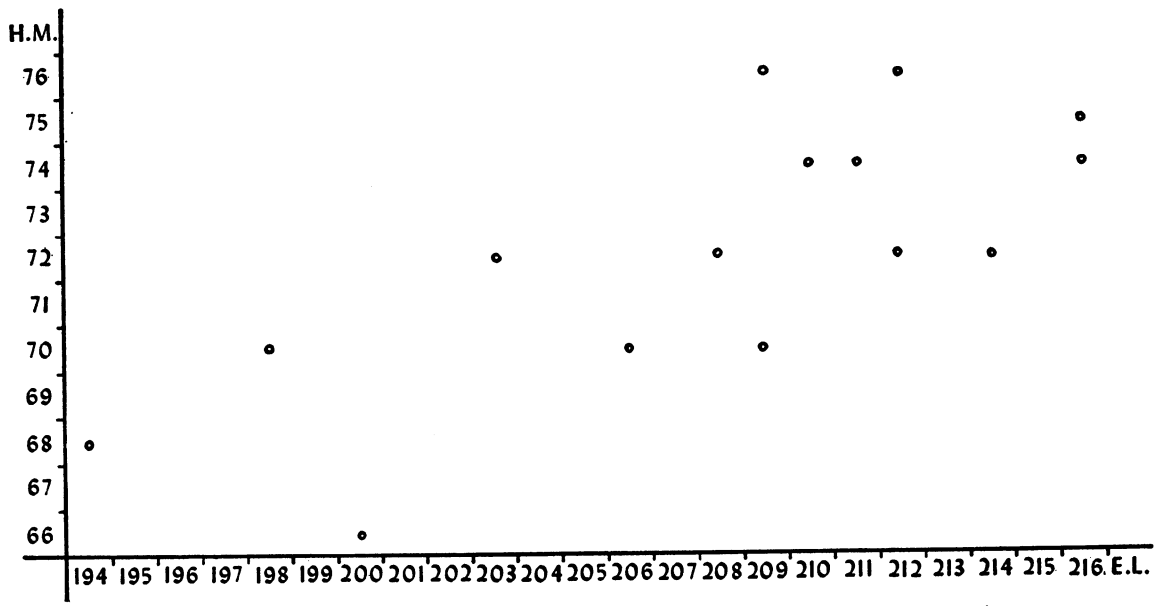


FIG. 9. Scatter diagram of 15 specimens (table 1, sample A) showing positive correlation between H.M. and E.L.

Scatter diagrams are used for showing graphically the relationship between two variables, not only the presence or absence but also the degree of relationship. To construct a scatter diagram, the systematist first selects the two variables to be compared and ascertains the minimum and maximum limits of the observed range of each. Using paper divided in many squares, he then begins in the lower left-hand corner with the minimum values of each character, running the scale for one variable vertically (from down to up) and the other horizontally from left to right. Figure 9 illustrates such a diagram using the data for E.L. and H.M. in table 1, sample A, in which 66 to 76 are the limits for H.M. and 194 to 216 for E.L. A scale is established for each variable, as was done in the frequency distributions, with a convenient number of classes (in this case with class-intervals of one unit for both), and the classes are entered along the scales to the maximum limit of each variable, with H.M. on the vertical scale and E.L. on the horizontal. The scales, as in frequency distributions, can be adjusted for any character with class-intervals of any convenient size. The more variation in a character, the larger the class-interval may be made to facilitate plotting. After the scales are established and entered on the graph paper, the individual entries are made in the squares made by the intersecting of the proper vertical and horizontal columns. For example, for specimen 1 in sample A of table 1 with an H.M. value of 74 units and E.L. of 216, the

scale value of 74 is found in the H.M. vertical scale, and in the square where the 74 column intersects column 216 on the E.L. scale a check is entered for this specimen. This process is repeated for each specimen in the sample. The sum of all the checks in the columns, added vertically as well as horizontally, gives a frequency distribution table for each variable; in figure 9 the sums of the horizontal columns are a frequency table for H.M., and of the vertical columns, for E.L. In this example the completed figure shows a positive relationship of H.M. and E.L., or, in other words, as the elytral length increases there is almost a proportional increase in the width of the hind thoracic margin.

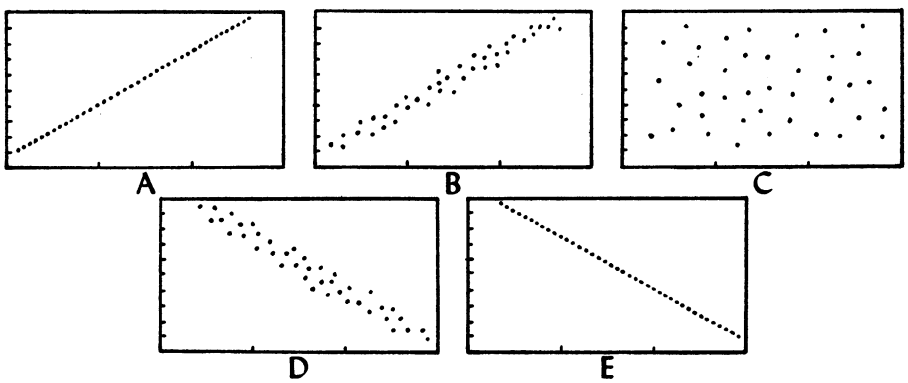


FIG. 10. Degrees of correlation shown by scatter diagrams. A. Perfect positive. B. Positive. C. Zero or none. D. Negative. E. Perfect negative.

Figure 10 illustrates the different degrees of relationship that can be shown by scatter diagrams. A illustrates a perfect positive relation between two variables such as weight and size. In biological studies it is very improbable that such a perfect relationship would ever exist, as it may in such relationships as weight and volume of pieces of metal. B shows a positive but imperfect relation due to the fact that although both variables tend to increase in general their proportions are not constant. This type of relationship is a very common one in biological material. The degree of scatter illustrates the relationship in that the more closely the distribution of the specimens approximates a straight diagonal line the more closely proportional are the characters. C shows the absence of relationship between two variables. This means that as one variable changes, the other one does not change in proportion to it. D illustrates a negative relationship, in which as one variable increases in value, the other tends to decrease. This type of relationship is also common in biological data. E shows a perfect negative relationship which, like the perfect positive, probably never exists in biological material. It indicates that as one variable increases the other one decreases in constant proportion.

In practical application these scatter diagrams enable the systematist to illustrate evolutionary trends within the species or among related species. The positive relationship indicates that there are parallel trends in the two variables or species and expresses the degree of this development. Negative relationship indicates that the trends are parallel but reversed. Figure 10C, showing lack of relationship, indicates that no parallel development is taking place and that development in these characters is progressing in different and unrelated ways.

Another application of this correlation method and probably the most important for the systematist is in establishing the relationship between morphological evolutionary trends and geographical distribution. This is done by using a measured variable and a logical distributional sequence of the analyzed samples, as will be shown in the next section. Population ranges calculated for each sample and arranged vertically in a consistent progression of localities so that the geographical variable is horizontal across the top or bottom and the measured variable scale ver-

tical will line up the series of ranges and will indicate evolutionary trends as well as direct environmental effects that are correlated with geographical distribution (clines). If there is no correlation between the character and the geographical sequence, then this character does not indicate geographical differentiation that could be used for subspecific differentiation. Most geographical distributional patterns do not show linear trends, but the correlation between the locality and structural change is none the less evident when the data are arranged in tabular form with a locality sequence.

COMPARISON OF SAMPLES

WHEN THE SYSTEMATIST compares two populations as represented by samples, qualitatively or quantitatively, he attempts to establish their classification status on the basis of their biological and morphological similarities and differences. The phylogenetic relationship is established primarily through consideration of the similarities, and the classification status through the differences between the samples. Having observed and measured the divergent characters separating two samples, the systematist is confronted with the problem of evaluating these differences according to their significance. That is, do these differences define specific or subspecifically distinct biological units or do they merely represent uninterrupted intraspecific evolutionary trends which have not resulted in biological differentiation.

Where the differences involve the presence of a fundamental character in one sample and its complete absence in another, without variation between, the solution is evident morphologically if the two samples are living together in the same region under the same environmental conditions without hybridization (sympatric; Mayr, 1942). If the two allied samples, with this same difference, do not live in the same territory and do not therefore have the opportunity to hybridize (allopatric; Mayr, 1942), there is no proof that they are reproductively incompatible and therefore that the morphological characters represent or are correlated with specific differences. Many systematists when describing allopatric species assume that if these allied species were brought together there would be no fertile hybridization other than that caused by the breaking down of a physiological or other barrier by unnatural laboratory or field conditions. Obviously, positive proof is wanting in such instances, the species being based entirely on selected morphological differences that may or may not indicate reproductive incompatability. The sympatric species condition appears to be the only indirect proof available of the existence of distinct species in nature and should be carefully considered in establishing the status of a sample.

Probably the most commonly used morphological characters are those that are not either present or absent but are present to a greater or less degree in the samples. In these characters the systematist recognizes a certain amount of variation and attempts to establish the population extremes by using the observed extremes in the available specimens. If there appears to be no overlap of the observed ranges of the samples, they have in many instances been called distinct species regardless of whether they are sympatric or allopatric. It can be seen at once that these continuous variants are less easily delimited than the present or absent discontinuous variants and that the problem of deciding what they represent resolves itself in part into one of determining the probability of finding variations outside the observed sample ranges and the calculated population ranges. The derivation of the probabilities of two samples representing morphologically distinct taxonomic units when their frequency distributions do not overlap involves the computation of the standard deviation of the sample which gives an estimate of the population variation. If the calculated ranges (mean, plus and minus three standard deviations) of the divergent characters in two samples are separated, there is little probability that specimens will or do occur outside these limits and they might, if biologically distinct, be called species. If the ranges are not separated, they cannot be called species on the basis of these characters since they do not indicate biological isolation.

When the calculated ranges overlap, the systematist has a difficult problem in determining the taxonomic status of the populations. He knows that any two samples, even though they may be from the same population or from the same area, etc., are different in greater or less degree. Also, he is aware that as a species spreads over a rather uniform area, the evolution of characters in local populations is gradual until either a barrier of some sort cuts off a portion of the original population or a more radical change is stimulated by varying conditions.

Figure 11 is a graphic presentation of character gradients (clines) of species that diminish in size over a north to south distribution. The heavy lines indicate the mean sizes of the

species and the lighter lines the limits of the calculated range of each. In species A there is a gradual reduction in size throughout the entire geographical distribution. It is obvious that any samples taken along this north to south distribution would contain a high proportion of individuals identical to those in samples taken on either side of it and that rier" causes an abrupt decrease in size ending in a smaller organism that continues from point 2 as BB. In this species all the intermediate sizes occur, but the maximum percentage of intergradation (overlap) between B and BB is small because of the effect of some environmental change in this narrow geographical zone (1 to 2). If samples were

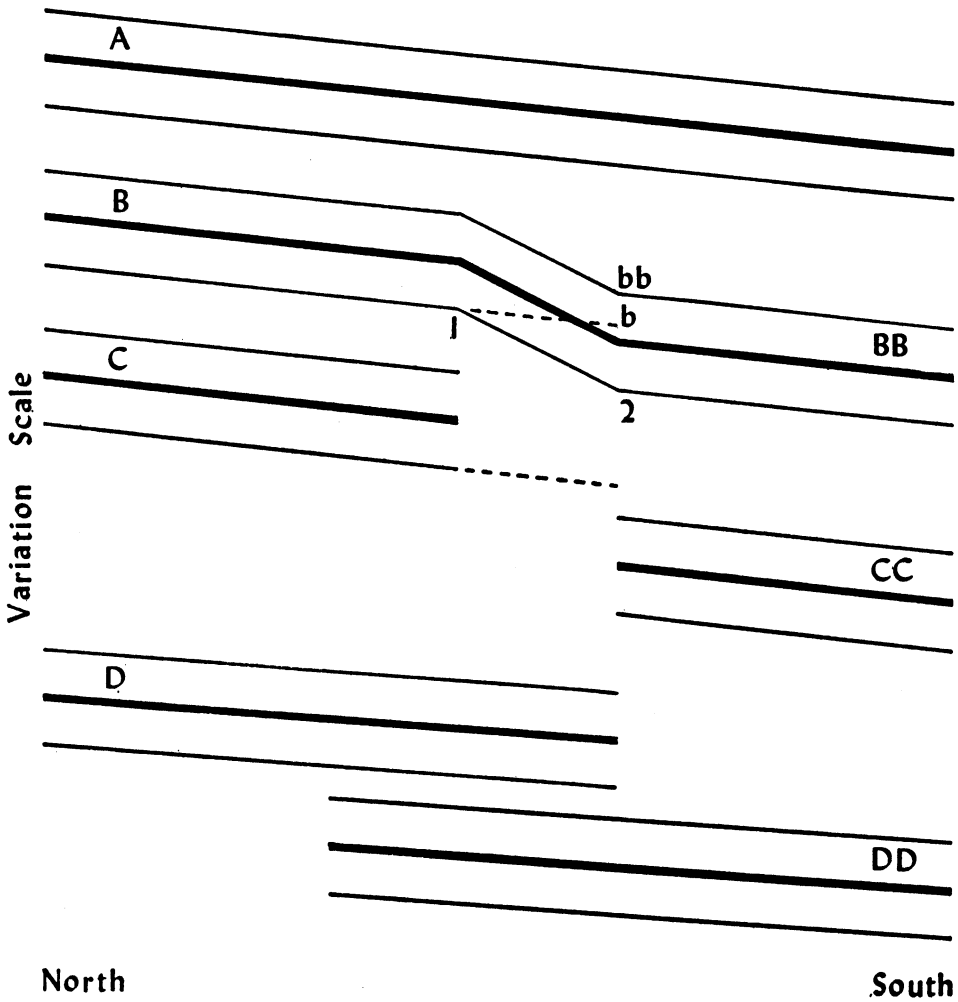


FIG. 11. Character gradients of species correlated with geographical distribution.

there would be no abrupt transition from one end of the gradient to the other. Any names proposed for samples along any such continuous gradient are obviously of little value in classification even though they might represent the extremes of this continuous variant.

In species B there is a gradual reduction in size until point 1 is reached, at which a "barselected from points 1 and 2, a small proportion of the specimens in each sample would fall within the range of the other (difference between b, the minimum of *B*, and bb, the maximum of BB) but not in large enough quantities to obscure the distinctness of B and BB. These are commonly called geographical subspecies. There is every gradation from A,

in which no abrupt geographical change is evident, to C, in which a barrier has isolated CC from C. For subspecific recognition it is necessary that character differences be accompanied by geographical or ecological isolation; that is, BB cannot be living in the same locality or under the same conditions as B, unless they converged after being isolated. If isolation of some sort had not occurred, divergence would have been obscured or would never have occurred because of unrestricted gene flow between the populations.

Species C represents the next evolutionary step from B in which C and CC have been separated from each other geographically and the intergrading forms have been eliminated. Here the probability of C overlapping CC in characteristics must be considered. If these probabilities are low, C and CC could be recognized as morphologically distinct species (allopatric). Proof of their biological distinctness would be available only if CC extended its distributional range northward (or C southward) into territory occupied by the other without hybridization, as shown in D and DD (sympatric). From many standpoints it is most practical that these allopatric, non-intergrading, but closely allied populations (C and CC) be recognized as species. It is known that they are evolving away from one another and that the samples were only from a particular stage in this evolution. A clear-cut distinction can be demonstrated between the two samples on the basis of one or more characters, morphological or otherwise. Furthermore, there is no proof that the populations will hybridize, and in many cases there is little possibility of eliminating the barrier that prevents their intermingling. Much systematic chaos would result if all of these allopatric entities were recognized as subspecies until such time as they were actually shown to be incapable of hybridization.

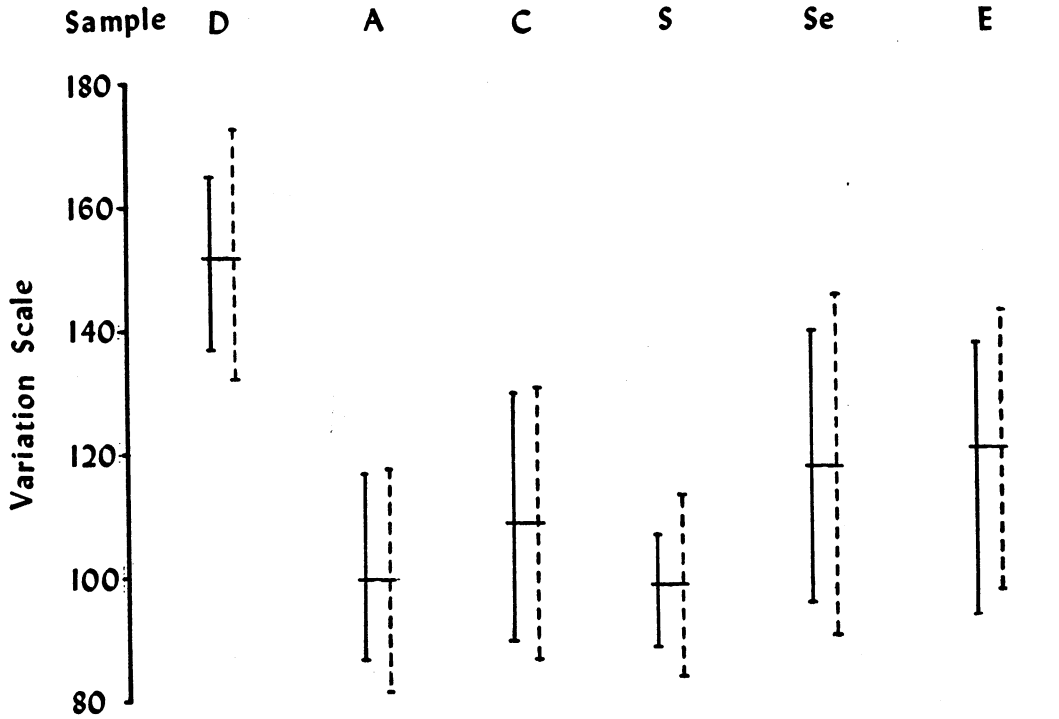
In actual practice, it is impossible for the systematist to ascertain whether he has a B-BB or a C-CC species if his samples do not show intergradation. If intergradation is evident, the species is polytypic as in B-BB. On the other hand, if no intergradation is exhibited in the samples, it may mean only that the collecting was incomplete or faulty. For this reason, it is desirable for the systematist to show what the chances actually are for the frequency distributions of two narrowly separated populations to overlap in their characters. If the chances for overlap in the characters are small, then it may well be assumed that they represent morphologically distinct species even though they appear at the moment to be allopatric. If the chances for overlap in the characters are great, then it is possibly best that they be recognized as subspecies until such time as they can be shown to be biologically isolated. This same probability analysis applies to sympatric samples suspected of being distinct species. If the data on the characters in two geographically distinct samples show an actual overlap in the observed range of variation, then it is necessary to obtain an estimate of the amount of overlap in their calculated ranges to see if it is feasible to recognize them as morphological subspecies. Biological confirmation of the morphological divergence is again necessary for a satisfactory conclusion in such instances.

GRAPHIC METHODS

Systematic interpretation is concerned primarily with comparisons of means and ranges of variation which generally require the presentation of long and complex series of measurements and calculations. Since modern publication costs prohibit the publishing of complete and largely unorganized raw data or frequency tables for each sample, it is necessary for the systematist to adopt some means of illustrating as concisely as possible all the information necessary to the understanding of the problem by the reader. This information, in its most desirable form, should convey to the reader a graphic picture of the facts and relationships as well as a brief numerical tabulation of the values of the pertinent mathematical measures involved.

One of the best methods for illustrating systematic data is shown in figure 12, in which the variation scale is vertical, the designations of the samples (by species, subspecies, or locality) are at the top of each column, and the number of specimens, mean, standard deviation, and the standard errors of these measures for each sample are at the bottom. The solid vertical lines in the graph indicate the observed ranges of the samples, the broken lines the calculated population ranges ($M \pm 3$ *S.D.*), and the horizontal

cross bars the means. Combined graphs and tables such as this convey most of the information relevant to the interpretation of the systematic problem. For additional methods of graphic presentation, the worker may consult various texts and papers, especially Simpson and Roe (1939, p. 305) and Simpson (1941a, 1941b, and 1942).



	D	A	C	S	Se	E
N	29	229	186	23	255	253
M	152.4 ± 1.1	99.8 ± .40	108.9 ± .54	99.0 ± 1.0	118.2 ± .58	120.7 ± .47
S.D.	6.8 ± .89	6.0 ± .28	7.3 ± .38	4.8 ± .70	9.2 ± .41	7.5 ± .33

FIG. 12. Comparison of *Omus* samples.

APPLICATION

THE MOST IMPORTANT RESULT desired from the application of the preceding technique is to give a more reliable method of evaluating character differences. The differences obtained can then be correlated with biological differences and used to supplement these in determining the classification status of the samples. The following discussion is presented to show some of the complexities involved in establishing the status of even properly analyzed (biologically and morphologically) samples.

Many contemporary writers in systematic zoology use mean differences in establishing criteria of significance to separate classification categories. Although this method has certain advantages, it has a number of serious disadvantages which, in the writers' opinions, prohibits or restricts its use, especially in groups represented by adequate numbers of specimens and reasonably continuous geographical samples. Differences based on means and standard errors gain their significance from statistical measures which do not, in a majority of cases, express biological differences. They give the systematist no indication of the total variation in the samples or of the percentage of identical individuals in each. They show the trend of the majority of the individuals but fail to indicate the more divergent specimens which are the most advanced and therefore, from an evolutionary standpoint, the best indicators of population divergence so long, of course, as the material is not biased by teratological specimens.

Probably the most serious single disadvantage of this mean difference method lies in the fact that the significance of differences based on the mean value and standard error of the mean vary inversely with increases in the number of specimens. That is, as the size of the sample is increased, the standard error gets smaller and the significance of the differences is increased rather than remaining fairly constant as it should in biological material, and as it does in the direct standard deviation of the variation method herein recommended.

If taxonomy is to reflect reality, species must be looked upon as natural discontinuous biological units, something tangible or real, as contrasted with the somewhat arbitrary delimitations of the continuously variable subspecies and arbitrarily delimited higher categories. If the systematist is going to progress in his study of phylogeny and taxonomy he must have certain "landmarks" from which studies can be oriented. The species is such a landmark. In order to have stability, it must be possible to recognize, on the basis of some inherent characteristics, all the individuals making up the specific population. In other words, it is required that there should be complete divergence between the values of the most divergent inherent characters separating two species, provided, of course, that in the case of morphological characters it can be demonstrated that the characters indicate accompanying genetic or biological isolation of the populations and not merely phenotypic expressions or individual freaks. The nature of the most divergent characters is not fixed; they can be morphological but may also be physiological, neurological, serological, ecological, or psychological, etc., and in addition they must be correlated with completely divergent biological characteristics. If two species are biologically (reproductively) isolated but no morphological characters show complete divergence, then the separation taxonomically can be made only on the divergent biological or genetical characters. If, on the other hand, the biologically distinct species have in addition correlated, morphologically divergent characters, then the more easily discernible morphological characters can be used to define and illustrate the more obscure and less easily definable genetic differences. Species based entirely on morphological characters lack the biological confirmation necessary for unquestioned acceptance taxonomically.

It is necessary that the morphological species concept be supplemented with an understanding of the biological species concept (see Mayr, 1942). As mentioned previously, the sympatric condition of species is indirect proof that they are biologically distinct. However, the allopatric species, although morphologically distinct at present, may not

prove to be biologically distinct at some future time. If the physical barrier between these strictly morphological species is eliminated (reinvasion), their biological compatability may allow interbreeding, in which case the morphological distinctness may cease to exist owing to the presence of morphological intermediates or hybrids. It is known, however, that there is considerable evidence supporting the assumption that geographical isolation precedes biological isolation and, therefore, that the allopatric species is one of the stages in the formation of biologically distinct species. From this it can be seen that there is a good possibility that even if the barrier separating the allopatric species were eliminated, and in most cases there is little if any possibility of this occurring, the species would be both morphologically and biologically distinct and, therefore, a sympatric species.

Schools of thought exist in which the allopatric morphological species is not recognized as such, all or part of them being considered as subspecies even though the two species are completely distinct in one or more characters, whether morphological, ecological, or otherwise. This school of thought appears to be denying itself the use of much evidence, as indicated above, in drawing conclusions, and such a procedure as it follows places the allopatric species on a basis similar to that of the subspecies. It assumes that geographical isolation as evidenced by morphological change does not necessarily indicate biological or genetic change and its resulting isolation. Although it is not possible to say that such a conclusion is false and unjustified, it can be pointed out that it is based on an assumption which disregards evidence of the effect of isolation on physiological and genetic characters and also the fact that the species involved are completely separable by means of various indicators.

The establishment of a subspecies involves the consideration of two features affecting the population. These are the evolutionary and geographical (including ecological) features; the first, evolutionary, is inherent in the organism and the second, geographical, in the environment. The consideration of the evolutionary features involves the detection of the divergent characters in the populations and the establishment of the amount of morphological and biological overlap existing between the geographically distinct samples. The geographical or ecological features are considered as the physical forces acting on the population as accelerators of genetic divergence and as partial isolating mechanisms, preventing the uninterrupted genetic exchange between two populations.

To determine what constitutes adequate physical barriers is a difficult problem, since they are different in different species. Separation in terms of distance only may not mean isolation, and closely proximate samples may be topographically or ecologically isolated. The establishment and justification of a subspecies involve a knowledge of the geography, topography, and ecology of the region occupied as well as of the degree of evolutionary divergence of the populations. Subspecies are always somewhat arbitrary and therefore more difficult to recognize than are species, and care should be exercised in their erection or in relegating species to subspecies.

SUMMARY

INASMUCH AS THE FREQUENCY DISTRIBUTIONS of many biological data approach the normal probability curve as developed by statisticians it is often possible for the taxonomist to adopt various statistical measures based on this type of distribution in the systematic analyses of biological material. The ones most commonly employed by the systematist to analyze normally distributed data are those of the central tendency (mean, median, and mode), of variation (standard deviation), of variability (coefficient of variability), and of reliability (standard errors).

The mean is an expression of the average tendency in the sample and serves as a point on the variation scale from which the measure of variation can be oriented. The median and mode are measures of central tendency used primarily in comparing the frequency distribution of the sample with the normal curve to reveal possible skewness. The standard deviation is the measure of variation with which the systematist estimates the range of variation in the population from which a particular sample was taken. The population range is calculated as the mean of the sample plus and minus three standard deviations of the sample ($M \pm 3$ *S.D.*) which gives the systematist the range of variation within which would occur approximately 100 per cent of the total population represented by that sample. The coefficient of variability enables the systematist to establish the relationship between the variation and the mean size of the sample, giving the relative variability which can then be used in making comparisons. The standard errors indicate the reliability of the preceding measures and are used to show the systematist the theoretical range of variation of any of these measures, within which range would be found the same measures of additional samples drawn from the same population. With these statistical tools and a knowledge of the biology and distribution of the population, the systematist should be able to establish more accurately the classification status of his samples.

REFERENCES

DAVENPORT, C. B.

1904. Statistical methods with special reference to biological variations. New York, John Wiley and Sons.

DAVENPORT, C. B., AND MERLE P. EKAS

1936. Statistical methods in biology, medicine and psychology. New York, John Wiley and Sons, Inc.

FISHER, R. A.

1935. Statistical methods for research workers. Biological monographs and manuals, V, vol. 1. Edinburgh and London, Oliver and Boyd.

KLAUBER, L. M.

1937. A statistical study of the rattlesnakes. Occas. Papers, San Diego Soc. Nat. Hist., no. 3, pp. 1–56.

1938. A statistical study of the rattlesnakes. *Ibid.*, no. 4, pp. 1–53.

1939. A statistical study of the rattlesnakes. *Ibid.*, no. 5, pp. 1–61.

1941. Four papers on the application of statistical methods to herpetological problems. Bull. Zool. Soc. San Diego, no. 17, pp. 1–95.

MAYR, ERNST

1942. Systematics and the origin of species. New York, Columbia University Press.

SIMPSON, GEORGE GAYLORD

1941a. Large Pleistocene felines of North America. Amer. Mus. Novitates, no. 1136, pp. 1–27.

1941b. Range as a zoological character. Amer. Jour. Sci., vol. 239, pp. 785–804.

1942. A standard frequency distribution method. Amer. Mus. Novitates, no. 1190, pp. 1–19.

SIMPSON, GEORGE GAYLORD, AND ANNE ROE

1939. Quantitative zoology. New York, McGraw-Hill Book Co.

SNEDECOR, GEORGE WADDEL

1946. Statistical methods. Applied to experiments in agriculture and biology. Ames, The Iowa State College Press.

THURSTONE, M. E.

1928. The fundamentals of statistics. New York, The Macmillan Co.

Many other useful references are given in the bibliographies of these books and papers.